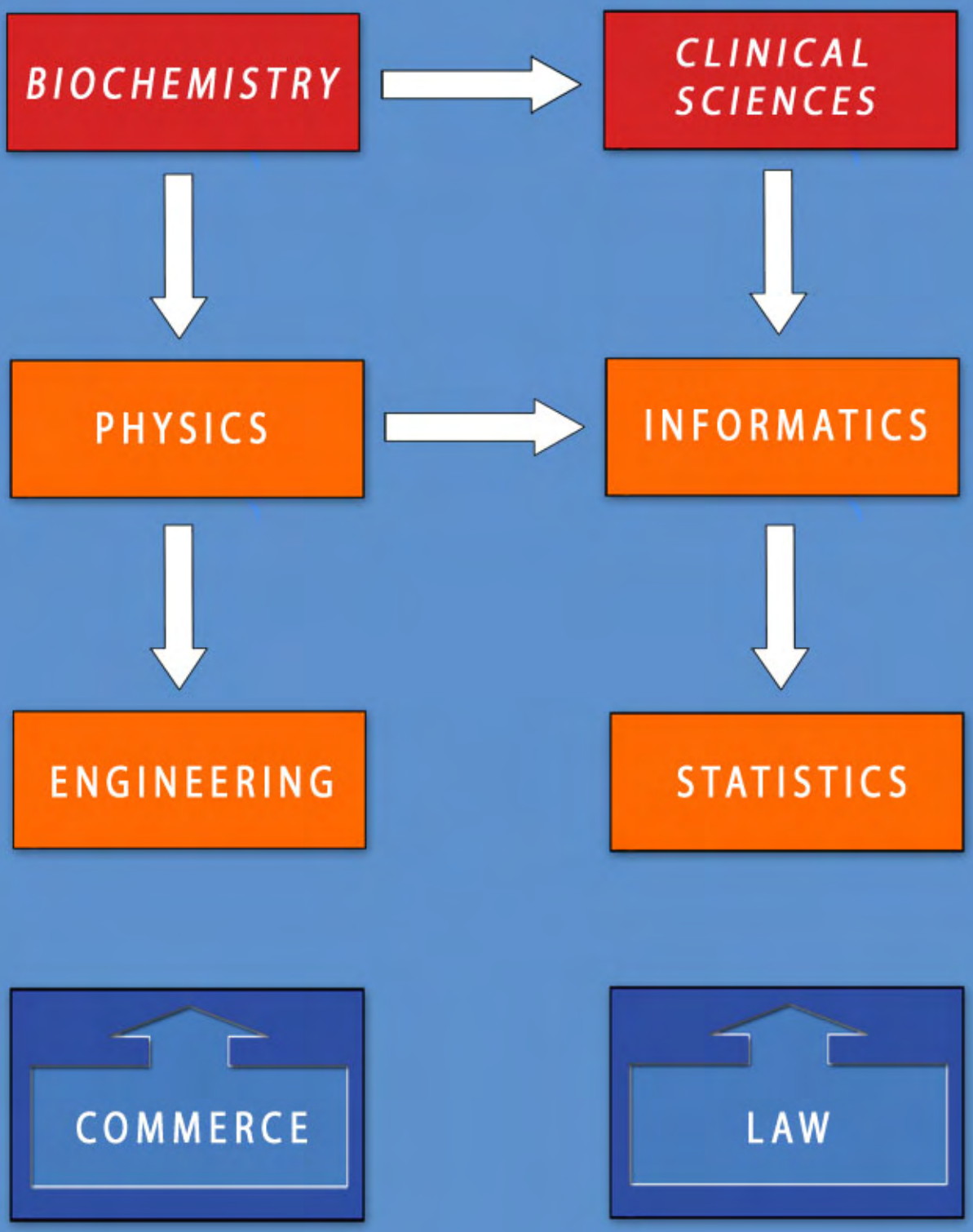
BIO-ALGORITHMS AND MED-SYSTEMS

Vol. 1, No. 1-2
2006



JOURNAL EDITED BY COLLEGIUM MEDICUM - JAGIELLONIAN UNIVERSITY – CRACOW – POLAND
http://www.bams.cm-uj.krakow.pl

SCOPE

Bio-Algorithms and Med-Systems is published semiannually by Collegium Medicum – Jagiellonian University through the Academic Library of the Jagiellonian University.

Bio-Algorithms and Med-Systems aims to cover a whole spectrum of interdisciplinary subjects, particularly those focused on medicine. The development of highly specialized disciplines has resulted in their isolation. Collaboration between them is needed to produce better practical applications.

Bio-Algorithms and Med-Systems is intended to present scientific problems that have developed in one discipline and which require collaboration with researchers in other fields.

At the end of their manuscripts, the authors of papers for the journal are encouraged to present a chart for collaboration among groups representing the disciplines needed as partners. An example of such a chart is presented on the cover of the volume.

The author of the opening paper may propose a system for collaboration, presenting the disciplines deemed necessary to accomplish the project. A partner who decides to join a particular system should cite the original paper. The partnership may be extended beyond the original system proposed by the author of the opening paper.

Bio-Algorithms and Med-Systems is published in traditional printed form as well as in an electronic version.

Standards for the printed version:

1. No color illustrations
2. Text length maximum 12 pages of 12 points Times New Roman, double spaced, including tables, figures and reference list
3. Text length for e-learning papers is limited to 4 pages of double columns: one column in English and the other in the author's native language.

Electronic version:

1. Any number of color figures allowed
2. No text length limitation
3. Include an e-learning description of the problem to facilitate the participation of non-specialist collaborators

The language of the journal is English, language correctness is the author's responsibility (only the first volume accepts Polish texts).

The E-learning papers as described above should satisfy the following expectations:

1. Project-related – aimed at mutual education of the participants in the project.
2. e-learning not related to the project should present new discoveries, historical analysis, important issues.
3. General e-learning publications should contain the following parts:

   A – why the problem was selected,
   B – what is new in the presented discipline,
   C – to whom the E-learning material is addressed,
   D – list of references not more than 4 years old,
   E – figures and tables common for both columns (English and author's native language)

All papers are evaluated and the reviewers' opinions are also printed to allow continuation of the discussion.

The example of the collaboration system shown on the cover page can be understood as follows:

1. An author specialist in biochemistry presents a new technique for medical diagnosis and patient monitoring during therapy (chemo- and/or radio-therapy)
2. To verify the results of the new technique, collaboration with a clinical medicine physician is sought – when one responds, he or she writes a new paper reporting how the method works
3. A specialist in physics is needed to construct a physical model in a version for mass production
4. A specialist in informatics is needed to prepare a computer program to evaluate the graphic version of the results (electrophoresis/chromatography images)
5. A specialist in statistics is needed to make a comparative (versus other techniques) and complex analysis of the newly introduced technique
6. On the diagram, the disciplines sought are distinguished from those already collaborating in the project
7. "E" symbols inform that the e-learning system is presented in an electronic version of the paper
8. The number of partners is not limited

The participation of companies is also possible. The conditions for collaboration are set individually according to legal regulations. Legal expertise will also be presented in the form of a paper.

The discussion network for the participants in each project can be organized according to the participants' expectations (in encrypted form if necessary).

The legal expertise as well as the discussion network (including encryption) is the editorial board's responsibility.

Additional information can be found at www.bams.cm-uj.krakow.pl

SŁOWO WSTĘPNE

W maju 2005 roku odbyła się w Krakowie kolejna konferencja z cyklu „Modelowanie Cybernetyczne Systemów Biologicznych”. Konferencja ta, organizowana co 5 lat, począwszy od 1975 roku rejestrowała wiele zmian, podążając za rozwojem różnych dyscyplin naukowych, od cybernetyki (a także filozofii) poczynając a na bioinformatyce i korzystającej z rozwiązań czysto technicznych i inżynierskich telemedycynie kończąc.

Cechą charakterystyczną wszystkich odbytych spotkań jest działalność uczestników w środowisku wysoce interdyscyplinarnym. Punktem centralnym jest jednak medycyna w jej czysto praktycznym znaczeniu, ale też w ujęciu naukowym – odwołującym się do analizy zjawisk na poziomie molekularnym, poszukującej mechanizmów funkcjonujących w układach żywych odpowiedzialnych za poszczególne funkcje biologiczne.

Podczas kolejnych sesji stawało się jasne, że wiele dzieje się w prezentowanych dyscyplinach. Okazało się też, że brakuje kontaktów pomiędzy dyscyplinami, które współpracując mogłyby doprowadzić do konkretnych efektów w postaci nowych technologii, nowych praktycznych rozwiązań w postaci sprzętu medycznego itp. Niektóre rozwiązania informatyczne mogłyby świetnie wspomagać pewne rozwiązania w dziedzinie diagnostyki klinicznej czy też projektowaniu sprzętu rehabilitacyjnego.

W końcowej, podsumowującej konferencję, dyskusji padł pomysł powołania pewnego interdyscyplinarnego forum, umożliwiającego wzajemne kontakty przedstawicieli różnych dziedzin naukowych, w tym także kontakty robocze. Nie wyklucza się także udziału w tymże forum firm, które – zainteresowane pewnymi rozwiązaniami technicznymi – mogą znaleźć miejsce dla siebie jako producenci sprzętu, bazującego na inżynierskich rozwiązaniach proponowanych przez specjalistów omawianych dyscyplin. Mamy nadzieję, że takim forum wymiany myśli o interdyscyplinarnym charakterze stanie się zaprezentowane czasopismo ***Bio-Algorithms and Med-Systems***.

Komitet Redakcyjny *Bio-Algorithms and Med-Systems* zaopatrzył publikacje dodatkowo w schematy, które w skrótowy sposób obrazują możliwość współpracy potrzebnej do pełnej realizacji projektu. Schematy te w sposób bardzo zwięzły i skrótowy wskazują na tematykę już podjętą i rozwijaną (co uwidocznione jest w postaci tekstu pisanego kursywą). Pozostałe okna (zwykły tekst) prezentują tematykę, której opracowanie jest oczekiwane przez potencjalnych partnerów i udziałowców w realizacji projektu. Struktura, w tym też liczba uczestników jest płynna i może podlegać stałej modyfikacji w zależności od rozwoju danego projektu.

Docelowo czasopismo będzie wydawane w języku angielskim. Wyjątkowo pierwszy tom obejmujący głównie prace związane z prezentacjami w ramach konferencji wydany został ze znaczną przewagą artykułów w języku polskim. Plany drukowania także recenzji będą realizowane, poczynając od drugiego tomu.

Mamy nadzieję, że czasopismo, przy znaczącym udziale Czytelników, osiągnie swój cel w postaci rozpropagowania aktywności naukowej w trudnej sferze działalności interdyscyplinarnej.

Zapraszamy potencjalnych partnerów do podejmowania tematyki zaprezentowanej na stronach tego czasopisma.

Irena Roterman-Konieczna

Kraków, 20 września 2005

Contents / Spis treści

OPENING PAPER / ARTYKUŁ OTWIERAJĄCY

1 Einstein's miraculous year 1905
Lucjan Piela

ARTIFICIAL INTELLIGENCE / SZTUCZNA INTELIGENCJA

3 Analiza możliwości zastosowania metod sztucznej inteligencji w medycynie sądowej
Elżbieta Filipowicz, Joanna Kwiecień, Małgorzata Kłys, Bogusław Filipowicz

9 Porównanie metody kNN i techniki sieci neuronowych w klasyfikacji badań spirometrycznych
Joanna Grabska-Chrząstowska, Wiesław Libuszowski, Waldemar Tomalak

13 Podobieństwa samooptymalizujących sieci neuronowych do biologicznych sieci neuronowych
Adrian Horzyk

21 The usefulness of pattern recognition methods and artificial neural networks in occupational medicine
Emilia Kolarzyk, Jacek Kwiatkowski, Jadwiga Helbin, Agata Pietrzycka

27 Badanie właściwości asocjacyjnych maszyny LSM złożonej z neuronów Hodgkina-Huxleya
Siemowit Muszyński, Wiesław A. Kamiński

31 Algorytmy ewolucyjne w realizacji systemów uczących się
Leszek Nowak, Mariusz Święcicki, Piotr Wais

37 Reinforcement learning algorithm based on psychological model of autonomous system
Mariusz Święcicki, Piotr Wais, Wiesław Wajs

45 Zastosowanie standardu HL7 dla automatycznego pozyskiwania wiedzy i przetwarzania jej w drodze analizy analogii przyczynowo-skutkowych
Michał Turek

51 Zastosowanie sieci immunologicznej do drążenia danych
Wiesław Wajs, Mariusz Święcicki, Piotr Wais, Hubert Wojtowicz, Paweł Janik

57 Abductive logic programming in allergy diagnosis
Piotr Walecki, Wojciech Lasoń, Grzegorz Porębski, Krystyna Obtułowicz

61 Systemowy algorytm ewolucyjny (SAE)
Jerzy Rudolf Tchórzewski

BIOCYBERNETICS / BIOCYBERNETYKA

65 Wirtualny układ oddechowy i jego wykorzystanie w testowaniu metod sztucznej wentylacji i wspomagania oddychania
Tomasz Gólczewski, Marek Darowski

73 Nowe spojrzenie na genezę pierwotnego nadciśnienia tętniczego – hipoteza cybernetyczna
Tomasz Gólczewski

81 Na autonomiczny układ nerwowy wpływa dynamika zmian przepływu krwi
Tomasz Gólczewski

87 Analiza aktywności fali wolnej w przewodzie pokarmowym szczura z zastosowaniem metody FFT
Aleksandra Jung, Grzegorz Królczyk, Piotr Thor

89 Patogenny deficyt motoryczny snu i koncepcja prewencji oddziaływaniami harmonicznymi
Maciej Pokora

99 Neuromodulacja błędna czynności motorycznych żołądka
Jacek Sobocki, Piotr Thor, Grzegorz Królczyk, Roman M. Herman

103 Tympanometria wieloczęstotliwościowa a cybernetyczna fazowość analizatora słuchowego
Paweł Trąbka-Zawicki, Elżbieta Reroń, Jan Trąbka

BIOINFORMATICS / BIOINFORMATYKA

107 Training set size in ensemble feature selection for clinical proteomics
Tomasz Arodź

111 Wykorzystanie bioinformatyki w biotechnologii
Anna Dobrowolska

113 Kompleksy białek i układów ciekłokrystalicznych. Możliwość zastosowań praktycznych
Anna Drozd, Leszek Konieczny, Janina Rybarska, Barbara Stopa, Barbara Piekarska, Irena Roterman, Marcin Król, Paweł Spólnik, Grzegorz Zemanek

117 Konstrukcja sekwencji konsensusowych rodziny białek β-spektryn o różnych parametrach progowych oraz weryfikacja ich użyteczności
Anna Fogtman, Jacek Leluk, Bogdan Lesyng

121 Implementacja oszacowania stopnia podobieństwa białek na poziomie struktury pierwszorzędowej oraz konstrukcja drzew filogenetycznych dla białek homologicznych
Elżbieta Gajewska, Jacek Leluk

125 Zmienność mutacyjna w homologicznych rodzinach kinaz
Jacek Kuśka, Jacek Leluk, Bogdan Lesyng

129 Analiza danych z mikromacierzy DNA – metody, narzędzia
Henryk Maciejewski, Łukasz Konarski, Anna Jasińska, Maria Drath

133 Tworzenie i rozdział kompleksów białek surowicy z bawnikami supramolekularnymi w elektroforezie dwukierunkowej – próba opracowania komputerowej techniki analizy obrazu rozdziałowego
Janina Rybarska, Paweł Spólnik, Anna Drozd, Leszek Konieczny, Barbara Piekarska, Barbara Stopa, Grzegorz Zemanek, Marcin Król, Irena Roterman

BIOMEDICAL ENGINEERING / INŻYNIERIA BIOMEDYCZNA

139 Tworzenie systemu wspomagającego organizowanie konferencji medycznych zgodnie z modelem kaskadowym cyklu życia oprogramowania
Michał Całka, Przemysław Guzik, Jan Kniat, Grzegorz Konieczka, Radosław Luboch, Karol Sawczyk, Edward Jacek Gorzelańczyk

143 Analiza biomechaniczna sztucznego krążka międzykręgowego
Paweł Borkowski

145 Modelowanie sztucznych układów mięśniowych z aktuatorami pneumatycznymi
Ryszard Dindorf

157 Zintegrowane symulatory chodu w zastosowaniu do syntezy algorytmów sterowania
Piotr Kaczmarek, Andrzej Kasiński

165 Wspomagany komputerowo wielościeżkowy system egzaminowania studentów medycyny
Andrzej Kononowicz, Marta Żabińska-Popiela, Irena Roterman-Konieczna, Józef Krzysiek

171 Zastosowanie wielorozdzielczej analizy sygnału do opisu aktywności mięśnia podczas chodu
Przemysław Mazurkiewicz

175 Metody badań ekspresji mowy u osób po usunięciu krtani
Rafał Pietruch, Marta Michalska, Wiesław Konopka, Antoni Grzanka

183 Metoda wymuszania ujemnych impulsów ciśnienia w pomiarach parametrów diagnostycznych układu oddechowego
Piotr Piwowar

191 Aplikacja oparta na bibliotece VTK wspomagająca zabiegi bronchoskopowe
Mirosław Socha, Mariusz Duplaga

197 Stabilność endoprotez cementowych stawu biodrowego w świetle badań histopatologicznych, numerycznych i eksperymentalnych
Jerzy Włodarski

MODELING AND SIMULATION / MODELOWANIE I SYMULACJA

205 Comparison of dialysis adequacy indices calculated according to one and two compartment model
Małgorzata Dębowska, Jacek Waniewski

209 Compartmental model of fluid and solute absorption in peritoneal dialysis
Magda Gałach, Andrzej Weryński

213 Symulacyjny model rodziny pszczelej
Andrzej Izworski, Andrzej Migacz, Marcin Michałek

217 Model zespołu synaps zapamiętujących komórki piramidowej hipokampa
Edward Jacek Gorzelańczyk, Michał Huflejt, Jan Kniat, Ewa Laskowska, Janusz Murakowski, Piotr Woźniak

221 Modele regresyjne uprawiania turystyki aktywnej przez osoby niepełnosprawne wzrokowo i ruchowo
Krzysztof Kaganek, Helena Stanuch

227 Modelowanie poziomu ekspresji genu TNFα,TNF RI,TNF RII w komórkach raka jelita grubego
Teresa Kokot, Ewa Nowakowska-Zajdel, Janusz Wierzgoń, Andrzej Kozowicz, Małgorzata Muc-Wierzgoń

233 Model symulacyjny wymiany gazowej tlenu
Wojciech Lasoń, Piotr Bogdał, Piotr Walecki, Wiesław Pyrczak

237 Terapia fotodynamiczna nowotworów upigmentowanych. Modelowanie własności optycznych melaniny metodami Monte Carlo
Zenon Matuszak, Aneta Sawow, Marta Wasilewska-Radwańska

241 Przystosowanie modelu systemu wzrokowego muchy do analizy ruchu drogowego
Zbigniew Mikrut, Michał Zajączkowski

247 Ontologia mammografii – struktura modelu, definicje i instancje pojęć
Teresa Podsiadły-Marczykowska, Anna Guzik

253 Neuronowy model doświadczenia Pawłowa. Badanie zespołu odruchów warunkowych
Krzysztof Sokalski

259 Bistability in epileptic phenomena
Piotr Suffczyński, Stiliyan Kalitzin, Fernando Lopes da Silva

267 ADD, HKD i ADHD w świetle hipotezy dipolowej
Piotr Walecki, Jan Trąbka

273 Modelowanie transportu obrazu modyfikowanymi włóknami polimerowymi
Mariusz Wójcik

277 Optymalizacja kodu Monte Carlo do symulacji czasu przelotu fotonów w strukturach tkankowych
Norbert S. Żołek, Adam Liebert, Roman Maniewski

PATTERN RECOGNITION / ROZPOZNAWANIE WZORÓW

281 Pola elektromagnetyczne w badaniach biomedycznych – układy ekspozycyjne i modelowanie numeryczne
Paweł Bieńkowski, Tomasz Długosz

287 Komputerowe metody detekcji nowotworów piersi w zdjęciach mammograficznych
Witold Dzwinel, Krzysztof Boryczko, Tomasz Arodź, Marcin Kurdziel

291 Analiza obciążeń pracą na podstawie wskaźnika wykorzystania rezerwy tętna i obciążeń statycznych metodą OWAS
Anna Groborz, Tadeusz Juliszewski, Marcin Gonciarz

297 Nowe metody pomiaru i analizy patologicznego drżenia kończyn górnych
Andrzej Izworski, Marcin Michałek, Monika Rudzińska

301 Are the earth magnetic field and Schumann resonance matter to global human activity?
Stanisław Micek, Grzegorz Micek

307 Lokalizacja kluczowych cech rozpoznawanych obiektów na podstawie analizy ścieżki wzrokowej obserwatora
Zbigniew Mikrut, Piotr Augustyniak

311 Wykorzystanie gramatyk grafowych do syntaktycznego opisu przestrzennych struktur naczyń wieńcowych
Marek R. Ogiela, Mirosław Trzupek

317 Komputerowa analiza obrazu w ocenie zmian zachodzących w strefach granicznych: kość-cement-implant
Jerzy Włodarski

321 Neuronal movement detector in the model of simulated mammalian visual system
Grzegorz M. Wójcik, Wiesław A. Kamiński

325 Problemy związane z topografią elektrod dla badań EEG – system ich digitalizacji
Tomasz Zyss

333 Zastosowanie videodermatoskopii w diagnostyce zmian barwnikowych skóry
Ewa Żabińska-Płazak, Anna Wojas-Pelc, Grzegorz Dyduch

339 Szacowanie istotności statystycznej zmian w mocy sygnału EEG związanego z bodźcem
Jarosław Żygierewicz, Piotr Durka, Joanna Mazurkiewicz, Nathan Crone, Piotr Franaszczuk

TELEMEDICINE / TELEMEDYCYNA

349 Automatyczny system zdalnego nadzoru kardiologicznego oparty na profesjonalnych relacjach wśród lekarzy
Piotr Augustyniak

355 Usprawnienie przewlekłej opieki nad chorymi na astmę oskrzelową przy pomocy narzędzi teleinformatycznych
Mariusz Duplaga, Grażyna Bochenek, Łukasz Kasper, Magdalenia Twardowska, Ewa Niżankowska-Mogilnicka, Andrzej Szczeklik

361 Telekonsultacje chirurgiczne na obecnym etapie rozwoju teleinformatyki medycznej w Polsce na przykładzie regionu Małopolska
Janusz Legutko, Ewa Woźniakiewicz, Stanisław Grzesik, Mirosław Szura

365 Zastosowanie kryptografii w telekonsultacji kardiologicznej
Krzysztof Sarapata, Grzegorz Bilo, Wiesław Pyrczak, Andrzej Kononowicz, Irena Roterman-Konieczna, Kalina Kawecka-Jaszcz

PHILOSPHY / FILOZOFIA

369 Cybernetyczne związki nauki ze sztuką
Jan Trąbka, Paweł Trąbka-Zawicki

373 Genetic Immortality in the light of Holistic Pedagogy
Andrzej Szyszko-Bohusz

375 Znaczenie modeli w hipotezie Gai
Honorata Korpikiewicz

381 O relacji liczby do rzeczy
Teresa Grabińska, Mirosław Zabierowski

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 1-2.

Einstein's miraculous year 1905

In the paths of humanity there are sometimes real breakthroughs of incredible consequences. The case of Albert Einstein is certainly of this kind.

Einstein was born in Ulm (Germany) in 1879 and graduated from the Technical University of Zurich. As a teenager he was similar to his peers, not showing any extraordinary abilities. Maybe,e he could be distinguished from colleagues by having a higher level of self-confidence and a strongly marked independence of mind. This is perhaps the reason why he started his career as a simple clerk in the Berne Patent Office. Being afraid of his supervisor, he used to read physics books hidden in a drawer, which he later called the "Department of Physics".

In these circumstances it is difficult to understand, how it was possible that this young man could have had ideas that would later overthrow nineteenth century physics. In 1905, at the age of 26, Einstein published three fundamental papers: the theory of relativity, Brownian motion, and the photoelectric effect. These papers were truly revolutionary, surpassing anything published by Einstein later. The first work introduced new understanding of space and time; the second explained a strange permanent motion of small particles by irregular collisions with molecules of the neighbourhood. The third paper treated light as... particles, whereas there existed thousands of experiments showing the wave-like characteristic of light. Robert Millikan decided to falsify experimentally Einstein's hypothesis about light. However, after ten years of investigations, Millikan acknowledged that he could no longer fail to support Einstein's explanation, "however absurd it may look". This conversion of a sceptic inclined the Nobel Committee to grant Einstein the Nobel Prize in 1923 ''for his work on the elementary charge of electricity and on the photo-electric effect". Today we know that light consists of particles (photons), we may even count them one by one (they may exhibit however their wave-like nature as well). It is worth noting that Isaac Newton was the first to think of light as particles, but his hypothesis seemed to be overthrown by the wave theory of light (diffraction, interference, etc.).

The consequences of Einstein's discoveries in the miraculous year 1905 are enormous, especially in our understanding of the Universe, but also in practical applications. As it turned out later, the very existence of atoms and molecules is related to their property called the spin, which is of purely relativistic nature. Thus, the world as we see it simply could not exist if the relativistic effects were switched off (i.e. if the speed of light was infinite). Einstein's relativity theory has practical every-day applications in nuclear physics (related also to nuclear medicine). Brownian motions are observed in microscopes, and therefore indirectly related to biological and medical practice. The photoelectric effect explained by Einstein has direct consequences in biology and medicine.

One hundred years after Einstein's miraculous year 1905, the photoelectric detectors are ubiquitous in high-tech medical equipment, so crucial in medicine.

What can we learn from Einstein's story? I think the most important message is that in science nothing should be treated as evident, and that human ideas are able to break courageously all limits and be incredibly bright and independent. This could open new possibilities, which, as in Einstein's case, may change profoundly mankind's perspectives.

Starting a new interdisciplinary scientific journal represents however an occasion to pose a related question to the medical and technical communities. This kind of question was confronted by Einstein years after his breathtaking theoretical discoveries, immediately after nuclear bombs were dropped in Hiroshima and Nagasaki. The destructive energy of the bomb comes from Einstein's famous $E=mc^2$. Should medical research be as courageous and absolutely independent, breaking anything we know, like Einstein's ideas did? In my opinion the honest answer should be "yes, it should be courageous, but not absolutely free, not at any price". In any science, and first of all in biology and medicine, we have to be *at the same time courageous and wise*, because permitting any experiment would open the path to experiments on you and me. May the scientists who will make the present journal future development think as independently as Einstein, and act in such a way as to preserve human dignity: that of their patients and their own.

Lucjan Piela
Professor of Chemistry
Warsaw University

Warsaw, 12th September, 2005

OPENING PAPER

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 3-8.

ANALIZA MOŻLIWOŚCI ZASTOSOWANIA METOD SZTUCZNEJ INTELIGENCJI W MEDYCYNIE SĄDOWEJ

ANALYSIS OF ARTIFICIAL INTELLIGENCE METHODS APPLICATION IN FORENSIC MEDICINE

ELŻBIETA FILIPOWICZ*, JOANNA KWIECIEŃ**, MAŁGORZATA KŁYS***, BOGUSŁAW FILIPOWICZ**

**Szpital Uniwersytecki, Collegium Medicum Uniwersytetu Jagiellońskiego, Kraków, filip@ia.agh.edu.pl*
***Katedra Automatyki Akademii Górniczo-Hutniczej, Kraków, filip@ia.agh.edu.pl*
****Katedra Medycyny Sądowej, Collegium Medicum Uniwersytetu Jagiellońskiego, Kraków, mpklys@cyf-kr.edu.pl*

Streszczenie. Artykuł opisuje zastosowanie nowych systemów bioinformatycznych w medycynie sądowej. Systemy biometryczne już znalazły zastosowanie między innymi w geometrycznej identyfikacji twarzy. Są one niezbędne w identyfikacji osobowej oraz w poszukiwaniu osób zaginionych. W ostatnich latach nastąpił bardzo szybki rozwój badań nad chromosomem Y, co pozwoliło na postęp w medycynie sądowej. Metody sztucznej inteligencji mogą być pomocne w badaniach spornego ojcostwa i ewolucji człowieka. Bioinformatyka oraz metody biologii molekularnej, metody multiplex PCR będą pełnić bardzo ważną funkcję w analizie sądowej DNA oraz w historycznych i genealogicznych badaniach.

Słowa kluczowe: medycyna sądowa, systemy biometryczne, chromosom Y, multiplex PCR, pojedyncze nukleotydowe polymorfizmy

Abstract. This article reviews new bioinformatic systems in forensic medicine. Biometric systems are already used in forensic medicine for face geometry identification. They are necessary in personal identification and missing persons investigations. The field of Y-chromosome analysis and its application to forensic science has undergone rapid improvement in recent years. Artificial Intelligence is usefull in paternity testing and in human evolutionary study. Bioinformatic systems and multiplex PCR assay will play an important role in the future of forensic DNA typing and historical and genealogical research.

Key words: forensic medicine, biometric systems, Y-chromosome, multiplex PCR, single nucleotide polymorphisms

1. Możliwości zastosowania metod sztucznej inteligencji w medycynie sądowej.

Systemy biometryczne mogą być stosowane do identyfikacji osobowej w medycynie sądowej. Systemy biometryczne są obecnie jednym z najszybciej rozwijających się działów informatyki na świecie. Kontrola biometryczna oparta jest na specyficznych cechach organizmu, które są charakterystyczne dla każdego człowieka [2], [3], [10], [11]. W chwili obecnej najpopularniejsze techniki biometryczne można podzielić na następujące grupy:

- systemy oparte o rozpoznawanie linii papilarnych,
- systemy oparte o rozpoznawanie geometrii twarzy,
- systemy oparte o rozpoznawanie mowy,
- systemy oparte o rozpoznawanie cech charakterystycznych tęczówki oka.

Aktualnie dostępne są nowe systemy:

- oparte o rozpoznawanie DNA,
- oparte o rozpoznawanie obrazu żył,
- oparte o rozpoznawanie dna oka.

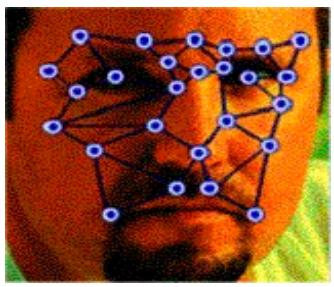
Systemy biometryczne są potrzebne w medycynie sądowej celem identyfikacji osobowej. Są proste w instalacji, niezawodne i tanie, jeśli chodzi o koszty utrzymania oraz konserwacji. Ogólne zasady biometryki opierają się na zapisywaniu w pamięci komputera określonego i niepowtarzalnego wzoru wybranej cechy, przechowywaniu tej informacji, a następnie dopasowaniu określonej cechy podczas procesu weryfikacji w oparciu o zapisany wzorzec [2], [3], [10], [11].

Identyfikacja w medycynie sądowej polega więc na automatycznym rozpoznaniu nieznanej osoby poprzez badanie jednej lub kilku jej cech biometrycznych. System porównuje aktualny obraz zapisany przez odpowiednie urządzenie z wzorcami zapisanymi w scentralizowanej bazie danych. Liczone jest prawdopodobieństwo przyporządkowane każdemu obrazowi. Jeśli przekracza ono ustaloną wartość krytyczną, system uznaje, że nieznana osoba została zidentyfikowana. Jeśli kilka obrazów przekracza ową wartość, przyjmuje się najbardziej prawdopodobny obraz za prawdziwy.

Linie papilarne są bardzo ważną cechą każdego ludzkiego organizmu, gdyż pozwalają jednoznacznie zidentyfikować człowieka. Podstawę identyfikacji odcisków palców stanowią tzw. minutie. Są to punkty, które zaznaczają początek i koniec charakterystycznych miejsc oraz miejsca przecięcia, widoczne nawet gołym okiem. Czytniki wykorzystujące systemy biometryczne pozwalają na szybkie i dokładne identyfikowanie danej osoby. Istnieją dwa rodzaje takich urządzeń: w jednym weryfikacja odbywa się w samym czytniku, w drugim zaś proces ten odbywa się w komputerze PC połączonym z małym skanerem, niewymagającym trudnego oprogramowania. Niektóre czytniki działają na zasadzie fotografowania opuszki palca przyłożonego do czytnika. Kod opisujący palec może być zapisany jako ciąg symboli liter i cyfr, a więc kodem ASCII. Możemy go bez trudu przechowywać w dowolnej bazie danych, a na dodatek zajmuje on bardzo mało miejsca w pamięci [10], [11].

Ludzkie oko posiada wiele cech, które mają zastosowanie w biometryce. Nie istnieją dwie tęczówki, których szczegółowy opis matematyczny byłby identyczny, nawet bliźnięta jednojajowe mają różne tęczówki. U każdego człowieka występują odmienne tęczówki w prawym i w lewym oku; są niezmienne począwszy od 18 miesiąca życia aż do śmierci. Unikalność jest jednym z najważniejszych problemów dotyczących systemów biometrycznych. Obrazem jest siateczka tkanki łącznej oraz innych widzianych elementów. Taki kod, zawierający skrócony opis punktów charakterystycznych, jest następnie porównywany z zapisanym obrazem w bazie danych [10], [11].

Systemy rozpoznające geometrię twarzy są najbardziej naturalnymi sposobami identyfikacji biometrycznej. Technologia rozpoznawania twarzy obecnie jest rozwijana w dwóch kierunkach: pomiaru twarzy i tzw. metody eigenface (właściwych twarzy). Technologia pomiaru twarzy polega na pomiarze specyficznych cech twarzy i relacji pomiędzy tymi pomiarami [10], [11]. Punkty pomiarowe pokazano na rys. 1.

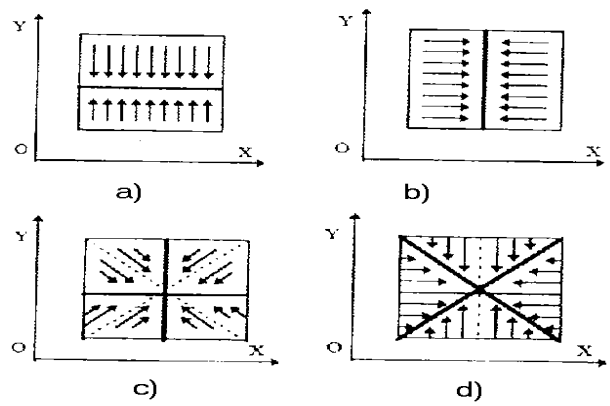


Rys. 1. Przykład punktów pomiarowych twarzy

Metoda eigenface polega na porównywaniu uzyskanego obrazu z gotowymi wzorcami umieszczonymi w pamięci. Jest podobna do metody stosowanej w kryminalistyce czyli portretów pamięciowych. Technologia identyfikacji opierającej się na eigenface jest w początkowym stadium rozwoju i jest bardzo obiecująca.

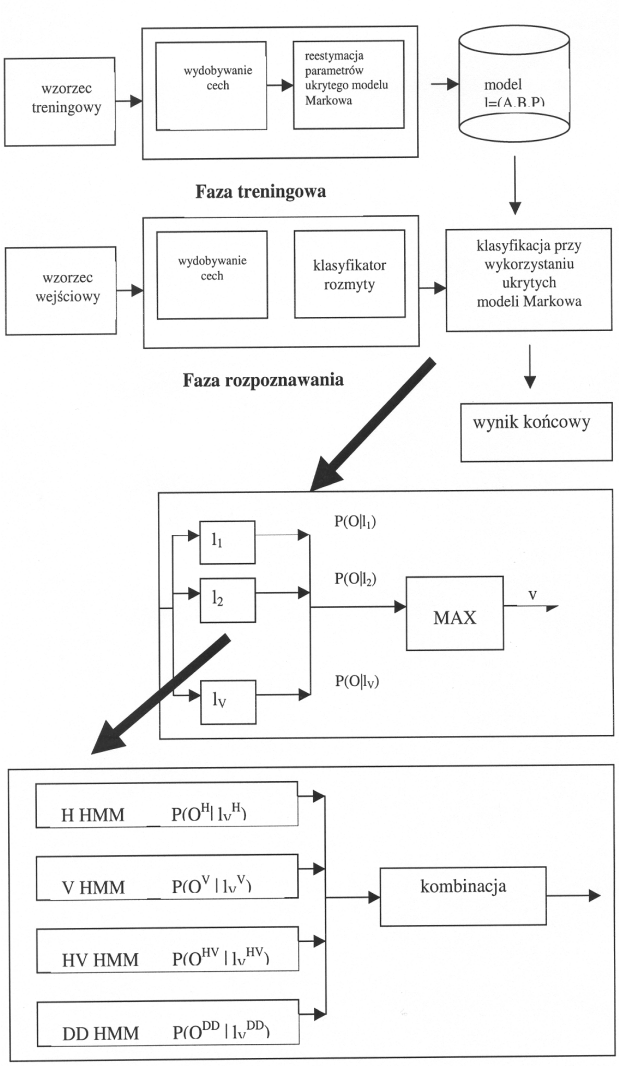
Tworzenie modeli stochastycznych jest elastyczną i najbardziej ogólną metodą wykorzystywaną do zagadnień systemów biometrycznych. Istnieje wiele nie w pełni jasnych i dających się z trudem zdefiniować aspektów, związanych z rozpoznaniem danej cechy. Modele probabilistyczne są bowiem najbardziej odpowiednie w rozwiązywaniu zagadnień związanych z występowaniem niepewnych bądź niekompletnych informacji. Modele stochastyczne są ukrytymi modelami Markowa, które uznane są za najbardziej przydatne do rozpoznawania właściwości cech. Gdy wzorzec sam służy jako odniesienie i jest przechowywany w pamięci, podejście bazujące na ukrytym modelu Markowa reprezentuje odniesienie przez model, tym samym wykazując wyższy stopień abstrakcji i elastyczności [2], [3], [4].

Ukryty model Markowa HMM jest procesem stochastycznym w podwójnym sensie, ze szczególnym uwzględnieniem procesu stochastycznego, który nie jest obserwowany a ukryty, a pewne informacje o nim można wydobyć, posługując się innym zbiorem procesów stochastycznych, które produkują sekwencje obserwowanych symboli. Ukryty model Markowa jest scharakteryzowany przez łańcuch Markowa o skończonej liczbie stanów i zbiór dystrybucji wyjściowych. Parametry przejścia pomiędzy stanami w łańcuchu Markowa modelują zmiany widmowe. Analiza tych dwóch typów zmian jest podstawą, na której opiera się rozpoznawanie właściwości cech tęczówki czy geometrii twarzy. W celu zdefiniowania symboli do utworzenia ukrytego modelu Markowa, musi być opracowany proces wydobywania cech oparty na analizie konturu. Aby ułatwić przetwarzanie wzorców, zastosowano technikę zwaną transformacją miejscowej projekcji konturu RPCT, która transformuje złożony wzór lub wzór wielokonturowy w jeden zewnętrzny kontur. Przy wykorzystaniu tej techniki miejscowa projekcja i przetwarzanie konturu są dokonywanie współbieżnie. Podstawowa zasada, na której się opiera ta metoda polega na tym, że wszystkie piksele wzorca są rzutowane na podstawy projekcyjne (ang. *projection bases).* Następnie dokonywana jest ekstrakcja z wzorca łańcucha konturu (ang. *contour chain*). Wszystkie te operacje przekształcają obraz w jeden kontur. Metoda RPCT może zostać podzielona na cztery typy technik transformacyjnych, w zależności od wyboru baz projekcyjnych, jak przedstawiono na rys. 2.



Rys. 2. Cztery typy przekształcenia RPCT: a) HRPCT – pozioma (ang. *horizontal RPCT*), b) VRPCT – pionowa (ang. *vertical RPCT*), c) HVRPCT – poziomo-pionowa (ang. *horinzontal-vertical RPCT*), d) DDRPCT – diagonalno-diagonalna (ang. *diagonal-diagonal RPCT*)

Rys. 3 przedstawia schemat blokowy ogólnej struktury systemu rozpoznawania znaków należących do zbiorów znaków o dużej liczności. Proces rozpoznawania znaków składa się z fazy treningowej i fazy klasyfikacji. Zaproponowano również użycie preklasyfikatora, który odgrywa istotną rolę w procesie wstępnej redukcji kandydatów do dalszego rozpoznawania.



Rys. 3. Struktura systemu rozpoznawania cechy

Podsumowując, ukryty model Markowa opiera się na grafie skierowanym o skończonej liczbie węzłów zwanych stanami, przy czym dla każdego ze stanów dane są prawdopodobieństwa zdarzenia, że proces właśnie od tego stanu wystartuje oraz dane są prawdopodobieństwa przejść do stanów następnych [2], [3], [4].

W każdym ze stanów modelu dokonywana jest generacja jednego z możliwych do zaobserwowania symboli elementarnych oraz podane są prawdopodobieństwa generacji dla każdego takiego symbolu. Dla każdego ze znaków należących do danego języka tworzony jest oddzielny ukryty model Markowa, dla którego prawdopodobieństwo generacji tego znaku, składającego się z pewnych elementarnych symboli jest największe. Na początku, po przypisaniu początkowych wartości prawdopodobieństw, ukryty model Markowa musi zostać poddany fazie treningowej, w której modyfikowane są wartości tych prawdopodobieństw w oparciu o dostarczony przez uczącego zbiór znanych znaków. Po wykonaniu odpowiedniej liczby iteracji, ukryte modele Markowa są już gotowe do rozpoznania zadanych znaków. Faza ta jest zwana fazą testującą. Dla każdego z ukrytych modeli Markowa obliczane jest prawdopodobieństwo generacji przez niego zaobserwowanego znaku, przy czym na końcu wybierany jest model dający największe prawdopodobieństwo tego zdarzenia. Z ukrytymi modelami Markowa związanych jest kilka bardzo ważnych zagadnień [2], [3], [4]:

- dobór kryterium parametrów, liczby stanów oraz wybór typu modeli,
- dobór liczby obserwowanych elementarnych symboli,
- problem klasyfikacji wyników dostarczonych przez poszczególne modele,
- problem treningu modeli.

Opisane modele mają więc zastosowanie w medycynie sądowej, aktualnie pracuje się nad zastosowaniem systemów biometrycznych w analizie DNA w hemogenetyce.

Zastosowanie bioinformatyki w technikach biologii molekularnej pozwoliło na szybki rozwój hemogenetyki w medycynie sądowej [1], [7], [13], [14].

2. Metody i algorytmy sztucznej inteligencji

Sztuczne sieci neuronowe, algorytmy genetyczne i sztuczne systemy immunologiczne reprezentują grupę metod i technik, które w pewnym stopniu są symulacją rozwiązań stworzonych przez naturę.

Algorytm genetyczny operuje na zakodowanej części informacji. Gen jako potencjalna cecha dla systemów biometrycznych jest to potencjalny bit w algorytmach genetycznych [4], [5]. Chromosom to binarny ciąg kodowy składający się z zer i jedynek. Każdy pojedynczy bit jest odpowiednikiem pojedynczego genu. Ze względu na sposób ułożenia genów w chromosomie można wyróżnić trzy podstawowe sposoby kodowania:

- klasyczne – geny są heterogeniczne, czyli geny na różnych pozycjach przechowują różne informacje. Stosuje się je wtedy, gdy mamy określone niejednorodne cechy osobnika i chcemy dobrać im optymalne wartości.
- permutacyjne – geny są homogeniczne, czyli przechowują podobne informacje i są wymienialne. Tego typu kodowanie stosuje się do rozwiązywania problemów kombinatorycznych.
- drzewiaste – chromosom nie jest liniowym ciągiem genów, ale złożoną strukturą drzewiastą. Kodowanie drzewiaste znajduje zastosowanie w tzw. programowaniu genetycznym, czyli wszędzie tam, gdzie ewolucji podlegają reguły matematyczne

Genotyp jest to zbiór (struktura) jednego lub więcej chromosomów. Genotyp może być już pojedynczym osobnikiem danej populacji. Fenotyp występuje w przyrodzie po interakcji ze środowiskiem. Inaczej fenotyp to zbiór parametrów, rozwiązanie, punkt.

Algorytmy genetyczne (AG) są to algorytmy poszukiwania oparte na mechanizmach doboru naturalnego oraz dziedziczności [4], [5]. Algorytm genetyczny zawiera w sobie elementy teorii ewolucji Darwina, która zakłada, że przeżywają tylko najlepiej przystosowane osobniki. W każdym pokoleniu powstaje nowy zespół sztucznych organizmów (ciągów bitowych), utworzonych z połączenia fragmentów najlepiej przystosowanych osobników poprzedniego pokolenia. AG wykorzystują efektywnie przeszłe doświadczenia do określania nowego obszaru poszukiwań o spodziewanej podwyższonej wydajności.

Populacja jest to pewien zbiór osobników. Każdy algorytm genetyczny rozpoczyna działanie od początkowej populacji

ciągów kodowych, po czym generuje kolejne populacje ciągów. Najczęściej populację początkową dla AG wybiera się drogą losową. W każdym kolejnym cyklu w algorytmach genetycznych populacja ma stały rozmiar, wszystkie chromosomy podlegają wymianie na nowe. Elementarny algorytm genetyczny jest skonstruowany z trzech następujących operacji: reprodukcji, krzyżowania oraz mutacji.

Reprodukcja jest to proces, polegający na powieleniu indywidualnych ciągów kodowych w stosunku zależnym od wartości, jakie przybiera funkcja celu. Istnieją różne sposoby powielania ciągów. Najczęściej spotykanym sposobem jest metoda ruletki.

Krzyżowanie to operacja genetyczna, składająca się z dwóch etapów. Najpierw kojarzymy w sposób losowy ciągi kodowe z puli rodzicielskiej w pary, a następnie każda para przechodzi proces krzyżowania.

Mutacja polega na wymianie pojedynczego bitu w chromosomie. Prawdopodobieństwo wystąpienia mutacji w AG jest bardzo małe. Mutacja jest błądzeniem przypadkowym w przestrzeni ciągów kodowych. Jest stosowana tylko na wypadek utraty ważnych składników rozwiązania.

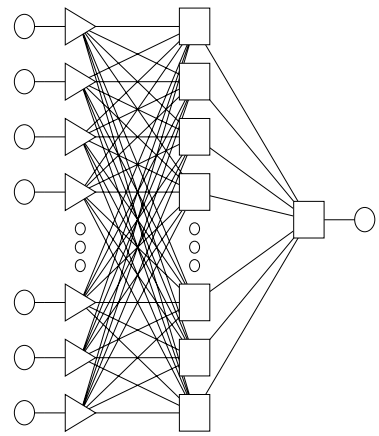
AG nie przetwarzają bezpośrednio parametrów zadania, lecz ich zakodowaną część [4], [5]. Prowadzą poszukiwania, wychodząc nie z pojedynczego punktu, ale z pewnej ich populacji. AG korzystają tylko z funkcji celu, nie zaś z jej pochodnych lub pomocniczych informacji. Ponadto stosują probabilistyczne, a nie deterministyczne reguły wyboru. Ważnym elementem w AG jest cel optymalizacji. Cel optymalizacji to zwiększenie efektywności aż do osiągnięcia pewnego optimum. Głównym celem optymalizacji jest ulepszenie. Czyli w AG optymalizacja sprowadza się do poszukiwania maksimum funkcji. W algorytmach tych dążymy do znalezienia globalnego maksimum, ale pewne odmiany potrafią znaleźć też optima lokalne [4], [5].

Pierwszym krokiem w AG jest utworzenie populacji początkowej. Polega on na wybraniu określonej liczby chromosomów, reprezentowanych przez ciągi bitowe określonej długości. Nasz algorytm rozpoczyna swoje działanie właśnie od tej wybranej populacji początkowej i generuje kolejne (z założenia coraz lepsze) populacje ciągów. Do nas należy ustalenie liczby populacji początkowej (należy pamiętać, że ta liczba nie może być zbyt mała ani zbyt duża). Po wprowadzeniu populacji początkowej przychodzi pora na ocenę osobników znajdujących się w populacji. Na tym etapie badamy cechy poszczególnych osobników [4], [5].

Krzyżowanie ma na celu wymianę materiału genetycznego pomiędzy dwoma osobnikami. Mutacja występuje z bardzo małym prawdopodobieństwem. Wystąpienie mutacji zależne jest od współczynnika mutacji.

Algorytm, teoretycznie, może się nie kończyć (działa w nieskończoność). Jednak przeważnie wprowadza się jakieś ograniczenia. Najczęściej stosowane ograniczenia to: uzyskanie wartości znanej wcześniej, określona liczbę iteracji oraz brak poprawy wyników. Zatrzymanie algorytmu zależy od zadania jakie wykonujemy [4], [5].

Sztuczne sieci neuronowe, będące bardzo uproszczonym modelem mózgu ludzkiego, składają się z dużej liczby jednostek – neuronów, posiadających umiejętność przetwarzania informacji. Każdy neuron wchodzący w skład sieci powiązany jest z innymi neuronami za pomocą łączy o parametrach (tzw. wagach synaptycznych) zmienianych w trakcie procesu uczenia i służących do komunikacji między neuronami. Na podstawie bieżącego stanu aktywacji neuronu i sygnałów wejściowych obliczany jest sygnał, jaki neuron wysyła do sieci poprzez jedno wyjście do pozostałych neuronów (węzłów sieci). W czasie transmisji sygnał ten podlega osłabieniu lub wzmocnieniu w zależności od charakterystyki łącza. Większość budowanych sieci neuronowych składa się z kilku warstw: wejściowej – służącej do wprowadzania do sieci danych wejściowych, wyjściowej – wyznaczającej końcowe rozwiązanie i ukrytych – przetwarzających sygnały w taki sposób, aby wydostać pewne dane pośrednie konieczne do wyznaczenia końcowego rozwiązania. Istnieje wiele rodzajów sieci, które różnią się między sobą strukturą i zasadą działania [8]. Najpopularniejszą obecnie strukturą sieci neuronowych są perceptrony wielowarstwowe (MLP), należące do grupy sieci z jednokierunkowymi połączeniami (rys. 4).



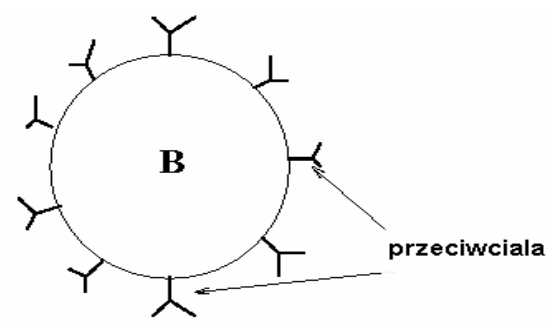
Rys. 4. Schemat trójwarstwowego MLP

Bardzo ważnym etapem całego procesu konstruowania tych sieci jest określenie właściwej liczby warstw i neuronów w warstwach. Okazuje się, że największe możliwości posiadają nieliniowe sieci neuronowe o co najmniej trzech warstwach, przy czym liczba neuronów w warstwach wejściowej i wyjściowej jest określona przez rozwiązywany problem, natomiast liczba neuronów w warstwie ukrytej zależy od złożoności problemu, od typu funkcji aktywacji neuronów tej warstwy, od algorytmu uczenia, od rozmiaru danych uczących. Najpowszechniejszą metodą uczenia perceptronów jest metoda wstecznej propagacji błędów. Chcąc właściwie przeprowadzić proces uczenia sieci, napotykamy na wiele trudności związanych m.in. z doborem odpowiedniego zbioru uczącego oraz parametrów uczenia sieci [8].

Powolność i uciążliwość procesu uczenia metodą wstecznej propagacji błędów w sieciach nieliniowych sprawiły, że pojawiły się również inne rodzaje sieci neuronowych, między innymi sieci rezonansowe (ART), sieci Hopfielda (ze sprzężeniem zwrotnym), sieci Kohonena, sieci o radialnych funkcjach bazowych (RBF), probabilistyczne sieci neuronowe (PNN) [8].

Jednym z głównych problemów rozwiązywanych za pomocą sieci neuronowych jest klasyfikacja i rozpoznawanie obrazów i dźwięków, wykorzystywane również w medycynie sądowej. Sieci neuronowe (m.in. RBF) mogą być również z powodzeniem stosowane do klasyfikacji chromosomów pod warunkiem istnienia olbrzymiego zbioru danych, pozwalających na uczenie i testowanie sieci [6].

W ostatnich latach wśród metod sztucznej inteligencji coraz większą popularnością cieszą się **sztuczne systemy immunologiczne**. Natura układu immunologicznego (m.in. selekcja klonalna, zdolność uczenia, pamięć immunologiczna, samoorganizacja, odporność) czyni sztuczne systemy immunologiczne przydatnymi do rozwiązywania problemów klasyfikacyjnych, w szczególności problemów związanych z rozpoznawaniem znaków i analizy DNA. Obiektami systemu immunologicznego są limfocyty typu T – wspomagające i wykorzystywane w detekcji anomalii oraz typu B – przeciwciała (rys. 5) wykorzystywane w systemach analizy danych, zdolne do rozpoznawania antygenów i reagowania na nie.

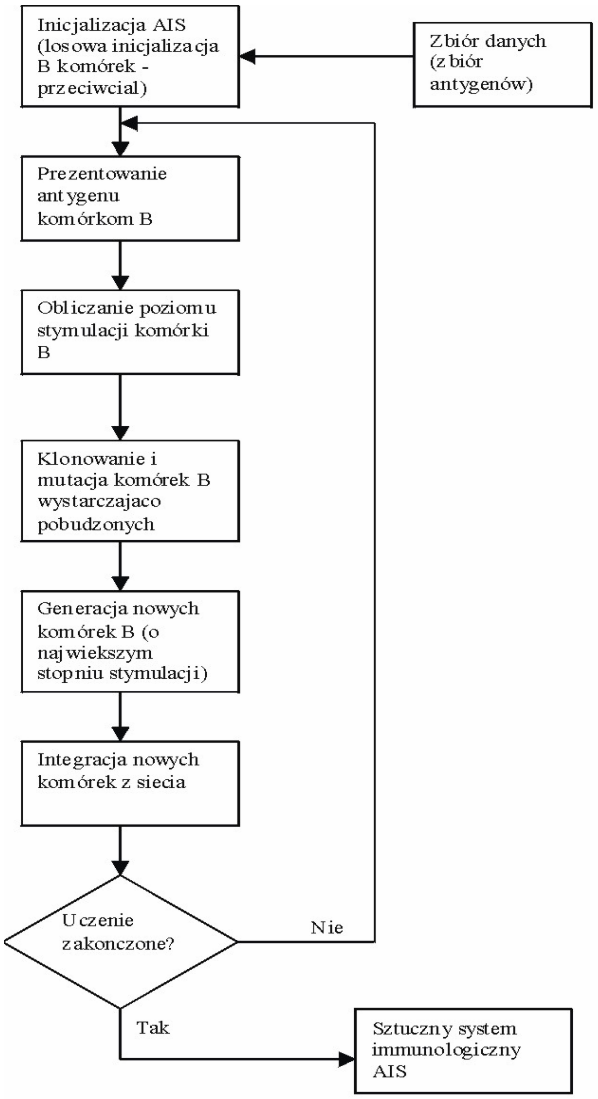


Rys. 5. Schemat limfocytu typu B

Limfocyty, reagując na określony antygen wychwytują istotne właściwości antygenu. Pobudzone limfocyty typu B są intensywnie klonowane w celu wychwycenia wszystkich antygenów (selekcja klonalna). Klony podlegają mutacji, dzięki której możliwe jest również wychwycenie antygenów podobnych. W sztucznych systemach immunologicznych antygeny tworzą zbiór danych klasyfikowanych, zaś przeciwciała tworzą wzorce tych danych. Przeciwciała tworzące sieć idiotypową modyfikowaną wraz z rozpoznawaniem kolejnych antygenów pogrupowane są na zasadzie podobieństwa. Jeżeli żadne z przeciwciał nie jest w stanie rozpoznać antygenu, produkowana jest nowa komórka na wzór antygenu. Każde nowo utworzone przeciwciało wprowadzane jest w sąsiedztwo przeciwciał do niego podobnych. Algorytm uczenia sieci idiotypowej przedstawiony jest na rys. 6. W procesie klasyfikowania określonej cechy (antygenu) ze zbioru przeciwciał wybieranych jest *n* elementów wraz z ich sąsiedztwem. Antygen porównywany jest z wybranymi przeciwciałami i obliczany jest stopień ich dopasowania oraz poziom stymulacji komórki. Przeciwciało, dla którego stopień dopasowania do antygenu przekracza próg stymulacji, nazywane jest węzłem klasyfikującym. Dla każdego węzła klasyfikującego badane jest również jego sąsiedztwo. Spośród wszystkich węzłów uporządkowanych malejąco względem dopasowania wybieramy przeciwciało o największym dopasowaniu, które stanowi rozwiązanie naszego problemu [9].

W większości przypadków stosowane jest kodowanie binarne przeciwciał i antygenów. Dla sekwencjonowania DNA stosujemy kodowanie ternarne, ponieważ łańcuch DNA zbudowany jest z 4 nukleotydów: A, T, G i C.

Istotną cechą odróżniającą sztuczne systemy immunologiczne od algorytmów genetycznych jest tzw. metadynamika (dzięki zapamiętywaniu przez sieć wyuczonych wzorców istnieje możliwość douczania sieci poprzez prezentowanie nowych danych). Poza tym o zmianach „gatunkowej różnorodności" w systemach immunologicznych decyduje selekcja klonalna i mutacja, natomiast w AG selekcja, krzyżowania i mutacja.



Rys. 6. Algorytm uczenia sieci idiotypowej

Dzięki opracowanej mapie chromosomu Y, zastosowaniu analizy pojedynczych polimorfizmów nukleotydowych (ang. *single nucleotide polymorphism – SNP*) oraz wykorzystaniu krótkich odcinków powtarzających (ang. *short tandem repeat – STR*) nastąpił znaczny rozwój w medycynie sądowej w następujących dziedzinach [1], [7], [13], [14]:

- badanie spornego ojcostwa,
- identyfikacja sprawców gwałtu,
- identyfikacja nieznanych zwłok oraz zaginionych osób.

Dzięki zastosowaniu techniki multiplex PCR – reakcji łańcuchowej polimerazy z wieloma starterami można uzyskać analizę 246 Y-SNP [12].

Aktualnie bardzo ważne jest tworzenie bazy danych DNA oraz rozwój wymiany informacji pomiędzy laboratoriami hemogenetyki [5], [9], [10], [12].

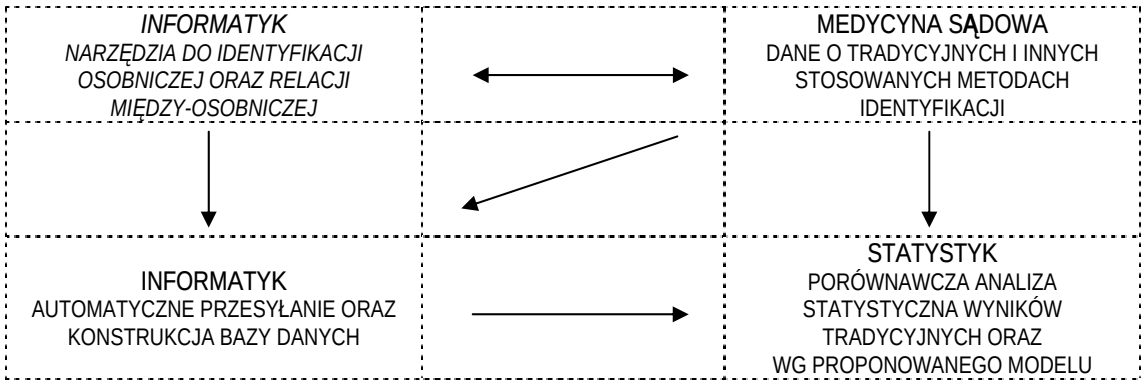
2. Podsumowanie

Nasz artykuł przedstawia kilka przykładów zastosowania systemów bioinformatycznych w medycynie sądowej. Urządzenia wykorzystujące systemy biometryczne są użyteczne w identyfikacji linii papilarnych, tęczówki oka czy też geometrii twarzy.

Tworzenie modeli stochastycznych jest elastyczną i najbardziej ogólną metodą wykorzystywaną w zagadnieniach systemów biometrycznych. Istnieje wiele nie w pełni jasnych i dających się z trudem zdefiniować aspektów związanych z rozpoznaniem danej cechy. Ponadto znajomość oraz zastosowanie algorytmów genetycznych wydaje się niezbędne w hemogenetyce, między innymi w identyfikacji DNA nieznanych zwłok oraz w oznaczaniu spornego ojcostwa.

Literatura cytowana

1. Butler J. M.: Recent developments in Y-single tandem repaet and Y-single nucleotide polymorphism analysis, Forensic Sci Rev, 15: 91-108, 2003.
2. Filipowicz B.: Modele stochastyczne w badaniach operacyjnych. Analiza i synteza systemów obsługi i sieci kolejkowych, WNT, Warszawa 1996.
3. Filipowicz B., Kwiecień J.: Optymalizacja sieci kolejkowych przy użyciu algorytmów genetycznych, AUTOMATYKA, Półrocznik, Uczelniane Wydawnictwa Naukowo-Dydaktyczne AGH, Kraków 2003.
4. Filipowicz B.: Modelowanie i optymalizacja systemów kolejkowych cz. 1. Systemy Markowskie, Przedsiębiorstwo Poligraficzne T. Rudkowski, Kraków 1995.
5. Goldberg D.: Algorytmy genetyczne i ich zastosowanie, WNT, Warszawa 1998.
6. Musavi M. T. i in.: Mouse chromosome classification by radial basis function network with fast orthogonal search, Neural Networks 11, 769-777, 1998.
7. Opolska-Bogusz B., Sanak M., Turowska M.: Genetic variation at STR-TH01 locus in the South Polish population, Arch Med Sadowej Krymino, 52(2): 99-101, 2002.
8. Tadeusiewicz R.: Sieci neuronowe, Akademicka Oficyna Wydawnicza, Warszawa 1993.
9. Wierzchoń S. T.: Sztuczne systemy immunologiczne. Teoria i zastosowania, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2001.
10. www.biometryka.com
11. www.logistyka.net.pl
12. http://ycc.biosci.arizona.edu
13. www.relive.genetics.com
14. http://genome.ucsc.edu



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 9-12.

PORÓWNANIE METODY KNN I TECHNIKI SIECI NEURONOWYCH W KLASYFIKACJI BADAŃ SPIROMETRYCZNYCH

A COMPARISON OF THE KNN METHOD AND NEURAL NETWORKS IN SPIROMETRIC TESTS CLASSIFICATION

JOANNA GRABSKA-CHRZĄSTOWSKA[*], WIESŁAW LIBUSZOWSKI[**], WALDEMAR TOMALAK[***]

[*]*Katedra Automatyki Akademii Górniczo-Hutniczej, Kraków, asior@agh.edu.pl*
[**] *Wydział Informatyki KWP, Rzeszów, wlibuszowski@plusnet.pl*
[***]*Instytut Gruźlicy i Chorób Płuc, Oddział w Rabce Zdroju, wtomalak@zpigichp.edu.pl*

Streszczenie. Praca przedstawia porównanie sieci neuronowych z metodą *k* najbliższych sąsiadów, zastosowanych do klasyfikacji wyników badania spirometrycznego. Sieci neuronowe okazały się lepszym klasyfikatorem niż metoda kNN.

Słowa kluczowe: metoda kNN, sieci neuronowe, klasyfikacja, badanie spirometryczne

Abstract. This paper presents a comparison of neural networks and the kNN method in spirometric tests classification. Neural networks proved to be a better classifier than the kNN method.

Key words: kNN method, neural networks, classification, spirometric test

1. Wstęp

Badanie spirometryczne polegające na pomiarze objętości i przepływów w układzie oddechowym jest najczęściej wykonywanym badaniem przy podejrzeniu jakichkolwiek zmian w układzie oddechowym spowodowanych przez proces chorobowy. Interpretacja wyniku badania polega najczęściej na porównaniu wyników pomiarów z wartościami należnymi (normami) i ewentualnym rozpoznaniu zaburzenia. Rejestrację krzywej przepływ-objętość i wartości z nią związanych można przeprowadzić w warunkach ambulatoryjnych. Automatyczne wspomaganie diagnozy na podstawie kilku charakterystycznych wartości umożliwiłoby wstępną kwalifikację pacjentów do bardziej szczegółowych badań. Celem tej pracy jest próba porównania popularnego klasyfikatora kNN i sieci neuronowych, zastosowanych do rozdzielenia wyników badań spirometrycznych pozostających w normie od wskazujących na zaburzenie. W pracy [2] opisano sieć neuronową użytą do wstępnej klasyfikacji wyników badania spirometrycznego i stwierdzono, że ze wszystkich parametrów antropometrycznych największy wpływ na poprawną decyzję ma parametr wzrostu i ta konkluzja została wykorzystana w opracowaniu następnej metody automatycznej diagnozy. Tak więc w pracy [3] przedstawiono metodę trzech sieci zestawionych w system wnioskujący. Testy przeprowadzono w oparciu o 1022 wyniki badań spirometrycznych (samych mężczyzn), które zostały wykonane w Zakładzie Fizjopatologii Układu Oddychania Instytutu Gruźlicy i Chorób Płuc w Rabce. Wszystkie doświadczenia z sieciami neuronowymi przeprowadzono przy pomocy pakietu Statistica firmy StatSoft. Program do klasyfikacji metodą kNN jest autorstwa dr A. Jóźwika.

Wyniki przedstawionych badań wskazują na większą przydatność sieci neuronowych do wstępnej klasyfikacji spirogramów.

2. Badanie spirometryczne

Do wykonanania badania spirometrycznego stosuje się najczęściej urządzenia mierzące przepływ powietrza w układzie oddechowym i umożliwiające określanie objętości lub jej zmiany na zasadzie integracji przepływu [4], [8], [9]. Jedną z procedur pomiarowych jest rejestracja manewru natężonego wydechu, który umożliwiaja rejestrację krzywej maksymalny przepływ-objętość oraz wartości z nią związanych.

Zaburzenia wartości charakterystycznych (zwłaszcza FEV_1 i FVC) mają wpływ na klasyfikację wyników badania i stanowią podstawę do formułowania wniosków i ustalania diagnozy.

Wartości uzyskane na podstawie badania i użyte jako wejścia do klasyfikatora to:

- FEV_1: objętość powietrza wydmuchnięta w czasie pierwszej sekundy natężonego wydechu,
- FVC: największa objętość powietrza wydmuchnięta przy maksymalnym wysiłku wydechowym po uprzednim największym możliwe wydechu,
- PEF: szczytowy przepływ wydechowy zarejestrowany w trakcie badania maksymalnie natężonego wdechu,

- MEFx%FVC: maksymalny przepływ powietrza, gdy x% FVC pozostało do wydmuchnięcia. Przepływ zmierzony podczas maksymalnie natężonego wydechu, w punkcie odpowiadającym x% pozostałej do wydmuchnięcia natężonej pojemności życiowej płuc (FVC).

3. Sieci neuronowe

Sieci neuronowe typu feedforward, uczone przy pomocy algorytmu ze wsteczną propagacją błędów, są chętnie stosowane do zadań klasyfikacji [7]. Aby znaleźć właściwą strukturę sieci trzeba określić liczbę jej wejść i wyjść. W efekcie badania spirometrycznego uzyskujemy 6 wartości charakteryzujących przepływ w drogach oddechowych. Dodatkowo pacjenta mogą charakteryzować dane antropometryczne: wzrost, waga i wiek. Dotychczasowe eksperymenty [1] wskazały na parametr wzrostu jako najważniejszy z wszystkich cech. Sieć neuronowa z 6 wejściami będącymi rezultatem badania spirometrycznego i dodanym parametrem wzrostu okazała się wstępnie najlepszym klasyfikatorem dla tego typu badań. W tab. 1 przedstawiono sposób prezentacji wyników, natomiast w tab. 2 rezultaty otrzymane dla najlepszej 7 – wejściowej sieci z jednym wyjściem (0 – norma, 1 – zaburzenie). Rozpatrywano jedynie wyniki mężczyzn, aby nie wprowadzać dodatkowych komplikacji związanych z płcią badanych. Wykorzystano 1022 wektory, w tym: 627 wskazujących na normę, 238 na chorobę obturacyjną, 76 na zaburzenie restrykcyjne i 81 na zaburzenie mieszane, czyli 395 wektorów uznanych za znajdujące się poza normą. W zbiorze testowym znajdowało się 511 losowo wybranych wektorów.

Tab. 1. Sposób przedstawiania wyników klasyfikacji

	SIEĆ	
EKSPERT	norma	zaburzenie
norma	a	c
zaburzenie	b	d

Uwagi: *a* – liczba przypadków klasyfikowanych w normie przez eksperta i przez sieć – przypadki prawdziwie ujemne,
b – liczba przypadków uznanych przez eksperta za zaburzenie, a przez sieć za normę – przypadki fałszywie ujemne,
c – liczba przypadków klasyfikowanych przez eksperta jako normę, a przez sieć jako zaburzenie – przypadki fałszywie dodatnie,
d – liczba przypadków kwalifikowanych przez eksperta i sieć jako zaburzenie – przypadki prawdziwie dodatnie.

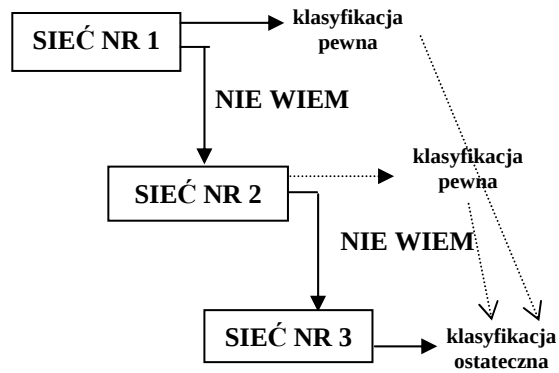
Tab. 2. Najlepsza pojedyncza sieć z dodanym parametrem wzrostu – próg 0,5

	SIEĆ	
EKSPERT	norma	zaburzenie
norma	300	20
zaburzenie	27	164

Przy interpretacji odpowiedzi sieci z założenia binarnej, a w praktyce ciągłej w przedziale <0,1>, stosowano zasadę przypisywania odpowiedzi zero (norma), gdy wartość wyjścia wynosiła mniej niż 0,5, i wartości 1 (zaburzenie), gdy rezultat przekraczał 0,5. Można to nazwać siecią o progu 0,5. Okazało się, że można sterować wartością przypadków fałszywie ujemnych (b) i fałszywie dodatnich (c), zmieniając ten wcześniej arbitralnie przyjęty próg binaryzacji.

Najważniejszym celem, oprócz otrzymania najlepszej jakości rozpoznania, jest otrzymanie jak najmniejszej liczby wyników fałszywie dodatnich (b). Takie przypadki nie mają szans na powtórną weryfikację, a nieleczenie osoby, u której występują zaburzenia oddechowe może prowadzić do pogorszenia stanu zdrowia pacjenta, przedłużającego się leczenia lub pogorszenia rokowania. Zdrowy pacjent uznany za chorego zostanie poddany dalszym badaniom i w końcu zostanie poprawnie zdiagnozowany lub – w najgorszym razie – poddany niepotrzebnemu leczeniu. Jednak zbyt duża liczba zdrowych pacjentów, kierowanych bez potrzeby do specjalisty dyskredytuje daną metodę. Wiarygodnymi wskaźnikami jakości testu oprócz procentowej liczby prawidłowych rozpoznań są więc czułość i swoistość testu. Poniżej zostanie przedstawione ostateczne porównanie wszystkich metod.

Dla pojedynczej sieci nie można jednocześnie zmniejszyć progu binaryzacji tak, aby otrzymać jak najmniej przypadków w polu b i zwiększyć progu, aby otrzymać mniejszą liczbę w polu c. Stąd pomysł systemu sieci z różnymi progami binaryzacji. Ideę systemu wnioskującego przedstawia rys. 1.



Rys. 1. Propozycja systemu wnioskującego

Do klasyfikacji pewnej dopuszcza się pewien błąd, którego przy takim podejściu nie da się uniknąć. Dobór progu przy interpretacji odpowiedzi sieci 1 i 2 ma w jak największym stopniu zmniejszyć ten wstępny błąd. Przy ostatecznej klasyfikacji (sieć nr 3) należy tak ustawić próg binaryzacji, aby możliwie najmniej przypadków pozostało w polu b przy rozsądnej liczbie przypadków w polu c.

Jako pierwszy element systemu wykorzystano sieć uczoną tylko 6 elementami badania spirometrycznego, oddzielającą normę od zaburzenia. Ustalono próg na 0,9 (duża wartość progu zmniejsza liczbę przypadków w polu c). Na wejście sieci nr 2 podaje się wektory z pól a i b, czyli uznane przez sieć nr 1 za będące w normie. W tym miejscu wprowadza się już wstępny błąd (około 2%) związany z wektorami, które zostały błędnie zaklasyfikowane i już nie zostaną zweryfikowane. Ta część potencjalnie zdrowych pacjentów zostanie wysłana do dalszych badań.

Na podstawie analizy błędnych rozpoznań ustalono, że sieć nr 2 oddzieli normę od obturacji, a sieć nr 3 umożliwi separację normy od restrykcji. Na wejściu sieci 2 i 3 dodano parametr wzrostu, który ułatwia uczenie sieci. W przypadku pokazanego systemu wnioskującego nie można stwierdzić, które wektory są uczące, a które testowe, gdyż dla kolejnych sieci są to inne wektory. Wyniki przedstawione w tabeli 3 dotyczą więc całego zbioru danych. Nie jest to nadużycie,

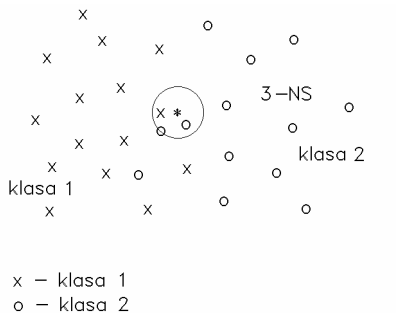
gdyż wartości błędów wszystkich sieci dla zbiorów uczących i testowych są podobne.

Tab. 3. Wyniki systemu wnioskującego

	SIEĆ	
EKSPERT	norma	zaburzenie
norma	574	53
zaburzenie	47	348

3. Metoda kNN

Reguła kNN (*k Nearest Neighbours*) została zaproponowana w pracy [1] i jest jednym z najlepszych klasycznych klasyfikatorów. Pewne jej modyfikacje [5], [6] mogą poprawić jakość klasyfikacji, ale ma to miejsce głównie w przypadkach więcej niż dwóch klas.



Rys. 1. Ilustracja reguły *k*-NS dla *k*=3. Źródło: [5]

Metoda polega na tym, że klasyfikowany obiekt jest zaliczany do klasy najliczniej reprezentowanej wśród jego *k* „najbliższych sąsiadów". Jeżeli w tej samej odległości, co *k*-ty „sąsiad" znajdą się jeszcze inne elementy, to wszyscy ci „sąsiedzi" biorą udział w głosowaniu. Działanie tej reguły dla *k*=3 i sztucznego małego dwuwymiarowego zbioru odniesienia zilustrowano na rys. 1.

Tab. 4. Wyniki metody 1NN uwzględniającej wszystkie 7 wejść

	SIEĆ	
EKSPERT	norma	zaburzenie
norma	276	38
zaburzenie	51	146

W przypadku wariantu metody najbliższego sąsiada uwzględniającej wszystkie 7 wejść otrzymano rezultaty, które ilustruje tab. 4. Najlepsze wyniki uzyskano jednak dla innego wariantu, w którym pozwalamy na redukcję cech wejściowych. Rezultaty dla *k*=1 i czterech wybranych cech przedstawiono w tab. 5. Wszystkie wyniki dotyczą zbioru testowego zawierającego 511 losowo wybranych przypadków z obu klas.

Tab. 5. Wyniki metody 1NN dla 4 wybranych cech

	SIEĆ	
EKSPERT	norma	zaburzenie
norma	283	31
zaburzenie	50	147

Inne wartości parametru *k* polepszały wyniki fałszywie dodatnie w pozycji c, pogarszając rezultaty fałszywie ujemne w pozycji b (tab. 1). Przykładowe wyniki pokazano w tab. 6.

Tab. 6. Wyniki metody 23NN uwzględniającej wszystkie 7 wejść

	SIEĆ	
EKSPERT	norma	zaburzenie
norma	304	10
zaburzenie	68	129

Takie wartości znacznie poprawiają swoistość testu, ale pogarszają czułość metody, co jest nie do przyjęcia. Wyniki dla *k*>1 przypominają rezultaty sieci neuronowej, gdy przy interpretacji wartości na wyjściu sieci przyjmiemy wartość progu mniejszą niż 0,5.

4. Porównanie wyników

Aby porównać wyniki uzyskane przez najlepszą pojedynczą sieć neuronową, neuronowy system wnioskujący i klasyfikator kNN, w tab. 7 przedstawiono trzy najważniejsze wskaźniki dobrego klasyfikatora: jakość rozpoznania, czułość testu i jego swoistość.

Czułość testu jest pojęciem odnoszącym się do zdolności wykrywania rzeczywiście chorych pacjentów. Natomiast swoistość testu pozwala określić, w jakim stopniu klasyfikator potrafi wskazać rzeczywiście zdrowe osoby.

Poszukiwaliśmy sposobu na sklasyfikowanie wyników badań spirometrycznych z największą czułością przy możliwie jak najwyższej liczbie prawidłowych rozpoznań i z możliwą do przyjęcia swoistością.

Tab. 5. Zestawienie wyników

	Jakość rozpoznania [%]	Czułość [%]	Swoistość [%]
pojedyncza sieć neuronowa (7 wejść)	90,8	85,9	93,8
neuronowy system wnioskujący	90,2	88,1	92,6
metoda 1NN wszystkie cechy	82,6	74,1	87,9
metoda 23NN wszystkie cechy	84,7	65,5	96,8
metoda 1NN z redukcją cech	84,2	74,6	90,1

System wnioskujący jest najlepszym klasyfikatorem, jeśli chodzi o czułość testu ustępuje nieco pojedynczej sieci w jakości rozpoznawania. Metoda kNN ustępuje każdej z pokazanych metod neuronowych.

5. Podsumowanie

Z przedstawionych rozważań i porównania wyników eksperymentów z sieciami neuronowymi i klasyfikatorem kNN można wyciągnąć wniosek o większej przydatności narzędzia neuronowego, przynajmniej przy próbach automatycznej klasyfikacji spirogramów. Jest to bardzo istotny wniosek dla ukierunkowania dalszych badań nad tym problemem. W przyszłości dysponując jeszcze większą liczbą danych będzie możliwe powtórne zweryfikowanie obu podejść.

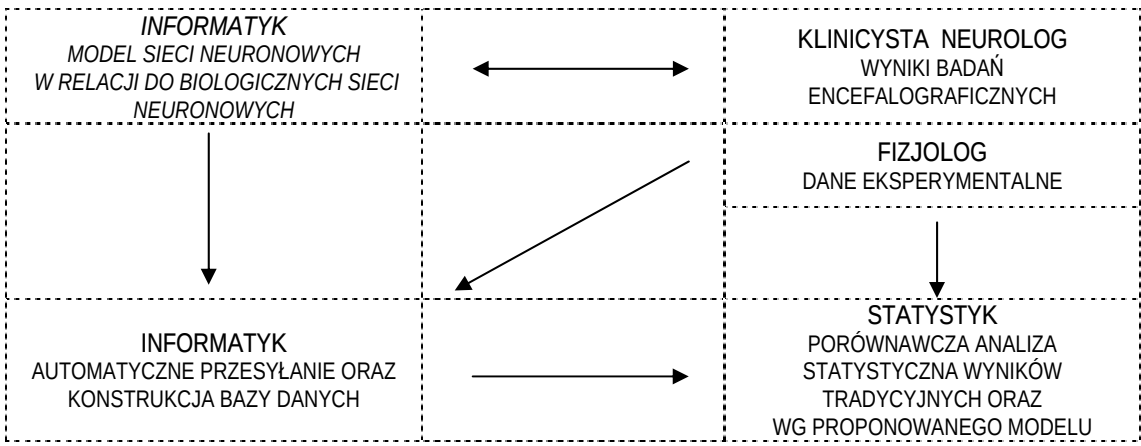
Podziękowania

Dziękujemy dr A. Jóźwikowi za cenne wskazówki i udostępnienie programu z zaimplementowaną metodą kNN.

Praca finansowana z funduszu badań statutowych (umowa AGH nr 11.11.120.24).

Literatura cytowana

1. Fix E., Hodges J. L.: Discriminatory analysis – nonparametric discrimination: consistency properties, Project 21-49-004, Report no. 4, USAF School of Aviation Medicine, pp. 261-279, Randolph Field, Tex. 1951.
2. Grabska-Chrząstowska J., Libuszowski W., Tomalak W.: Analiza wpływu parametrów wejściowych na klasyfikację wyników badania spirometrycznego przy pomocy sieci neuronowych, Automatyka, Półrocznik, tom 8, zeszyt 3, str. 469-475, Zeszyty Naukowe AGH, ISNN 1429-3447, Kraków 2004.
3. Grabska-Chrząstowska J., Tomalak W.: Zastosowanie neuronowego systemu wnioskującego do klasyfikacji wyników badania spirometrycznego, Materiały konferencyjne SIIB, Kraków 2004.
4. Hałuszka J., Willim G., Tomalak W., Kapustianyk I.: Komputerowo wspomagana ocena stanu czynnościowego płuc, Instytut Matki i Dziecka OT, Rabka 1990.
5. Jóźwik A., Serpico S., Roli F.: A parallel network of modified 1-NN and k-NN classifiers – application to remote-sensing image classification, Pattern Recognition Letters 19, pp. 57-62, 1998.
6. Jóźwik A., Janecki J., Demczuk M.: A new type of two stage NN classifier and NN classifiers nets. Application to cardio-circulatory events prediction, Proceedings of the 8th International IMEKO Conference on Measurement in Clinical Medicine, pp. 11-9 till 11-12, Dubrownik – Croatia 1998.
7. Tadeusiewicz R.: Sieci neuronowe, PWN, Warszawa 1993
8. Tomalak W.: Badania czynnościowe układu oddechowego, Materiały III Sympozjum Modelowanie i Pomiary w Medycynie, 37-49, Krynica 2001.
9. Willim G., Kurzawa R., Mazurek H., Hałuszka J., Jędrys-Kłucjasz U., Baran B., Radliński J.: Wartości należne wskaźników oddechowych dzieci i młodzieży, IGiChP, Rabka 1998.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 13-20.

PODOBIEŃSTWA SAMOOPTYMALIZUJĄCYCH SIECI NEURONOWYCH DO BIOLOGICZNYCH SIECI NEURONOWYCH

SIMILARITIES OF SELF-OPTIMIZING NEURAL NETWORKS TO BIOLOGICAL NEURAL NETWORKS

ADRIAN HORZYK

Katedra Automatyki Akademii Górniczo-Hutniczej, Kraków, horzyk@agh.edu.pl

Streszczenie. Pragnienie poznania procesów zachodzących w otaczającym nas świecie, ale również w naszym wewnętrznym świecie myśli, chęć ich kontroli, dążenie do zaspokojenia potrzeb, jak również możliwość uwolnienia się od pewnych przyzwyczajeń oraz myśli skłania naukowców do ciągłego pogłębiania tajemnicy naszego umysłu. Rozwijająca się technologia, sprzęt badawczy, algorytmy umożliwiają w różnoraki sposób odkrywać coraz to głębsze tajemnice naszego wewnętrznego świata. Poznanie prawdy o nas samych daje możliwość prostszego wyzbycia się dręczących ludzkość spraw oraz szybszą adaptację i naukę pożądanych umiejętności. W tym artykule opisano podobieństwa biologicznych sieci neuronowych oraz Samo Optymalizujących Sieci Neuronowych (SONN), których sposób tworzenia i własności klasyfikacyjne są częściowym odzwierciedleniem procesów biologicznych zachodzących w naturalnych sieciach neuronowych.

Słowa kluczowe: ontogeniczne sieci neuronowe, biologiczne sieci neuronowe, plastyczność, optymalizacja

Abstract. A natural human desire of perception of different processes in the surrounding world and our inner world of thoughts, a desire to control them, an endeavor to satisfy our needs, an ability to make our thoughts free from certain exterminated habits make scientists continue the plumetting of our brains. The newly developed technology, hardware, and algorithms make it possible to detect still deeper secrets of our inner world. Recognition of truth about us makes us simply free from tormenting question and enables faster adaptation and training of desirable skills. This paper describes similarities between natural neural networks (NNNs) and Self-Optimizing Neural Networks (SONNs). The paper describes development, adaptation, and classification properties of SONNs that are partially reflected in processes of the natural neural networks.

Key words: ontogenic neural networks, natural neural networks, plasticity, optimization

1. Wstęp

Systemy biologiczne i ich zdolności efektywnego przetwarzania informacji były i są natchnieniem dla wielu badaczy. Na bazie systemów biologicznych powstawały najróżniejsze modele teoretyczne i praktyczne, np. sztuczne sieci neuronowe, algorytmy genetyczne, metody ewolucyjne. Wszystkie te systemy ewoluowały i ewoluują do dnia dzisiejszego, próbując z jednej strony coraz bardziej zbliżać się do ich biologicznego pierwowzoru, z drugiej strony poprawiać ich zdolności adaptacyjne [1], [2], [3], [5], [6]. Ponadto dążenia wielu naukowców zmierzają do przyspieszenia procesów adaptacji biocybernetycznych systemów obliczeniowych wespół z poprawieniem i udowodnieniem poprawności ich działania również w przypadkach nieokreślonych. Poprawne działania biocybernetycznych systemów – tzw. zdolność do uogólnień – determinuje możliwość ich praktycznego wykorzystania do poważnych problemów z zakresu medycyny, przemysłu, ekonomii itp. [8], [11], [12], [13], [14], [15].

W tym artykule zaprezentowano i porównano jeden z nowych rodzajów ontogenicznych sieci neuronowych, tzw. Samooptymalizujące Sieci Neuronowe (SONN) [7], [8], pod kątem ich biologicznych podobieństw do naturalnych sieci neuronowych zarówno w sposobie ich adaptacji, jak również ich działania. Zaprezentowane ontogeniczne sieci neuronowe (mimo wykonywania wielu różnych obliczeń mających na celu zoptymalizowanie różnych parametrów sieci, tj. topologii sieci i parametrów wagowych), charakteryzują się dużą szybkością działania w fazie ich adaptacji. Ich naturalna (wbudowana w metodę) zdolność adaptacji do szerokiej klasy problemów klasyfikacji umożliwia ich szerokie stosowanie w praktyce. Na szczególną uwagę zasługuje prostota ich wykorzystania, która nie jest okupiona koniecznością stosowania przeróżnych współczynników, jak to się dzieje w wielu innych metodach, umożliwiając nawet mniej doświadczonym badaczom łatwe wykorzystywanie opisanych sieci. Sposób adaptacji sieci SONN znajduje ponadto odzierciedlenie w biologicznych procesach tworzenia naturalnych sieci

neuronowych. Celem adaptacji SONN jest zbudowanie sieci zoptymalizowanej pod kątem zadanego zbioru danych uczących. Optymalizacja sieci SONN dokonywana jest na dwóch podstawowych płaszczyznach: topologii sieci oraz parametrów sieci. Sieci SONN tworzą ciekawą niepełną wielowarstwową strukturę o połączeniach pomiędzy sąsiednimi warstwami, jak również pomiędzy bardziej odległymi warstwami, jak to się dzieje również w naturalnych sieciach neuronowych. W sieciach SONN występuje ponadto bardzo ściśle zdefiniowany moment zatrzymania procesu adaptacyjnego, który z założenia ma oznaczać zbudowanie sieci neuronowej będącej w stanie w 100% poprawnie i jednoznacznie sklasyfikować dane uczące zgodnie z informacjami zawartymi w zbiorze danych uczących. Sieci SONN posiadają również pewne unikalne własności, rzadko spotykane w innych metodach, np. umiejętność automatycznego rozpoznawania i poprawnej klasyfikacji negatywów (np. w przypadku obrazów), możliwość wykorzystania w fazie adaptacji danych niepełnych, czyli o niepełnych wektorach cech. Jedną z najistotniejszych cech sieci SONN jest ich przewidywalna zdolność do uogólniania zdobytej wiedzy o problemie. Sieci SONN są przykładem efektywnej i syntetycznej reprezentacji wiedzy o problemie, reprezentowanej w sieci neuronowej poprzez niewielką ilość połączeń.

W artykule poruszono również bardzo istotną kwestię wyboru odpowiedniego algorytmu adaptacyjnego w zależności od tego, czy dane są w pełni określone przed rozpoczęciem procesu adaptacji, czy też napływają stopniowo w czasie. Okazuje się bowiem, iż w przypadku sztucznych sieci neuronowych możliwe jest ominięcie pewnych ograniczeń biologicznych układów neuronowych i zastosowanie bardziej efektywnych metod adaptacyjnych w pewnych dobrze określonych przypadkach, stanowiących dużą grupę problemów, z jakimi spotykamy się na co dzień.

2. Opis metodologii budowy sieci SONN

Topologia ontogenicznych sieci SONN jest bardzo ściśle powiązana z konkretnym zbiorem uczącym, na podstawie którego powstaje. W związku z tym sposób budowy sieci SONN będzie wykorzystywał odpowiednio przetworzone dane z takiego zbioru.

Zbiór uczący $U = \left\{ \left(u^1, C^{m_1}\right), \ldots, \left(u^N, C^{m_N}\right) \right\}$ składa się z wielu wektorów wejściowych $u^n = \left[u_1^n, \ldots, u_K^n\right]$, których poszczególne cechy $u_k^n \in \{-1, 0, +1\}$ oraz klas $C^m \in \left\{ C^1, \ldots, C^M \right\}$ takich, że $u^n \in C^{m_n}$. Sposób adaptacji sieci SONN wykorzystuje technikę adaptacji nadzorowanej („z nauczycielem"), tzn. znane jest przyporządkowanie wzorców uczących do klas. Proces optymalizacji topologii sieci oraz jej parametrów (wag) oparty jest o zaproponowany współczynnik dyskryminacji [4], [7], obliczany dla każdej cechy wszystkich wzorców uczących:

$$\forall_{n\in\{1,\ldots,N\}} \forall_{k\in\{1,\ldots,K\}} \quad d_k^n = p_k^n \cdot q_k^n \cdot r_k^n = \begin{cases} \dfrac{\left(P_k^m\right)^2 \cdot N_k}{\left(P_k^m + N_k^m\right) P_k \cdot \left(P_k + N_k - 1\right)} & gdy\ u_k^n > 0\ \&\ u^n \in C^m \\ 0 & gdy\ u_k^n = 0 \\ \dfrac{\left(N_k^m\right)^2 \cdot P_k}{\left(P_k^m + N_k^m\right) N_k \cdot \left(P_k + N_k - 1\right)} & gdy\ u_k^n < 0\ \&\ u^n \in C^m \end{cases} \tag{1}$$

gdzie:

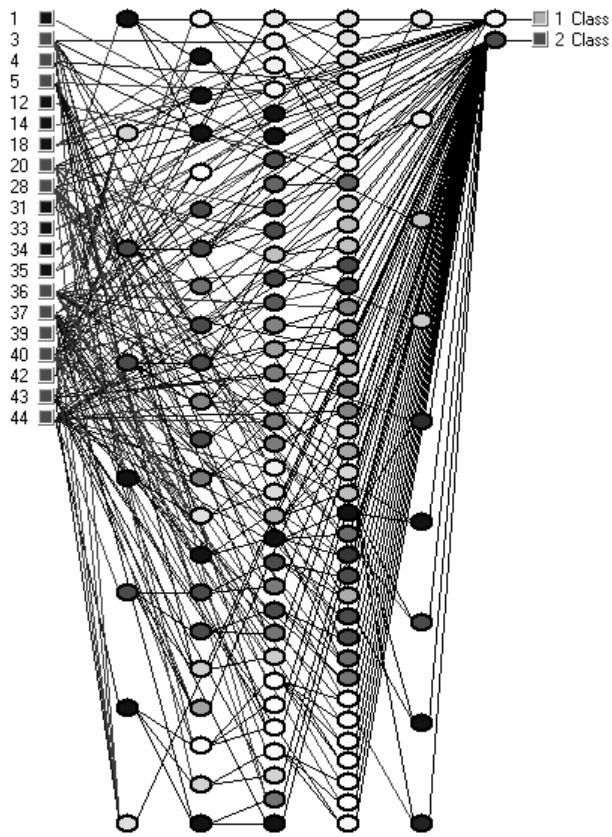
$$\forall_{m\in\{1,\ldots,M\}, k\in\{1,\ldots,K\}} P_k^m = \sum_{u_k^n \in \left\{u_k^n : n\in\{1,\ldots,N\} \& u^n \in U \cup C^m \& u_k^n > 0\right\}} u_k^n \tag{2}$$

$$\forall_{m\in\{1,\ldots,M\}, k\in\{1,\ldots,K\}} N_k^m = \sum_{u_k^n \in \left\{u_k^n : n\in\{1,\ldots,N\} \& u^n \in U \cup C^m \& u_k^n < 0\right\}} -u_k^n \tag{3}$$

$$\forall_{k\in\{1,\ldots,K\}} \quad P_k = \sum_{u_k^n \in \left\{u_k^n : n\in\{1,\ldots,N\} \& u^n \in U \& u_k^n > 0\right\}} u_k^n \tag{4}$$

$$\forall_{k\in\{1,\ldots,K\}} \quad N_k = \sum_{u_k^n \in \left\{u_k^n : n\in\{1,\ldots,N\} \& u^n \in U \& u_k^n < 0\right\}} -u_k^n \tag{5}$$

Współczynnik dyskryminacji $d_k^n \in [0;1]$ wyraża stopień istotności danej cechy *k* konkretnego wzorca *n* w procesie rozróżnienia wzorów klasy, do której należy, od pozostałych wzorców nie należących do tej klasy. Im większy współczynnik dyskryminacji, tym lepiej wyróżnia (dyskryminuje) on daną klasę wzorców.



Rys. 1. Przykład topologii klasycznej wersji sieci SONN. Z lewej strony wykorzystane cechy wejściowe (20 z 47), z prawej – dwie klasy określone w zbiorze uczącym

W klasycznej wersji sieci SONN (rys. 1) dla każdej klasy wzorców budowana jest oddzielna rozłączna podsieć, będąca „ekspertem" w jej rozpoznawaniu na podstawie wybranych najistotniejszych cech danej klasy [7]. W rozszerzonej wersji sieci SONN (rys. 2) możliwe jest zbudowanie jednej spójnej sieci dla wszystkich klas, co umożliwia lepsze zoptymalizowanie struktury sieci, wykorzystując również podobieństwa pomiędzy klasami. Rozszerzona wersja sieci SONN umożliwia

także automatyczne znalezienie minimalnej topologii sieci SONN (rys. 2) będącej w stanie poprawnie i jednoznacznie klasyfikować wszystkie wzorce uczące. Minimalna architektura sieci SONN może jednak wykazywać ograniczone zdolności do uogólniania, dlatego rozszerzona metoda SONN automatycznie zakłada wykorzystanie również innych mocno dyskryminujących cech przy budowie sieci – celem osiągnięcia pożądanych zdolności uogólniających. Takie postępowanie jest zbieżne ze sposobem działania biologicznych sieci neuronowych, w których występuje również pewna naturalna redundancja, która implikuje zdolności do uogólniania oraz uodparnia system neuronowy na ewentualne uszkodzenia, ułatwiając zarazem regenerację powypadkową. Należy podkreślić, że przy budowie sieci SONN redundancja nie jest tworzona w sposób sztuczny lub przypadkowy, lecz jest wynikiem obliczeń mających na celu zbudowanie struktury sieci tylko i wyłącznie w oparciu o wszystkie najistotniejsze, najlepiej rozróżniające cechy, które zarazem zapewniają własność poprawnego uogólniania przez sieci SONN.

Tab. 1. Sposób obliczania cechy podziału *DF* dla przykładowego zbioru danych (literek) z rys. 2

	WspółczynnikDyskryminacji * u_i * *10 (dla wszystkich wzorców uczących)*																*Ilość*		*Zgodne*		*Zaoszczędzone*	*Suma*
Cecha	*1*	*9*	*2*	*10*	*3*	*11*	*4*	*12*	*5*	*13*	*6*	*14*	*7*	*15*	*8*	*16*	*<0*	*>0*	*<0*	*>0*	*połączenia*	*W.D.*
1	-2,22	-2,22	0,80	0,80	-2,22	-2,22	0,80	0,80	0,80	0,80	0,80	0,80	-2,22	-2,22	0,80	0,80	6	10	6	7	93	2,13
2	-9,33	-9,33	0,19	0,19	0,19	0,19	0,19	0,19	0,19	0,19	0,19	0,19	0,19	0,19	0,19	0,19	2	14	22	4	74	2,13
7	0,44	0,44	0,44	0,44	-4,00	-4,00	0,44	0,44	0,44	0,44	0,44	0,44	-4,00	-4,00	0,44	0,44	4	12	24	1	83	2,13
DF = 8	-4,00	-4,00	0,44	0,44	0,44	0,44	0,44	0,44	0,44	0,44	0,44	0,44	0,44	0,44	-4,00	-4,00	4	12	12	7	113	2,13
9	9,33	9,33	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	14	2	5	25	90	2,13
10	-0,19	-0,19	-0,05	2,33	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,05	2,33	14	2	2	19	45	0,70
11	-0,26	0,43	0,43	-0,26	-1,04	-1,04	-1,04	-1,04	1,71	1,71	1,71	1,71	-1,04	-1,04	0,43	-0,26	9	7	6	5	78	1,51
12	5,00	-0,02	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	15	1	3	28	42	0,63
14	-2,22	-2,22	0,80	0,80	0,80	0,80	0,80	0,80	0,80	0,80	-2,22	-2,22	0,80	0,80	-2,22	-2,22	6	10	5	8	97	2,13
15	0,19	0,19	0,19	0,19	0,19	0,19	0,19	0,19	0,19	0,19	0,19	0,19	0,19	0,19	-9,33	-9,33	2	14	21	1	34	2,13
17	-0,19	-0,19	-0,05	2,33	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,05	2,33	14	2	2	19	45	0,70
18	-0,33	0,33	0,33	-0,33	-1,33	-1,33	-1,33	-1,33	1,33	1,33	1,33	1,33	-0,33	0,33	0,33	-0,33	8	8	7	3	70	1,33
19	5,00	-0,02	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	15	1	3	28	42	0,63
21	-2,22	-2,22	0,80	0,80	0,80	0,80	0,80	0,80	0,80	0,80	-2,22	-2,22	0,80	0,80	-2,22	-2,22	6	10	5	9	106	2,13
22	-2,93	-2,93	0,61	0,61	0,61	0,61	0,15	-0,73	0,61	0,61	0,61	0,61	0,61	0,61	-2,93	-2,93	5	11	6	8	104	1,87
23	5,78	5,78	-0,31	-0,31	-0,31	-0,31	-0,08	1,44	-0,31	-0,31	-0,31	-0,31	-0,31	-0,31	-0,31	-0,31	13	3	5	13	86	1,68
24	-0,19	-0,19	-0,05	2,33	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,05	2,33	14	2	2	19	45	0,70
25	-0,26	0,43	0,43	-0,26	-1,04	-1,04	-1,04	-1,04	0,43	-0,26	0,43	-0,26	1,71	1,71	0,43	-0,26	9	7	1	7	50	1,10
26	5,00	-0,02	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	-0,09	15	1	3	28	42	0,63
27	-0,19	-0,19	-0,19	-0,19	-0,19	-0,19	-0,05	2,33	-0,19	-0,19	-0,19	-0,19	-0,05	2,33	-0,19	-0,19	14	2	1	18	31	0,70
28	-1,71	-1,71	1,04	1,04	1,04	1,04	0,26	-0,43	1,04	1,04	-1,71	-1,71	1,04	1,04	-1,71	-1,71	7	9	2	12	108	1,93
29	-0,80	-0,80	-0,80	-0,80	-0,80	-0,80	-0,80	-0,80	2,22	2,22	2,22	2,22	-0,80	-0,80	2,22	2,22	10	6	2	10	68	2,13
30	-1,71	-1,71	1,04	1,04	1,04	1,04	0,26	-0,43	-1,71	-1,71	-1,71	-1,71	1,04	1,04	1,04	1,04	7	9	5	8	94	1,93
31	1,71	1,71	0,43	-0,26	-1,04	-1,04	1,71	1,71	-1,04	-1,04	-1,04	-1,04	-1,04	-1,04	1,71	1,71	9	7	9	3	90	1,93
32	1,04	1,04	-0,43	0,26	-1,71	-1,71	1,04	1,04	-1,71	-1,71	-1,71	-1,71	1,04	1,04	1,04	1,04	7	9	14	2	100	1,93
33	0,61	0,61	0,61	0,61	-0,73	0,15	0,61	0,61	-2,93	-2,93	-2,93	-2,93	0,61	0,61	0,61	0,61	5	11	15	1	70	1,87
34	0,61	0,61	0,61	0,61	0,61	0,61	0,15	-0,73	-2,93	-2,93	-2,93	-2,93	0,61	0,61	0,61	0,61	5	11	14	1	66	1,87
35	2,22	2,22	-0,80	-0,80	-0,80	-0,80	-0,80	-0,80	2,22	2,22	-0,80	-0,80	-0,80	-0,80	2,22	2,22	10	6	8	3	87	2,13
Klasa:	A	A	B	B	C	C	D	D	E	E	F	F	G	G	H	H						

Współczynnik *DF*=8 dzieli zbiór wzorców na dwa podzbiory *F*={1,8,9,16} oraz *T*={2,3,4,5,6,7,10,11,12,13,14,15}, pozwalając zaoszczędzić 113 połączeń sieci, reprezentując 132 zgodnych cech, które są zagregowane i reprezentowane (bez żadnej utraty informacji!) w sieci neuronowej przez zaledwie 12+7=19 połączeń i dwa neurony w pierwszej warstwie sieci SONN (rys. 2).

Optymalizacja topologii sieci SONN wykorzystuje tabele (tab. 1) służące do obliczania ilości zaoszczędzonych połączeń przy wyborze dowolnej dostępnej cechy z przestrzeni wejściowej zbioru wektorów uczących. Ilość zaoszczędzonych połączeń ZP_k dla cechy *k* obliczamy następująco:

$$ZP_k = ZgF_k \cdot (IlF_k - 1) + ZgT_k \cdot (IlT_k - 1) \qquad (6)$$

gdzie

ZgF_k – ilość zgodnych cech (co do wartości) dla wszystkich wzorców należących do grupy wzorców mających wartość ujemną (*F*) dla cechy *k*

ZgT_k – ilość zgodnych cech (co do wartości) dla wszystkich wzorców należących do grupy wzorców mających wartość dodatnią (*T*) dla cechy *k*

IlF_k – ilość wzorców posiadających ujemną cechę *k*

IlT_k – ilość wzorców posiadających dodatnią cechę *k*.

Następnie obliczana jest cecha podziału *DF* będąca cechą, która oszczędza maksymalną możliwą ilość połączeń reprezentujących cechy rozważanych wzorców uczących bez upraszczania informacji o nich w tworzonym modelu matematycznym:

$$DF = \max_{k=1,\dots,K} \{ZP_k\} \tag{7}$$

Obliczone cechy *DF* stanowią zarazem minimalną topologię SONN (rys. 2) będącą w stanie po skończonym procesie jej tworzenia poprawnie rozróżnić wszystkie wzorce uczące i przyporządkować je jednoznacznie do klas, które reprezentują. Ze względu na osiągnięcie pożądanej jakości uogólnienia metoda preferuje automatyczne oparcie topologii sieci również o inne cechy, które charakteryzują się dużym współczynnikiem dyskryminacji. Te dodatkowe cechy mogą być przy okazji podziału zbioru wzorców na dwa podzbiory (na podstawie cechy podziału *DF*) zastąpione również pojedynczymi połączeniami. Takie oszczędne redundantne zwiększenie topologii sieci umożliwia osiągnięcie bardzo dobrych wyników uogólniania i posiada także swoje odzwierciedlenie w biologicznych układach neuronowych. Naturalne sieci neuronowe również wykazują się pewną uzasadnioną redundancją pozwalającą na jakościowe uogólnianie, niewrażliwość na drobne uszkodzenia i regenerację powypadkową. Dla kontrastu – minimalna topologia jest wrażliwa na najdrobniejsze uszkodzenia destabilizujące jej poprawną pracę i posiada niewielkie zdolności do uogólniania. W przypadku sieci SONN, dzięki możliwości agregacji cech o dużym współczynniku dyskryminacji udało się otrzymać rozsądny poziom redundancji (rys. 2), zapewniający wysoką jakość uogólniania przy bardzo niskim koszcie dodatkowych „redundantnych" obliczeń.

Topologia sieci SONN składa się z wielu warstw neuronów, spośród których możemy wyróżnić:

- warstwę wyjściową – neuronów obliczających maksimum z wyjść wybranych neuronów warstw poprzednich i określających stopień podobieństwa wejścia do poszczególnych klas [7]:

$$y^m = f_{SGNMAX}(Y) = \begin{cases} y_{MIN} & gdy\ |y_{MAX}| < |y_{MIN}| \\ y_{MAX} & gdy\ |y_{MAX}| \geq |y_{MIN}| \end{cases} \tag{8}$$

gdzie: $Y = \{y^1,\dots,y^T\}$, y^t jest wyjściem pewnego neuronu z warstwy poprzedniej oraz

$$y_{MIN} = \min\{y_1,\dots,y_T\} \tag{9}$$

$$y_{MAX} = \max\{y_1,\dots,y_T\} \tag{10}$$

- warstw pośrednich – neuronów (ukrytych) odpowiedzialnych za stopniową dyskryminację wejścia z punktu widzenia poszczególnych klas. Neurony te grupują również cechy podobne dla wzorców uczących poszczególnych klas (w wersji podstawowej metody) lub wszystkich wzorców uczących (w wersji rozszerzonej metody) [7]:

$$y = f(u) = \sum_{j \in J} w_j \cdot u_j \tag{11}$$

Neurony warstw pośrednich posiadają rzadkie połączenia pomiędzy sąsiednimi oraz bardziej odległymi warstwami sieci. Wsteczna analiza aktywacji neuronów w sieci SONN pozwala na prześledzenie ścieżek aktywacji i określenie cech, które zadecydowały o klasyfikacji danego wejścia. Taka cecha jest czasami bardzo cenna, gdyż pozwala w przypadkach szczególnych na odpowiednie rozszerzenie lub ograniczenie zbioru badanych cech zbieranych w doświadczeniach empirycznych.

Parametry wagowe sieci SONN obliczane są również w specyficzny sposób, żeby odzwierciedlały przetworzoną już informację z warstw poprzednich sieci [7]. Wagi są dobierane adekwatnie do wartości współczynników dyskryminacji cech, jakie w sieci reprezentują. Przetworzone już wejścia w warstwach poprzednich sumują swoje współczynniki dyskryminacji d_0 i sumarycznie wpływają na przetwarzanie informacji w warstwach następnych. Obliczanie wag odbywa się równocześnie z konstrukcją struktury sieci następująco:

$$\forall_{i \in J}\ w_i = \frac{u_i^n \cdot \hat{d}_i}{\sum_{j \in J} \hat{d}_j} \tag{12}$$

gdzie

$$\hat{d}_i = \begin{cases} d_i & gdy\ cecha\ i\ uczestniczy\ w\ adaptacji \\ 0 & w\ odwrotnym\ przypadku \end{cases}$$

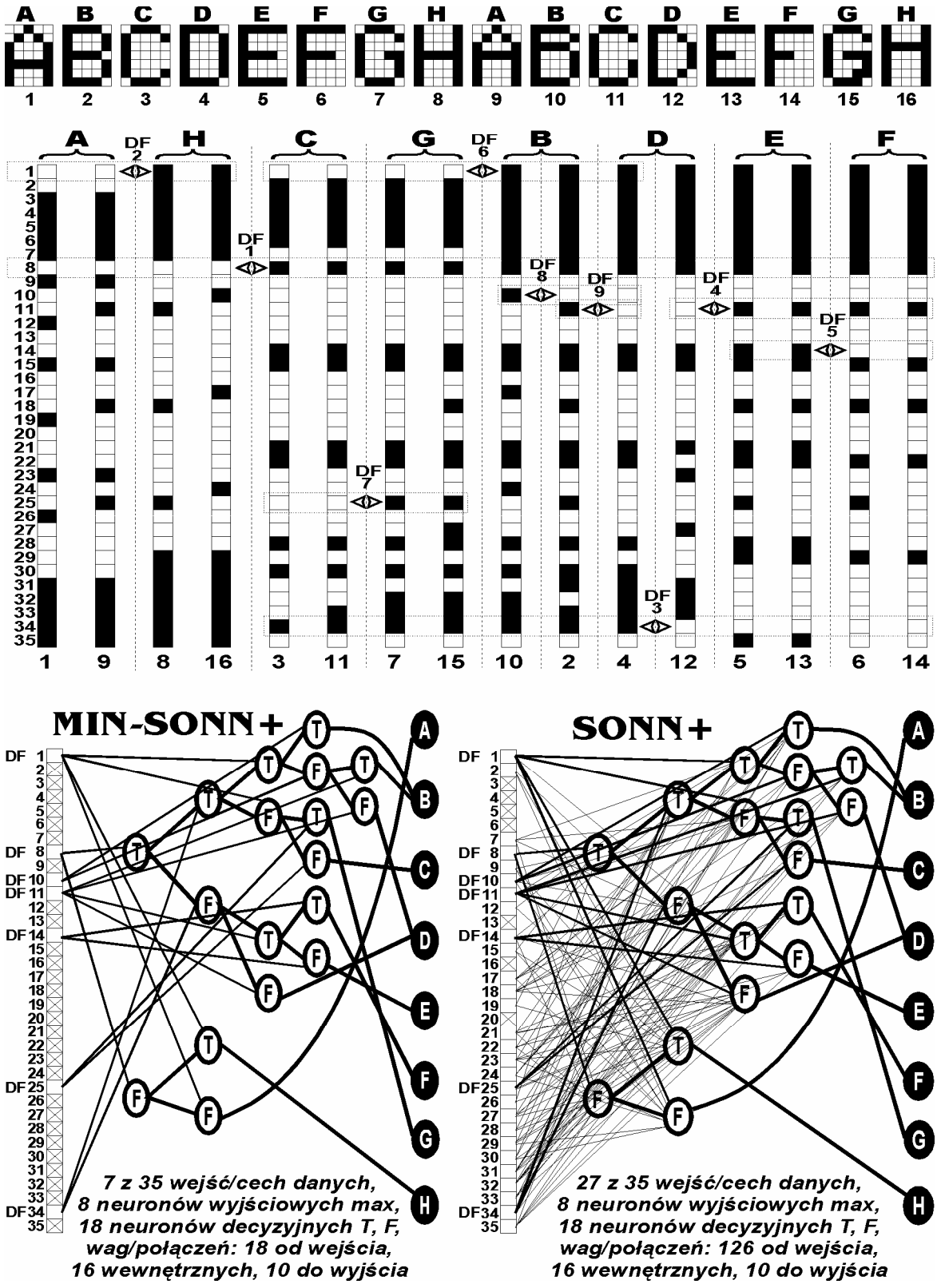
$J = A \cup \{0\}$ – zbiorem cech przekształconych na połączenia dla rozważanego neuronu,

$A \subset \{1,2,\dots,K\}$ – zbiorem zgodnych cech, nieprzekształconych w poprzednich warstwach na połączenia dla rozważanego neuronu.

$$\forall_{n \in \{1,\dots,N\}}\ u_0^n = \begin{cases} 0 & dla\ neuronów\ wejsciowych \\ 1 & dla\ neuronów\ ukrytych \end{cases}$$

$$d_0 = \begin{cases} 0 & dla\ neuronów\ wejsciowych \\ \sum_{l \in L} d_l & dla\ neuronów\ ukrytych \end{cases}$$

$L \subset \{1,2,\dots,K\} \cup \{0\}$ – zbiorem cech poprzednich neuronów, które zostały przekształcone na połączenia.



Rys. 2. Konstrukcja topologii rozszerzonej wersji sieci neuronowej SONN dla przykładowego zbioru danych (literek) z uwidocznieniem minimalnej topologii oraz normalnej topologii sieci. Cechy *DF* są to cechy podziału wybrane do skonstruowania wymaganej minimalnej struktury sieci. Wejścia sieci neuronowej (cechy) niewykorzystane (zredukowane) w procesie adaptacji sieci oznaczone są krzyżykiem

Sieci SONN wykorzystują fakt, iż dane uczące znane są w całości w momencie rozpoczęcia procesu adaptacji sieci. Dzięki temu mogą zebrać na podstawie danych uczących wszystkie potrzebne informacje o cechach najlepiej rozróżniających poszczególne klasy wzorców od siebie i podjąć globalną optymalizację architektury sieci, jak również obliczyć w sposób jednoznaczny optymalne parametry wagowe. Taka własność odróżnia możliwości sztucznych układów neuronowych od ich biologicznych odpowiedników, które działają w oparciu o dane napływające w czasie i stopniowo dostrajają do nich swoje parametry.

3. Opis podobieństw i różnic sieci SONN i biologicznych sieci neuronowych

Opisana metoda SONN umożliwia zbudowanie i zoptymalizowanie sieci neuronowej dla zadanego problemu klasyfikacji. Również biologiczne sieci neuronowe dostosowują swoją strukturę (z wykorzystaniem mechanizmów plastyczności synaptycznej) [10] pod kątem problemów, jakie będą musiały rozwiązać w trakcie swojego istnienia. Budowa i optymalizacja sieci SONN przebiega równocześnie.

Nietypowa wielowarstwowa topologia sieci SONN o częściowych połączeniach pomiędzy różnymi warstwami sieci (rys. 1-2) przypomina biologiczne sieci neuronowe, które również często są wielowarstwowe o różnorodnych niepełnych połączeniach pomiędzy różnymi warstwami lub pomiędzy neuronami nieuporządkowanymi w warstwy. W sieciach biologicznych występują jeszcze połączenia rekurencyjne, które nie występują w sieciach SONN.

Sieci SONN nie wymagają inicjacji ani topologii, ani parametrów sieci, podobnie jak ma to miejsce w biologicznych sieciach neuronowych, których przepuszczalności synaptyczne w postaci ilości i wrażliwości pomp jonowych, jak również kanałów dla różnego rodzaju neurotransmiterów różnicują się podczas życia [10].

Proces adaptacji sieci SONN rozpoczyna się od prostych struktur, które stopniowo rozrastają się do bardziej skomplikowanych w miarę potrzeb wynikających ze zbioru danych uczących. Także biologiczne układy neuronowe, począwszy od ich rozwoju prenatalnego rozwijają się z zarodka (neuronowych komórek macierzystych) [10] w kierunku sieci bardziej skomplikowanych, celem doskonalenia potrzeb przystosowawczych do otoczenia. Również w trakcie życia oraz uszkodzeń biologicznych sieci neuronowych odnotowujemy możliwości rekonfiguracji sieci neuronowej lub umiejętne rozszerzanie przetwarzania istotnych informacji do sąsiednich struktur, które są zagospodarowane przez inne zadania i czynności psychomotoryczne [10]. Ponadto plastyczność neuronowa umożliwia istniejącym neuronom rozszerzanie swojej siatki połączeń, ich degenerację – czyli ich rekonfigurację w zależności od zmieniających się potrzeb organizmu. W podobny sposób sieci SONN kreują najpierw podstawową strukturę w oparciu o najistotniejsze cechy danych, stopniowo zwiększając swoją architekturę umożliwiają coraz lepsze rozróżnianie nawet bardzo zbliżonych wzorców różnych klas aż do momentu ich poprawnego rozróżnienia w pełni, pod warunkiem, że dane uczące nie są wzajemnie sprzeczne.

Sposób adaptacji i zdobywania wiedzy przez sieci SONN jest bardzo zbliżony do sposobu jej zdobywania przez człowieka. Początkowe fazy rozwoju sieci SONN (a zarazem jej pierwsze warstwy) odwzorowują najistotniejsze i najlepiej rozróżniające cechy poszczególnych klas, zaś następne fazy jej rozwoju (odwzorowane w następnych jej warstwach) skupiają się na coraz to drobniejszych cechach wzorców o tyle, o ile są one konieczne do rozróżnienia wzorców różnych klas od siebie. Klasy dobrze rozróżnialne (separowalne) są reprezentowane w sieci SONN przez mniejszą ilość elementów (neuronów i połączeń), zaś klasy trudniejsze do odseparowania będą reprezentowane przez większą ich ilość. Naturalna niechęć do podejmowania zbędnego wysiłku (np. badania cech wyraźnie nieistotnych) przez człowieka podczas zdobywania wiedzy oraz twórczej pracy także w naturalny sposób prowadzi do wydobywania z danych tylko tych najistotniejszych cech, stwarzających wrażenie niezbędnych do poprawnej klasyfikacji. Jeżeli proces adaptacji sieci SONN zostanie przerwany w dowolnym momencie, skutek będzie łatwo przewidywalny i zbieżny z przerwaniem nauki przez człowieka (biologiczną sieć neuronową), tzn. część wzorców będzie poprawnie klasyfikowanych, a część niepoprawnie, również zdolności do uogólniania wiedzy też będą ograniczone i częściowo błędne. Biologiczne sieci neuronowe podczas uczenia zwracają najpierw uwagę na czynniki dobrze rozróżniające klasy, a gdy to jest niewystarczające, zwracana jest uwaga na szczegóły umożliwiające poprawną klasyfikację.

W procesie adaptacji sieci SONN wybierane są tylko najistotniejsze (tzn. nie wszystkie dostępne) spośród cech wzorców celem ich rozróżnienia. W podobny sposób postępują biologiczne sieci neuronowe, zwracając uwagę na to, co istotne, a nie na wszystkie cechy wzorców. W przypadkach mniej oczywistych, ilość wybranych cech do rozróżnienia klas może być częściowo redundantna, tj. brana jest pod uwagę większa ilość cech niż wskazywałoby na to minimum. Taki sposób postępowania zazwyczaj poprawia zdolności sieci do uogólnień i jest w sposób naturalny dostępny w sieci SONN, jak również w biologicznych sieciach neuronowych.

Proces adaptacji sieci SONN automatycznie zatrzymuje się w momencie osiągnięcia przez sieć jednoznacznego rozróżnienia pomiędzy klasami, nie dobudowując zbędnych elementów sieciowych i nie zwracając uwagi na nieistotne szczegóły danych. Taki sposób postępowania wpływa na poprawę zdolności do uogólniania, gdyż każda zbędna ilość szczegółów w sieci stwarza niebezpieczeństwo uczenia się sieci „na pamięć" zamiast modelowania wybranego zjawiska z możliwością rozszerzenia jej działania również na wzorce nowe. Jeżeli z jakichkolwiek powodów okazałoby się konieczne rozróżnianie wzorców poszczególnych klas przy pomocy większej ilości cech (szczegółów), możliwe jest kontynuowanie procesu adaptacji sieci SONN, który jednak rzadko prowadzi do poprawy zdolności uogólniających. Podobne zjawiska zachodzą również u człowieka. Czasami nadmiar wiedzy na określony temat utrudnia człowiekowi twórcze myślenie, gdyż zdobyta wiedza toruje w jego umyśle swoiste ślady, skłaniając jego uwagę w kierunku zdobytej już wiedzy. Dlatego też dzieci, które nie są przesiąknięte wiedzą, zadziwiają starszych ludzi swoją twórczą wyobraźnią i niesamowitą zdolnością do wymyślania rzeczy nowych.

Stosując sieci SONN mamy każdorazowo pewność, iż dane uczące zostaną w pełni sklasyfikowane poprawnie i jednoznacznie, za wyjątkiem sytuacji, w której dane są

sprzeczne, błędne lub niejednoznacznie określone w zbiorze uczącym.

Sieci SONN zdolne są również do patrzenia na dane przez pryzmat odwrotności. Ze względu na to, iż wejścia sieci SONN (zakres zmienności poszczególnych cech danych) znormalizowane są do zakresu od –1 do +1, a w fazie uczenia do zbioru {-1;0;1}, można mówić o odwrotności danych. W przypadku rozpoznawania np. obrazów często zdarza się, iż obraz podany jest w negatywie. W takim przypadku zarówno sieci biologiczne jak również SONN dostrzegą w sposób automatyczny ten fakt i będą zdolne do poprawnej klasyfikacji wzorca. Biologiczne neurony także w swoim działaniu wyróżniają dwa podstawowe stany: stan spoczynku (narastania potencjału czynnościowego) oraz stan aktywacji (potencjału czynnościowego), które – zależnie od rodzaju neurotransmitera – mogą oddziaływać na neuron postsynaptyczny pobudzająco (+1) lub hamująco (-1) [1], [3], [10].

W rzeczywistości często zdarza się, iż dane pochodzące z różnych doświadczeń empirycznych mogą posiadać pewne cechy nieokreślone. Sieci SONN, na wzór sieci biologicznych, są w stanie wykorzystać również takie nie w pełni określone wzorce w procesie adaptacji sieci. W takich przypadkach model matematyczny reprezentowany przez sieć SONN oparty jest tylko o znane cechy takich wzorców.

Dużą zaletą sieci SONN jest ich szybkość adaptacji do zadanego problemu. Ze względu na to, iż zbiór uczący dla sieci SONN jest w pełni określony w momencie rozpoczęcia procesu adaptacji, można przyjąć, iż taka optymalna sieć SONN istnieje i można ją wyznaczyć. Inaczej jest w przypadku biologicznych sieci neuronowych, które opierają swoje działanie na danych napływających w czasie i zmuszone są do ciągłej adaptacji swoich parametrów. Ze względu na ten fakt, sztuczne sieci neuronowe posiadają często pewną przewagę nad ich naturalnymi odpowiednikami, szczególnie gdy dane są w pełni określone przed rozpoczęciem procesu adaptacji. W takim przypadku sieć biologiczna (ze względu na sposób swojego działania i możliwości adaptacyjne) i tak zmuszona jest nauczyć się tych danych w taki sposób, jakby dane napływały stopniowo w czasie. Taki sposób nauki jest niekiedy dla człowieka niewygodny, przede wszystkim wtedy, gdy występuje potrzeba szybkiego zapamiętania danych i zbudowania na ich podstawie pewnego modelu, np. aproksymacyjnego. Sztuczne systemy neuronowe mogą zaś wykorzystać algorytmy szybkiej adaptacji w oparciu o informację o wszystkich wzorcach uczących dostępnych od razu tak, jak dzieje się to w sieciach SONN.

4. Podsumowanie

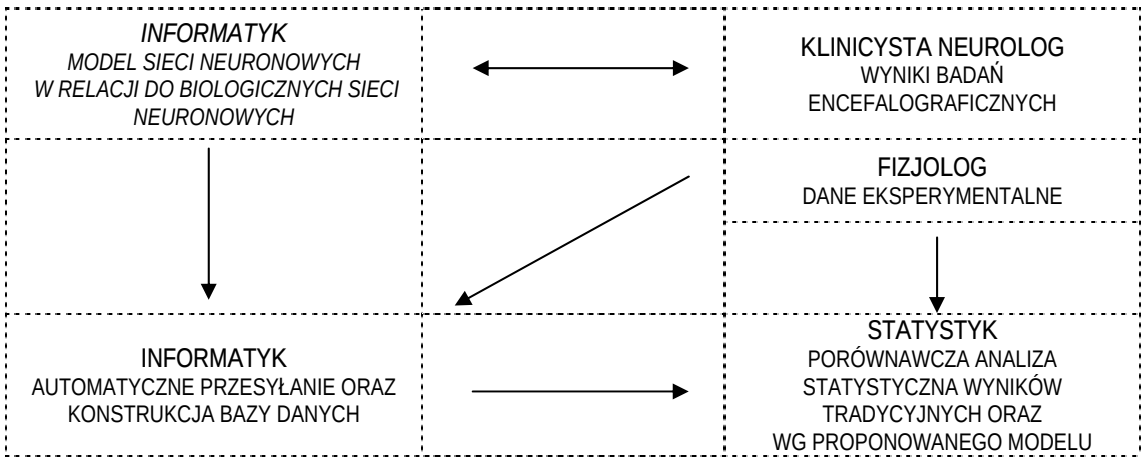
W artykule opisano sieci neuronowe SONN w wersji klasycznej, jak i rozszerzonej, z punktu widzenia ich biologicznego podobieństwa do naturalnych układów neuronowych. Opisane sieci posiadają wiele wspólnych cech topologicznych, adaptacyjnych i funkcjonalnych z biologicznymi układami nerwowymi. Wysunięto wniosek, iż możliwość szybkiej adaptacji do danych uczących występuje w przypadkach, gdy wszystkie dane uczące są od razu dostępne, tzn. nie jest wymagane oczekiwanie na dane uczące w czasie, jak ma to miejsce w przypadku biologicznych układów neuronowych. Taka właściwość umożliwia konstruowanie sztucznych sieci neuronowych, stosując bardziej efektywne mechanizmy adaptacji – jak to pokazano na przykładzie sieci SONN, które z założenia wykorzystują w procesie adaptacji informacje o dokładnie wszystkich danych uczących na raz. Efektywność szybkiej adaptacji sieci SONN do zbioru uczącego okupiona jest jednak brakiem możliwości douczenia takiej sieci nowych wzorców napływających w czasie. W takim przypadku występuje konieczność zebrania wszystkich wzorców razem i rozpoczęcia procesu adaptacji SONN od początku. Fakt ten można by było postrzegać jako wadę sieci SONN i ich algorytmu adaptacji, jednak większość przypadków, z jakimi spotykamy się w praktycznych zastosowaniach, to właśnie takie zadania, dla których zbiór uczący znany jest w całości na początku i potrzebna jest metoda efektywnej i poprawnej adaptacji. Bardzo duża szybkość adaptacji sieci SONN do dowolnego problemu nie nastręcza zaś dużych trudności nawet w przypadkach konieczności ponownej adaptacji sieci SONN po rozszerzeniu zbioru danych uczących. Dla kontrastu można by było traktować jako wadę biologicznych sieci neuronowych ich ograniczone możliwości adaptacji do danych w pełni określonych w danej chwili czasowej. Jednak większość problemów biologicznych układów neuronowych jest rozłożona w czasie, a sposób pozyskiwania danych – cząstkowy i wymagający czasu na ich zebranie. Taki sposób nauki polegający na stopniowym (w pewnym sensie ograniczonym co do szybkości) pozyskiwaniu wiedzy budującym doświadczenie osobnicze – stabilizuje zarazem wiedzę i tworzy charakter, np. człowieka. Umożliwienie biologicznym sieciom neuronowym szybkiego sposobu adaptacji, jaki występuje w sieciach SONN, destabilizowałby proces zdobywania wiedzy, ewolucji myśli i płynności życia organizmów żywych, gdyż pozwoliłoby im dokonywanie zbyt szybkich zmian pod wpływem nowych danych, bez zachowania jakiejkolwiek ich ciągłości i spójności. W przypadku sztucznych sieci neuronowych oczekujemy natomiast szybkiej adaptacji do zagadnień, których rozwiązanie jest nam potrzebne, dla których człowiek określił zbiór uczący jako względnie poprawny.

Podsumowując rozważania zawarte w tym artykule można wysunąć wniosek, iż obydwa sposoby adaptacji są potrzebne i zależne od zastosowań, a wybór optymalnego rozwiązania w przypadku sztucznych sieci neuronowych należy do człowieka. Dzięki temu, że kryterium wyboru sposobu adaptacji (dane w pełni określone w danej chwili czasowej czy też dane napływające w czasie) jest bardzo dobrze określone, w przyszłości wybór metody może być dokonywany również zupełnie automatycznie.

Literatura cytowana

1. Duch W., Korbicz J., Rutkowski L., Tadeusiewicz R. (red.): Biocybernetyka i Inżynieria Biomedyczna. Sieci Neuronowe, EXIT, Warszawa 2000.
2. Dudek-Dyduch E., Horzyk A.: Analytical Synthesis of Neural Networks for Selected Classes of Problems, Knowlege Engineering and Experts Systems, Bubnicki Z., Grzech A. (eds): OWPN, Wrocław 2003.
3. Fiesler E., Beale R. (eds): Handbook of Neural Computation. IOP Publishing Ltd and Oxford University Press, Bristol & New York 1997.
4. Hellwig Z.: Elementy rachunku prawdopodobieństwa i statystyki matematycznej, PWN, Warszawa 1993.
5. Horzyk A.: New Efficient Ontogenic Neural Networks for Classification Tasks. Advanced Computer Systems, Soldek J., Drobiazbiewicz L. (eds) INFORMA, Szczecin 2003.

6. Horzyk A.: New Ontogenic Neural Network Classificator Based on Probabilistic Reasoning. Advances in Soft Computing. Neural Networks and Soft Computing, Rutkowski L., Kacprzyk J. (eds): Physica Verlag, Springer-Verlag Company, Heidelberg 2003.
7. Horzyk, A., Tadeusiewicz, R.: Self-Optimizing Neural Networks. Advances in Neural Networks, In: Yin F., Wang J., Guo C. (eds): ISNN 2004, Springer-Verlag, Berlin Heidelberg 2004.
8. Horzyk A.: Self-Optimizing Neural Networks as a New Computational Tool in Biomedicine, Proc. of SIIB 2004, preTEXt, Kraków 2004.
9. Novak M., Faber J., Kufudaki O.: Neuronove site a informacni systemy zivych organismu, GRADA, Praga 1992.
10. Konturek S.: Fizjologia człowieka. Neurofizjologia, AGAT PRINT, Kraków 1992.
11. Mendyk A., Horzyk A., Jachowicz R., Polak S.: Development of new microemulsions systems using ontogenic neural networks, Proc. of SIIB 2004, preTEXt, Kraków 2004.
12. Polak S., Horzyk A., Mendyk A., Skowron A., Brandys J.: Perspective of use self-optimizing neural networks in pharmacy – modeling survival time of III-B and IV-th stage Non-Small Cell Lung Cancer patients, Proc. of SIIB 2004, preTEXt, Kraków 2004.
13. Tadeusiewicz R., Wszolek W., Izworski A., Wszolek T.: Utilization of Artificial Intelligence Methods for Assistance in Interpretation of Acoustic Signals, Brambilla G., Ianniello C., Maffei L. (eds): Proc. of the 5-th European Conference on Noise Control, Naples 2003.
14. Tadeusiewicz R., Ogiela M.R.: Machine Perception and Automatic Understanding of Medical Visualization. Automatic Image Processing in Production Process, Damczyk M. (eds): Second Polish-German Seminar, CAMT, Wrocław 2003.
15. Tadeusiewicz R., Mikrut Z.: Neural-Based Object Recognition Support – From Classical Preprocessing to Space-Variant Sensing, Heiss M. (eds): Proc. of the International ICSC/IFAC Symposium on Neural Computation NC'98, ICSC Academic Press Canada/Switzerland 1998.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 21-26.

THE USEFULNESS OF PATTERN RECOGNITION METHODS AND ARTIFICIAL NEURAL NETWORKS IN OCCUPATIONAL MEDICINE

PRZYDATNOŚĆ ZASTOSOWANIA ANALIZY SKŁADOWYCH GŁÓWNYCH ORAZ SZTUCZNYCH SIECI NEURONOWYCH W MEDYCYNIE PRACY

EMILIA KOLARZYK[*], JACEK KWIATKOWSKI[*], JADWIGA HELBIN[*], AGATA PIETRZYCKA[**]

[*]*Zakład Higieny i Ekologii, Wydział Lekarski Collegium Medicum Uniwersytetu Jagiellońskiego 31-034 Kraków, ul. Kopernika 7*
[**]*Samodzielna Pracownia Radioligandów, Wydział Farmaceutyczny, Collegium Medicum Uniwersytetu Jagiellońskiego, 30-688 Kraków, ul. Medyczna 9/119*

Abstract. OBJECTIVE: Environmental exposition in the workplace of welders to free radicals and Reactive Oxygen Species (ROS) is multifactor and the response of their organisms to this exposition may be complex. This is why a wide range of clinical, biochemical and spirometrical measurements are performed to reveal early abnormalities in welders' health status.
In order to estimate the relations between obtained results of examinations and to reduce the dimensionality of the relatively large set of data, two methods have been used:
1. Pattern recognition methods: principal components analysis (PCA) and mutual contributions analysis (MCA).
2. Artificial Neural Networks (ANNs).
METHODS: There were 94 welders employed in Huta Sendzimira in Kraków, Poland (men only, aged 29–57, all active smokers) occupationally exposed to O_3 and NO_x under examination. They underwent biochemical measurements including: Total Antioxidant Status (TAS) and anti-oxidative defense enzymes level: Superoxide Dismutase (SOD) and Catalase (CT); biomineral levels: Fe, Cu, Zn, Mg in blood serum and in hair, the concentrations of albumins, bilirubin and uric acid in blood. The determination of respiratory efficiency was based on a "flow-volume" curve and spirometry. The dependent variables for ANNs were: TAS, SOD, and CT.
RESULTS: Both methods were useful for reduction of the dimensionality and were suitable for complex analysis and relatively large set of parameters when it was unknown which of these parameters were related. PCA showed that the most relevant parameters were: FEV_1, MEF_{50}, bilirubine, Fe, Zn, Cu (all elements in blood, not in hair) TAS and SOD. Only the relations of MEF_{50}, Cu and TAS were statistically valid. ANNs found statistically valid relationship between TAS and Zn and Cu in blood and two pulmonary parameters: FEV_1 and $MEF_{25/75}$.
CONCLUSIONS: Both methods proved the existence of less known relation between antioxidant defense and the efficiency of the respiratory tract in smoking welders. This relation is important for occupational medicine

Key words: pattern recognition methods, principal components analysis, mutual contributions analysis, artificial neural networks, respiratory tract, total antioxidant status

Streszczenie. Źródło ekspozycji na wolne rodniki i reaktywne pochodne tlenu w środowisku pracy spawaczy jest wieloczynnikowe, a odpowiedź organizmu na tę ekspozycję może być ogólnoustrojowa. W związku z tym dla określenia charakteru i stopnia natężenia tej odpowiedzi wykonywany jest szeroki zakres badań klinicznych, biochemicznych i laboratoryjnych. Uzyskane z badań diagnostycznych wyniki wykazują różny stopień wzajemnych zależności. W celu uściślenia relacji między nimi oraz wyselekcjonowania parametrów istotnie ze sobą powiązanych zastosowane mogą być 2 metody:
1. PCA (*Principal Components Analysis*) – Analiza Składowych Głównych i MCA (*Mutual Contributions Analysis*) – Analiza Wzajemnych Udziałów.
2. Sztuczne Sieci Neuronowe (ANNs).
Badaniami objęto 94 spawaczy w wieku 29-57 lat zatrudnionych w Hucie im. T. Sendzimira w Krakowie. Wszyscy badani mężczyźni byli palaczami papierosów, a w miejscu pracy byli wystawieni na działanie ozonu i tlenków azotu w stężeniu przekraczającym NDS. W ramach badań okresowych mieli wykonane pomiary równowagi oksydacyjno-antyoksydacyjnej, obejmujące całkowitą wydolność antyoksydacyjną (TAS) i aktywność enzymów: dysmutazy ponadtlenkowej (SOD) i katalazy (CT); oznaczenia stężeń składników mineralnych: Fe, Cu, Zn, Mg w krwi i we włosach oraz oznaczenia stężeń albuminy, bilirubiny i kwasu moczowego we krwi. Oznaczono również parametry określające wydolność układu oddechowego, które uzyskano z krzywej „Przepływ-objętość" i spirometrii. Zmiennymi zależnymi dla ANNs były TAS, SOD, CT.
Obie metody okazały się użyteczne w redukcji rozproszenia rezultatów uzyskanych z badań diagnostycznych. Były również przydatne w kompleksowej analizie dużej liczby wyników, a jednocześnie w uściśleniu tych relacji, które są ważne z medycznego punktu widzenia. Metoda PCA wykazała, że najistotniejszymi parametrami były: TAS i SOD, bilirubina, Fe, Zn, Cu (oznaczane we krwi, a nie we włosach) oraz FEV_1, MEF_{50}. Tylko relacja między MEF_{50}, Cu i TAS była istotna statystycznie. Metodą sieci neuronowych uzyskano istotny związek między TAS oraz Zn i Cu we krwi i dwoma parametrami wentylacyjnymi: FEV_1 i $MEF_{25/75}$.
Przy użyciu obu metod udowodniono mniej znany związek między wydolnością antyoksydacyjną i sprawnością wentylacyjną układu oddechowego spawaczy będących czynnymi palaczami tytoniu. Znajomość tych relacji jest ważna dla lekarzy medycyny przemysłowej.

Słowa kluczowe: Analiza Składowych Głównych, Analiza Wzajemnych Udziałów, Sztuczne Sieci Neuronowe, układ oddechowy, wydolność antyoksydacyjna

1. Introduction

Environmental exposition in the workplace to free radicals and Reactive Oxygen Species (ROS) is multifactor and the response of human organism to this exposition may be complex. In case of welders the occupational environment may influence both their respiratory system efficiency as well as their antioxidant ability. Clinical observation and literature date have confirmed the disturbances of the respiratory tract connected with the exposure to oxidant gases and fumes generated during welding [1], [2]. This occupational exposure may also lead to the impairment of oxidant/antioxidant balance. Both ozone and nitrogen oxides are capable of depleting antioxidant resources of the organism [3], [4]

The welders have undergone preventive, periodic examinations at the Industrial Health Centre. Clinical and laboratory tests have been performed within the annual, obligatory, medical check-up and a wide range of results have been obtained. In order to estimate the relations between obtained results of examinations and to reduce dimensionality of relatively large set of data the using of two methods: Pattern recognition methods: principal components analysis (PCA) and mutual contributions analysis (MCA) [5] and Artificial Neural Networks (ANNs) [6], [7], [8], [9], 10], [11] seems to be helpful.

The aim of this study was to explain the relations between the most important anti-oxidative defense parameters, including Total Antioxidant Status (TAS), catalase (CT) and superoxide dismutase (SOD), and respiratory parameters, some biominerals concentration in blood serum and their content in hair, depending on age and seniority of work.

Material and methods

The subjects of the study were 94 welders aged 41.2 ±10.0 years, employed in Huta Sendzimira, Krakow. All were active smokers, smoking at least 20 cigarettes a day for more than 5 years. All of them worked for 10 or more years in the same workplace. They were engaged both in the manual electric arc or gas welding without an inert gas blanket. They were exposed to nitric oxides and ozone in concentrations exceeding the threshold limit values.

The control group, matched by gender and seniority of work to the group of welders, consisted of 115 healthy workers aged 40.8 ±10.2 years, performing virtually the same tasks, except welding.

The antioxidant defense parameters were measured in saliva. The total antioxidant status (TAS) was measured by using the spectrophotometric Randox reagent kits (Randox Laboratories Ltd. U.K.) and anti-oxidative defence enzymes activity: Catalase-CT (according to Aebi's method) and Superoxide dismutase -SOD (according to the adrenaline method). Biominerals: Fe, Cu, Zn, Mg in blood serum and hair were measured using the absorption atomic spectrometry method (by Perkin Elmer AA 800 USA). The concentrations of albumin, bilirubin, uric acid in fasting, venous blood serum were measured by standard methods of clinical chemistry, using biochemical analyser (Cone-Pro, Finland). Lung function parameters (FVC – Forced Expiratory Vital Capacity; FEV_1 – Forced Expiratory Volume after 1 sec.; FEF_{75} – Forced Expiratory Flow at 75% of VC; PEF – Peak Expiratory Flow; FEF_{25} – Forced Expiratory Flow at 25% of VC; FEF_{50} – Forced Expiratory Flow at 50% of VC; $MEF_{25/75}$ – Middle Expiratory Flow at 25 -75% of VC) were obtained from the flow-volume curves and spirometry by computer aided system (Lungtest MES, Poland). In order to eliminate the influence of age, height and body mass on spirometric parameters, all the obtained values were expressed as the percentage of the predicted values (%N) derived from the appropriate equations of regression [12].

Pattern recognition methods: principal components analysis (PCA) and mutual contributions analysis (MCA) and Artificial Neural Networks (ANNs) were used to achieve a reduction of dimensionality and to study the influence of age and seniority of work on the data's structure.

Results

The descriptive statistics of the obtained laboratory data and respiratory parameters taken to the further statistical analysis are presented in Table 1 and Table 2.

Biochemical data were within normal limits. The mean values of basic spirometric parameters: (FEV_1, FVC, PEF, FEV_1%FVC, FEF_{25}, FEF_{50}, $FEF_{25/75}$) expressed as the percentage of predicted values were also within the normal limits (>80%N), however about 1/3 of examined welders had small or central bronchi obstruction in a mild form. There were no persons with a pure form of restrictive respiratory disorders.

Pattern recognition methods

Applying PCA to the data presented in the Tables 1 and 2, resulted in the reduction of dimensionality. The most relevant parameters were: FEV1, MEF50, bilirubin, Fe, Zn, Cu (all elements in blood), TAS, SOD (Fig. 1).

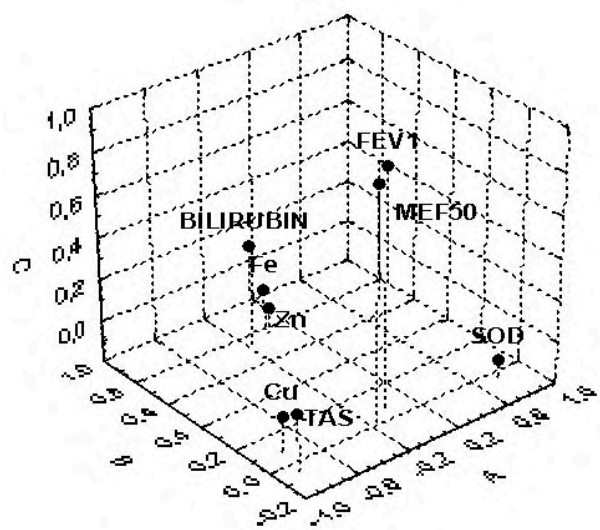


Fig. 1. PCA, plot of first three principal components.
A - 1st Principal Component; B - 2nd Principal Component;
C - 3rd Principal Component.

Table 1. Descriptive statistics of age, seniority of work and laboratory parameters

Parameters	X	SD	Median	Minimum	Maximum
Age [years]	42.09	10.1	44.0	29.0	57.0
Seniority of work [years]	23.70	8.89	25.0	10.0	36.0
TAS [mmol/L]	1.15	0.815	1.06	0.169	4.02
SOD [U/g Hb]	767.62	733.3	1160.7	237.0	4702.7
CT [U/g Hb]	18.1	14.01	26.72	1.11	76.8
Albumin [g / L]	45.0	5.65	44.8	30.9	54.6
Uric Acid [μmol / L]	281.64	84.91	309.0	117.0	422.0
Bilirubin [μmol / L]	9.53	4.32	7.9	4.2	21.4
Fe in blood serum [μmol / L]	21.52	6.81	20.8	9.71	39.1
Cu in blood serum [μg / 100 ml]	106.30	40.27	100.8	24.3	250.0
Zn in blood serum [μg / 100 ml]	124.75	38.9	121.0	23.4	249.5
Mg in blood serum [mmol / L]	0.73	0.06	0.73	0.65	0.88
Cu in hair [μg / g hair]	18.28	7.49	18.4	5.94	40.03
Zn in hair [μg / g hair]	166.67	82.59	145.2	67.9	392.7
Mg in hair [μg / g hair]	46.67	24.32	40.21	15.67	113.6

Table. 2 Spirometric parameters in welders

Spirometric parameters	x	SD	Median	Minimum	Maximum
FVC [% N]	99.71	8.61	98.58	84.96	114.25
FEV_1 [% N]	97.82	11.81	98.23	53.75	105.46
MEF_{75} [% N]	98.61	19.35	104.13	39.49	125.42
PEF [% N]	103.68	23.7	98.71	60.41	144.85
$FEV_{1\%}FVC$	82.25	8.07	82.07	56.17	96.45
MEF_{25} [% N]	94.34	23.28	91.49	31.39	147.95
MEF_{50} [% N]	99.31	23.66	97.05	30.25	140.31
$MEF_{25/75}$ [% N]	94.93	20.67	91.50	33.26	131.15

The most relevant parameters were chosen according to the following criteria:

1) the proportion of the total variation explained by the first *k* PC (principal component);

2) the coefficients of linear correlations between PC and original parameters;

3) the squared multiple correlation coefficients obtained from linear regression of each parameter on the first *k* PC.

K-L projection showed that age and seniority of work did not exert any "segregating" effects on the welders.

The results of mutual contribution analysis (MCA) are presented in Fig 2.

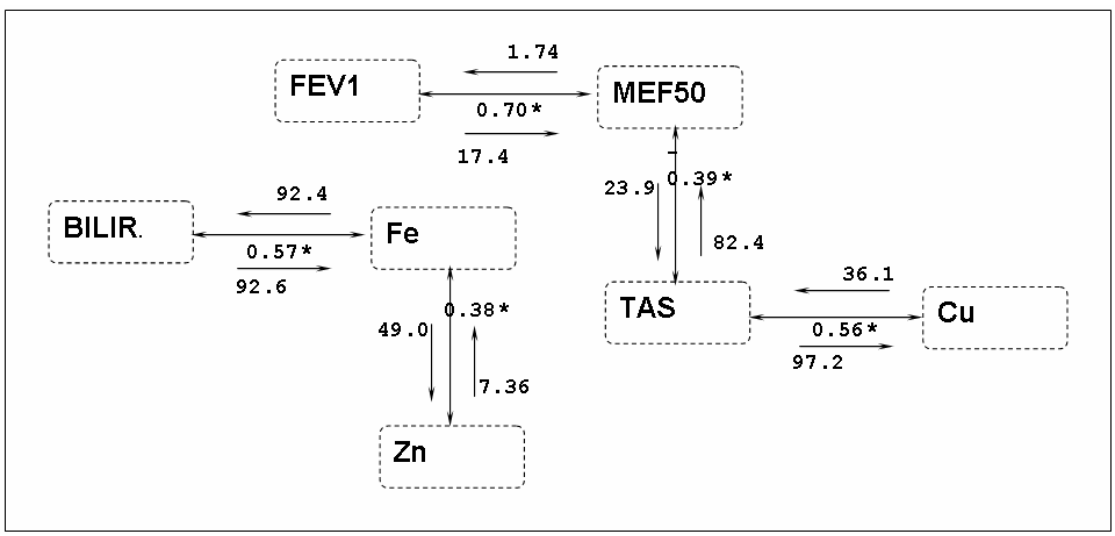


Fig. 2. MCA, visualization of the mutual contributions of the selected parameters to their variability. The values of the mutual contribution coefficients are expressed in per cent and normalized to 100%

* The partial correlation coefficient is significant at $p \leq 0.05$

It was revealed that in the studied population of welders significant ($p<0.05$) and asymmetrical interactions exist between TAS and Cu ($\delta_{Cu\text{-}TAS}$=97.2%, $\delta_{TAS\text{-}Cu}$=36.1), as well as between TAS and MEF50 ($\delta_{MEF50\text{-}TAS}$=82.4%, $\delta_{TAS\text{-}MEF50}$=23.9%). The first correlation is positive (r=0.556) while the latter is negative (r=-0.386). However, for the pair Fe-Bilirubin the interaction is significant, positive (r=0.569) but symmetrical ($\delta_{Fe\text{-}Bilir.}$=92.6%, $\delta_{Bilir.\text{-}Fe}$=92.4%).

Artificial Neural Networks (ANNs).

Artificial neural networks (ANNs) are non-linear, information processing systems, which are designed in a manner that is similar to the biological neural structures [6], [11]. This similarity can be revealed in the structural and the functional composition of ANNs as well. There are artificial neurons connected hierarchically by so-called "weights" simulating biological synapses in their structures (Fig. 3).

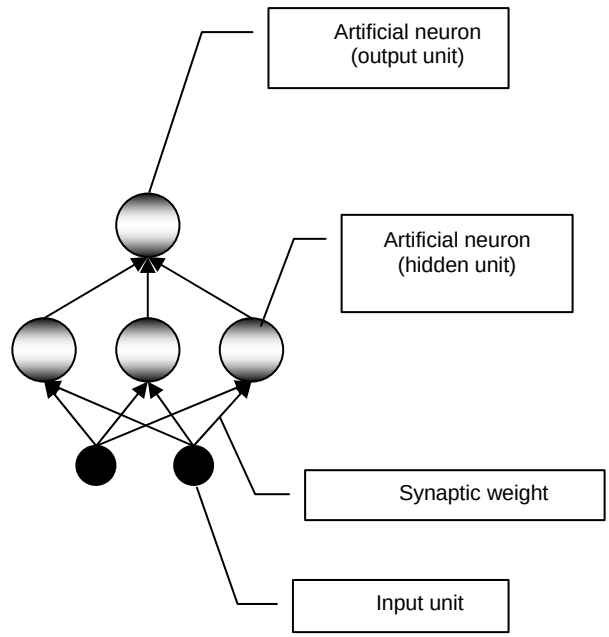


Fig. 3. Artificial Neural Network: chart

In the study, three-stage method to create an appropriate neural model was applied. The dependent variables of the model for ANNs were three parameters of antioxidative balance: TAS, SOD and CT. The variables representing pulmonary parameters, level of minerals in blood and in hair, protein and uric acid from blood samples were used to feed the ANNs. Neural models for each output variable were created separately.

During the first stage of this work it was impossible to find reasonable reduction of dimensions for CT and SOD. It was also proved that relying on available variables is only possible to create neural model for TAS. The comparison of TAS values which were obtained from examinations of welders and control group is shown in Fig. 4.

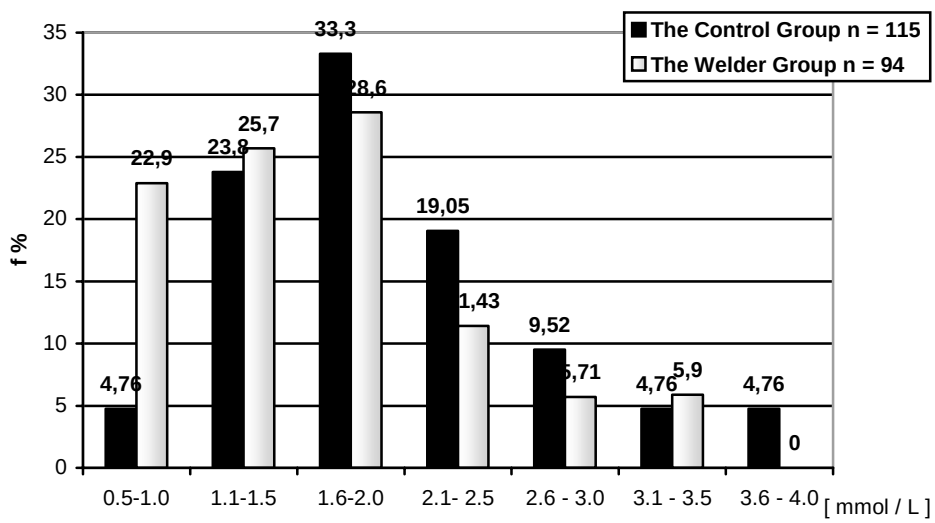


Fig. 4. Total Antioxidant Status (TAS) in groups of welders and controls

In the lowest range of activity of TAS, the frequency of welders was significantly higher than in controls in the same range of values.

The best neural model architecture is presented in Fig. 5

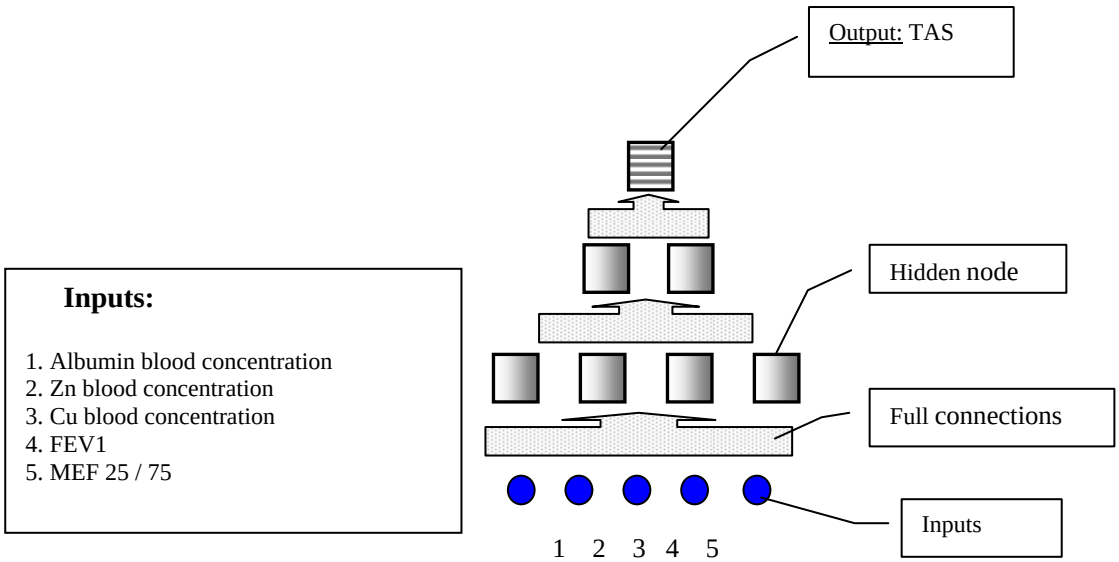


Fig. 5. The best neural model architecture

Statistically valid relationship between TAS and Zn and Cu in blood and two pulmonary parameters: Forced Expiratory Volume after 1 sec (FEV_1) and Middle Expiratory Flow of 25-75% of Vital Capacity ($MEF_{25/75}$) was found. Zn concentration almost linearly influenced TAS. For Cu a sigmoid curve was obtained.

Discussion

Our examinations confirmed that occupational exposure to nitrogen oxides and ozone lead to the impairment of oxidant/antioxidant balance and to the depletion of antioxidant ability. The exposure of airways to these oxidant gases may result in pulmonary injury. It may be characterized by the damage of the epithelium of the upper airways, bronchi, terminal bronchioles and alveoli. Finally, as a consequence of inflammation, chronic obstructive lung disease (COPD) and other various diseases of respiratory tract may occur [1], [2]. Our examinations indicated that in comparison to the controls, the antioxidant system was affected in workers chronically exposed to welding fumes and gases, which are thought to be oxidant pollutants. COPD and small airway disease in the welders group was more frequent than in the control group [13]

Obtained results indicate that biological response of welders to occupational exposure and cigarette smoking was multifactor but antioxidant resources of the organism and respiratory efficiency remained in closed relation [13], [14].

Conclusion

Pattern recognition methods: principal components analysis (PCA) and mutual contributions analysis (MCA), and Artificial Neural Networks (ANNs) were useful for reduction of the dimensionality and were suited to complex analysis and relatively large set of parameters when it was unknown which of these parameters was related.

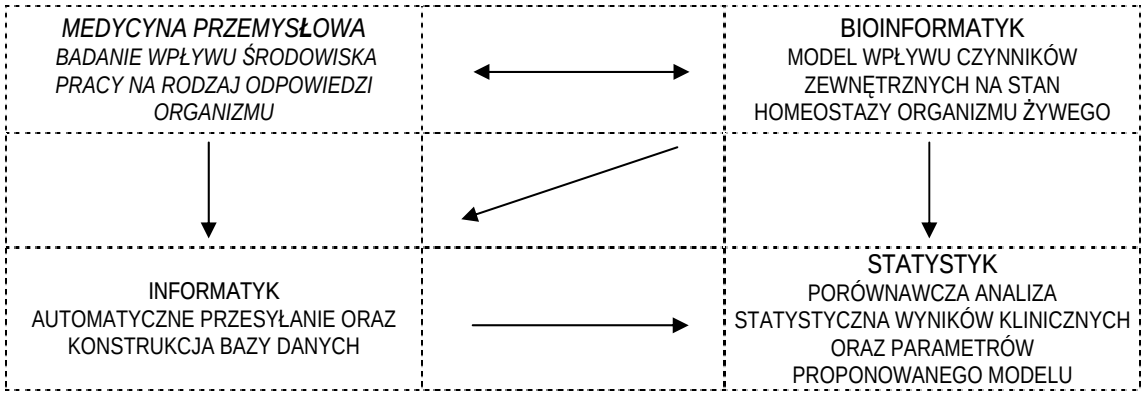
Both methods proved the existence of relation between antioxidant defense and the efficiency of the respiratory tract in smoking welders. This relation, although less known to investigators of antioxidant defense and pulmonologists seems to be important for occupational medicine.

References

1. Wolf C., Pirich C., Valic E., Waldhoer T.: Pulmonary function and symptoms of welders, Int Arch Occup Environ Health, 69: 350-3, 1997.
2. Sobaszek A., Edme JL., Boulenguez C., Shirali P., Mereau M., Robin H., at al.: Respiratory symptoms and pulmonary function among stainless steel welders, J Occup Environ Med, 40: 223-9, 1998.
3. Abrashitova N. F.: Indices of lipid peroxidation and antioxidants in workers of ecologically unfavorable productions Aviakosm, Ekol. Med, 33: 71-74, 1999.
4. MacNee W.: Oxidants/Antioxidants and COPD, *Chest* 117, 303S-317S, 2000.
5. Wold S.: Principal component analysis, Chemokkjmed. Intel. Lab. Systems, 2: 37-52, 1987.
6. Baxt S., William G.: Application of artificial neural networks to clinical medicine, Lancet, 364: 1135-1138, 1995.
7. Bilski J.: The back propagation learning with logarithmic transfer function, Proceedings of the 5th Conference Neural Networks and Soft Computing, 71, Zakopane 2000.
8. Bourquin J., Shmidli H., van Hoogevest P., Leuenberger H.: Advantages of artificial neural networks (ANNs) as alternative modelling technique for data sets showing non-linear relationship using data from a galenical study on a solid dosage form. Eur. J Pharmac., Science 7: 5-16, 1998.
9. Tadeusiewicz T.: Elementary introduction to neural networks technique with exemplary programme, pp. 1-51, PLJ Warszawa 1999.
10. Yager R. R., Filev D. P.: Polish edition of: Essentials of fuzzy modelling and control, John Wiley & Sons, Inc. WNT Warszawa 1994.

11. Hertz J., Krogh A., Palmer R. G.: Introduction to neural calculations theory. 2nd ed. Warszawa, WNT 1995.
12. European Community for Coal and Steel. Standardisation of lung function tests. Bulletin of European Physiopathology and Respiration, 19, suppl 5, 1989.
13. Stępmiewski M., Kolarzyk E., Pietrzyka A., Kitliński M., Helbin J., Brzyszczan K.: Antioxidant enzymes and pulmonary function of welders employed in steel mill. A simple description. Int. J Occup. Med. Environ. Health, 16: 41-47, 2003.
14. Stępniewski M., Kolarzyk E., Zagrodzki P., Solarska K., Pietrzyka A., Kitliński M., Schlegel-Zawadzka M., Batko B.: Pattern Recognition Methods in evaluation of the structure of the laboratory data biominerals, antioxidant enzymes, selected biochemical parameters, and pulmonary function of welders, Biol. Trace Elem, Res., 93: 39-46, 2003.

Author to whom reprint requests and all correspondence are to be sent:
Assoc. Prof. Emilia Kolarzyk ScD, PhD
Department of Hygiene and Ecology, Medical College, Jagiellonian University
7 Kopernika Str. 31-034 Kraków, Poland
Phone/fax: +48 (12) 4223720
E-mail: mykolarz@cyf-kr.edu.pl



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 27-30.

BADANIE WŁAŚCIWOŚCI ASOCJACYJNYCH MASZYNY LSM ZŁOŻONEJ Z NEURONÓW HODGKINA-HUXLEYA

STUDIES ON THE ASSOCIATION PROPERTIES OF THE LSM NETWORK OF THE HODGKIN-HUXLEY NEURONS

SIEMOWIT MUSZYŃSKI*, WIESŁAW A. KAMIŃSKI**

** Katedra Fizyki, Akademia Rolnicza, Lublin, smuszyn@ursus.ar.lublin.pl*
*** Zakład Fizyki Systemów Złożonych, Uniwersytet Marii Curie-Skłodowskiej, Lublin, kaminski@neuron.umcs.lublin.pl*

Streszczenie. Praca zawiera wyniki symulacji, przeprowadzonych w czasie badania podstawowych zdolności asocjacyjnych sieci Liquid State Machine (LSM), zbudowanej na neuronach Hodgkina-Huxleya. Przebadano również możliwość zastosowania algorytmu Hebba w procesie uczenia takiej sieci.

Słowa kluczowe: Liquid State Machine, neurony Hodgkina-Huxleya, uczenie nienadzorowane, odległość Hamminga

Abstract. The aim of this work was to examine the Hebbian learning capabilities of the new neural microcircuit model called Liquid State Machine (LSM) build on Hodgkin-Huxley neurons. LSM network does not require a task-dependent construction of a neural circuit, and has the real-time computing capability.

Key words: Liquid State Machine, Hodgkin-Huxley neurons, unsupervised learning, Hamming distance

1. Wstęp

Przez całe lata za podstawową jednostkę obliczeniową w mózgu uważano komórkę nerwową. Ostatnie badania wykazały jednak, że w układach biologicznych pojedynczy neuron nie jest w stanie przetwarzać wszystkich docierających do niego informacji. Rolę podstawowej jednostki obliczeniowej spełniają układy neuronów, tworzące tak zwane mikroobwody, których istnienie zakładało wiele teorii psychologicznych już na początku XX wieku [1]. Do badania mikroobwodów stosuje się obecnie głównie symulacje komputerowe, których analizę uważa się teraz za równoważną obserwacjom biologicznym [2], [3]. Architektura Maszyny LSM zaproponowana została przez informatyka Wolfganga Maassa i neurobiologa Henry'ego Markrama [4]. Cechą charakterystyczną tej sieci jest jej uniwersalność – nie wymagane jest dostosowywanie budowy sieci indywidualnie do każdego zadania. Drugą ważną cechą jest fakt, że sieć taka pracuje w sposób ciągły i na bieżąco analizuje dostarczany jej strumień danych.

Sieć LSM zbudowana jest z trzech elementów. Pierwszą cześć stanowi warstwa neuronów wejściowych (moduł wprowadzający), na którą podawany jest sygnał zewnętrzny. Przekazywany jest on do wybranych neuronów środkowej kolumny, zwanej kolumną cieczową. Tu dokonywane są obliczenia neuronowe. Sygnał, który opuszcza kolumnę środkową stanowi odpowiedź pewnej grupy neuronów tej kolumny. Kierowany jest on do neuronów kolumny odczytującej (moduł odczytujący), gdzie następuje analiza i interpretacja obliczeń wykonanych w kolumnie cieczowej. Oczywiście, możliwe jest zbudowanie kilku modułów odczytujących, każdy połączony z innymi neuronami i wyspecjalizowany do analizy innej właściwości sygnału generowanego w kolumnie cieczowej.

W obrębie kolumn wejściowej i wyjściowej neurony nie są ze sobą połączone. Połączenia między neuronem *A* i neuronem *B* w kolumnie środkowej generowane są z prawdopodobieństwem

$$P(D) = C \cdot e^{\frac{-D(A,B)}{\lambda}} . \qquad (1)$$

D(A,B) oznacza euklidesową odległość miedzy neuronami *A* i *B*. Parametr *λ* określa gęstość połączeń. W zależności od tego, czy połączenie tworzone jest pomiędzy neuronami pobudzającymi (*Ex*) a hamującymi (*Inh*) współczynnik C przyjmuje wartości: $C_{Ex\text{-}Ex}=0,3$, $C_{Ex\text{-}Inh}=0,2$, $C_{Inh\text{-}Inh}=0,1$ i $C_{Inh\text{-}Ex}=0,4$. Udział neuronów hamujących w sieci, zgodnie z wynikami Markrama wynosi 30%.

Widać wiec, że w przypadku LSM nie można mówić o sieci z zadaną z góry architekturą. Co więcej, aby z takiej sieci korzystać, nie musimy znać dokładnej struktury połączeń w kolumnie cieczowej. Interesuje nas tylko odpowiedź wybranych neuronów na wprowadzony do niej sygnał. Sieć LSM nie jest więc siecią z propagacją aktywności do przodu, lecz siecią rekurencyjną. Wprowadzenie neuronów hamujących powoduje, że w sieci takiej istnieje skończony czas samowzbudzeń (sprzężeń zwrotnych).

2. Cel pracy i metody

W swoich badaniach twórcy LSM używali neuronów *Integrate-and-Fire*. Proces generowania przez taki neuron odpowiedzi na impuls pobudzający jest diametralnie inny niż w przypadku neuronu Hodgkina-Huxleya (HH). Celem pracy było więc porównanie zdolności asocjacyjnych maszyny LSM zbudowanej z neuronów HH z wynikami opisanymi w pracy [4]. Dodatkową różnicą było to, że podawany do sieci sygnał nie był wzorcem zakodowanym czasowo, lecz był sygnałem zakodowanym przestrzennie.

W celu zbadania aktywności sieci w wyniku wprowadzenia do niej dwóch różnych ciągów impulsów wykorzystano analizę odległości Hamminga. Pozwala ona określić zdolności asocjacyjne nieuczonej sieci, jak również ocenić efektywność i przebieg procesu jej uczenia. Odległość Hamminga $H_{W1W2}(t)$ definiowana jest jako liczba neuronów, których stan w danym kroku czasowym *t* po wprowadzeniu wzorca W1 różni się od stanu tych neuronów w przypadku pokazywania wzorca W2:

$$H_{W1W2}(t) = \frac{1}{N}\sum_{i=1}^{N}\left|y_i^{W1}(t) - y_i^{W2}(t)\right|, \qquad (2)$$

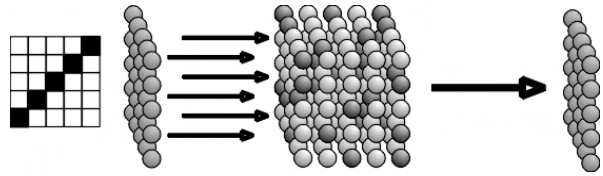
gdzie N oznacza liczbę obserwowanych neuronów, $y_i^{W1}(t)$ i $y_i^{W2}(t)$ oznaczają kolejno stan *i*-tego neuronu w czasie *t* po wprowadzeniu do sieci wzorca W1 i W2. Odległość Hamminga $H_{W1W2}(t)=0$ oznacza całkowitą zgodność stanów w czasie *t*. Odległość $H_{W1W2}(t)=1/2$ mają stany zgodne w połowie, a $H_{W1W2}(t)=1$ oznacza, że wzorzec W1 jest „negatywem" wzorca W2.

Dla sieci rekurencyjnych, a taką jest sieć LSM, nie można określić błędu generowanego na poszczególnych neuronach w kolejnych krokach czasowych. Dlatego dla neuronów z modułu cieczowego proces uczenia jest procesem nienadzorowanym i przebiega według algorytmu Hebla. W tym przypadku wagi połączenia synaptycznego $\Delta\omega_{AB}$ pomiędzy neuronami *A* i *B* zmieniają się zgodnie z formułą [5]:

$$\Delta\omega_{AB} = \beta\Delta y_B(t)\Delta y_A(t), \qquad (3)$$

gdzie β jest współczynnikiem uczenia, $\Delta y_B(t)$ i $\Delta y_B(t)$ oznaczają zmiany aktywności neuronów *A* i *b* w chwili *t*.

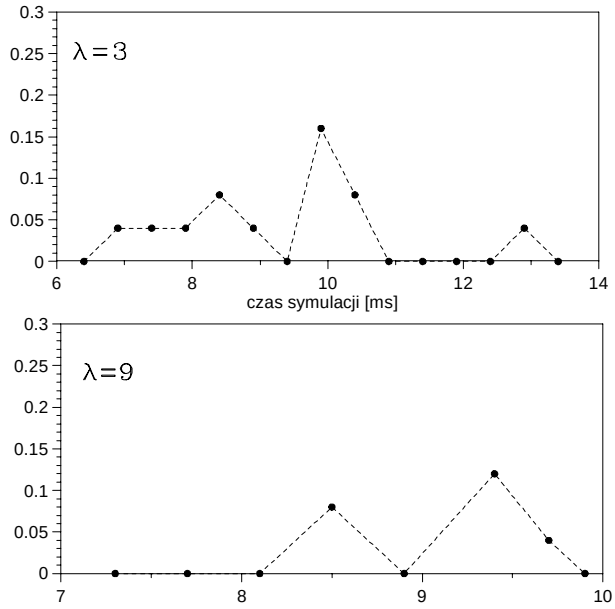
Schemat sieci LSM przedstawia rys. 1 Na każdy z 25 neuronów modułu wejściowego podawana jest jedna składowa impulsu pobudzającego (wzorca). Kolumna cieczowa posiadała 125 neuronów ułożonych w sześcian o bokach 5x5x5. Rozpatrywane były dwa przypadki połączeń miedzy kolumnami. W pierwszym przypadku neurony kolumny cieczowej, na które podawany był sygnał i neurony, których aktywność była obserwowana na kolumnie wyjściowej, wybierane były w sposób losowy. Tak zbudowana sieć nosi nazwę *modelu losowego*. W drugim przypadku potraktowano kolumnę cieczową maszyny LSM jak „skrzynkę" - sygnał był podawany na jedna ścianę i zbierany ze ściany przeciwnej (*model ścianowy*). W obu przypadkach rozpatrywane były dwie wartości parametru λ określającego gęstość synaps w kolumnie cieczowej. Czas symulacji wynosił *20 ms*. Wszystkie symulacje przeprowadzone zostały przy użyciu symulatora sieci GENESIS [6].



Rys. 1. Schematyczny obraz sieci LSM. Na moduł wejściowy podawany jest wzorzec, przekazywany następnie do kolumny cieczowej. Wynik obliczeń zbierany jest przez moduł odczytujący

3. Wyniki

Na początku badano odległości Hamminga na kolumnie odczytującej pomiędzy wzorcem W1 i wzorcem zaburzonym W1U różniącym się od wzorca W1 jedną składową (4%). Istniało podejrzenie, że układ LSM jako sieć o losowych połączeniach będzie się zachowywał chaotycznie - nawet mała zmiana impulsu wejściowego prowadzić mogła do całkowicie innej aktywności sieci.



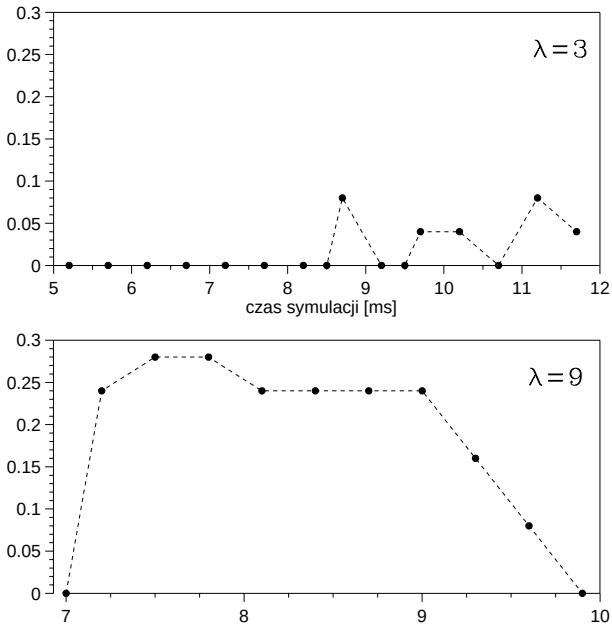
Rys. 2. Odległości Hamminga między wzorcami W1 i W1U dla *modelu ścianowego*. Panel górny odpowiada sieci o mniejszej liczbie synaps (λ=3) w stosunku do sieci z wykresu dolnego (λ=9)

Rys. 2 przedstawia wyznaczone odległości Hamminga w *modelu ścianowym* dla dwóch różnych gęstości połączeń synaptycznych. W obu przypadkach przez cały okres symulacji sieć dobrze kojarzy wzorzec zaburzony W1U ze wzorcem W1. Dla mniejszej liczby synaps (*λ=3)* odległość maksymalna $H_{max}=0,16$, co odpowiada różnicy stanów czterech neuronów.

Zwiększenie liczby synaps (*λ=9*) powoduje zmniejszenie maksymalnej liczby niezgodnych neuronów do trzech ($H_{max}=0,12$). Różne wartości przedziałów czasowych, podczas których obserwowano aktywność sieci, wynikają z faktu, że aktywność sieci zależy od jej konfiguracji, tj. od struktury połączeń synaptycznych w kolumnie cieczowej.

Odwrotna sytuacja ma miejsce w przypadku *modelu losowego*, co ilustruje rys. 3. Zwiększenie liczby synaps

(zmiana parametru *λ=3* do *λ=9*) powoduje wzrost maksymalnej odległości Hamminga od wartości $H_{max}=0,08$ (dwa neurony niezgodne) do wartości $H_{max}=0,28$ (siedem neuronów). Widać też, że dla takiej sieci podczas prawie całego czasu symulacji utrzymują się duże wartości *H*.



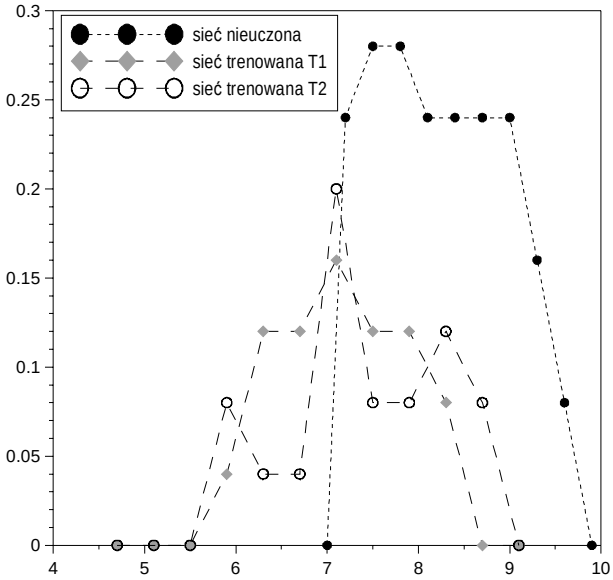
Rys. 3. Odległości Hamminga między wzorcami W1 i W1U dla *modelu losowego*. Panel górny odpowiada sieci o mniejszej liczbie synaps (λ=3), panel dolny mniejszej liczbie synaps (λ=9)

Neurony kolumny cieczowej sieci generującej największe wartości H_{max} (*model losowy*, *λ=9*) poddane zostały uczeniu w treningiu Hebba. Rozważone zostały dwa przypadki. Pierwszy można określić jako lekki trening Hebba, drugi przypadek jako silny trening Hebba.

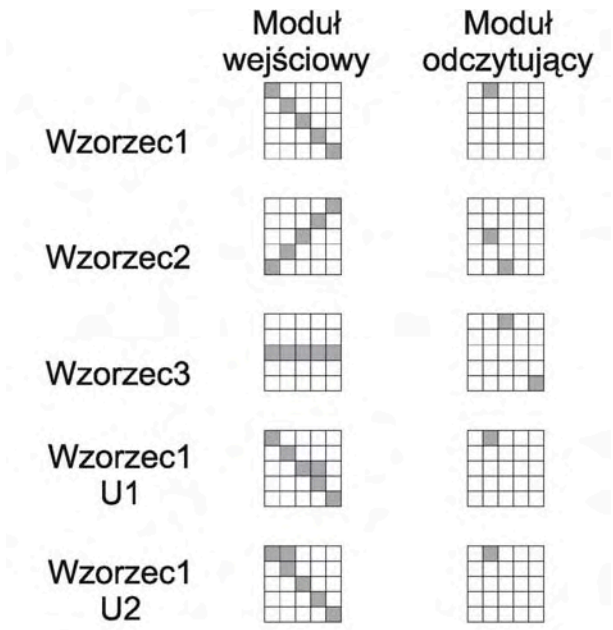
W przypadku treningu lekkiego synapsy były uaktualniane zgodnie ze wzorem (2) dla niewielkich wartości parametru ß. Rys. 4 przedstawia odległości Hamminga między wzorcem W1 i W1U przed i po zmianie wag dla dwóch różnych wartości parametru ß (sieci T1 i T2). Widać wyraźne zmniejszenie się liczby niezgodnych neuronów. Różne czasy aktywności sieci wynikają z faktu, że architektura sieci w wyniku uczenia uległa modyfikacji – po treningu impuls szybciej dociera do neuronów warstwy odczytującej, co interpretować można jako dowód wytworzenia się mikroobwodów wewnątrz kolumny cieczowej.

Podczas treningu silnego wartości współczynnika ß były o wiele większe niż w poprzednim przypadku. Celem symulacji było wymuszenie tak dużych zmian w architekturze sieci, że jednemu wzorcowi odpowiadała aktywność wybranych neuronów modułu odczytującego. Wyniki treningu Hebba dla wybranych wzorców zarówno pierwotnych, jak i zaburzonych przedstawia rys. 5.

Wybór wartości parametru ß w obu przypadkach dobierany był drogą prób i błędów. Do każdego problemu (wzorzec uczący, struktura połączeń) parametr ß musi być dobierany indywidualnie. Warunek ten dotyczy zwłaszcza przypadku treningu silnego, ponieważ w treningu lekkim w większości przebadanych wartości parametru ß zaobserwowano zmniejszenie się odległości Hamminga pomiędzy wzorcem pierwotnym i wzorcem zaburzonym.



Rys. 4. Odległości Hamminga między wzorcami W1 i W1U dla *modelu losowego* o gęstości synaps λ=9. Odległości H wyznaczono dla sieci nieuczonej, oraz dla sieci uczonej algorytmem Hebba dla dwóch różnych wartości parametru ß (sieci T1 i T2)



Rys. 5. Odpowiedzi modułu odczytującego dla sieci *losowej* poddanej silnemu treningowi Hebba. Przedstawione zostały wyniki dla czterech wzorców uczących oraz przykładowych dwóch wzorców testujących U1 i U2

4. Podsumowanie

W pracy przebadano w symulacjach komputerowych podstawowe cechy maszyny LSM. Stosując miarę odległości Hamminga wykazano, że sieci tego typu posiadają zdolności asocjacyjne. Zdolności te występowały we wszystkich przebadanych konfiguracjach sieci. Zastosowanie treningu Hebba pokazało, że możliwe jest również wymuszenie takich zmian w architekturze sieci, aby w obszarze kolumny cieczowej zachodziło klasyfikowanie wzorców.

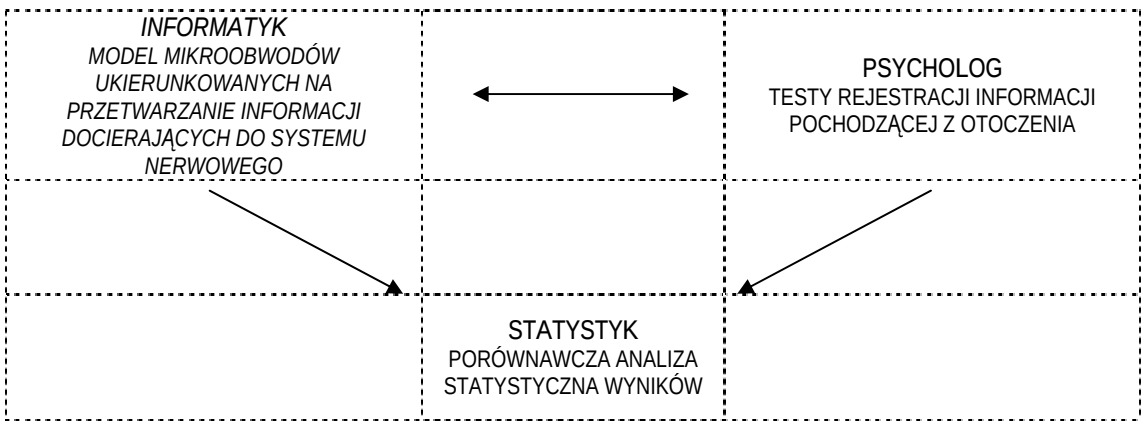
Dalsze badania nad maszyną LSM powinny skoncentrować się na problematyce uczenia takiej sieci. Preferowaną formą uczenia neuronów kolumny cieczowej powinien być trening Hebba. Zagadnieniem otwartym jest takie optymalizowanie architektury sieci, aby pojemność maszyny LSM była zbliżona jakością do klasycznych sztucznych sieci neuronowych [7]. W tym celu można zbudować w obszarze kolumny cieczowej hebbowskie zespoły komórkowe. Alternatywnym rozwiązaniem jest stworzenie wielu modułów odczytujących, wyszkolonych w analizie jednej cechy sygnału podawanego na moduł wejściowy sieci. W takim przypadku modyfikowane powinny być połączenia synaptyczne między kolumną cieczową a modułem odczytującym. Jest to o tyle łatwiejsze, że można wtedy zastosować algorytmy uczenia nadzorowanego.

Kolejną czynnością powinno być określenie sposobu magazynowania informacji przez sieć [8]. Miejscem, gdzie sieć rekurencyjna magazynuje informacje o pokazywanych jej wzorcach wydają się być synapsy, a dokładnie zmiany zachodzące pod wpływem uczenia w wagach połączeń synaptycznych [9], [10].

Architektura sieci LSM zaproponowana w [4] stała się już podstawą kilku ambitnych projektów neuroobliczeniowych, opisanych w pracach [11], [12], [13], [14]. Przeszkodą w realizacji tego typu symulacji pozostają jednak w dalszym ciągu zdolności obliczeniowe komputerów.

Literatura cytowana

1. Obrach J.: The Neuropsychological Theories of Lashley and Hebb, University Press of America, Lanham 1998.
2. Amit D. J.: Simulation in neurobiology – theory or experiment?, TINS, 21: 231–237, 1998.
3. Hansel D., Sompolinsky H.: Chaos and synchrony in a model of a hypercolumn in visual cortex, Journal of Computational Neuroscience, 3: 7-34, 1996.
4. Maass W., Markram H.: Real-time computing without stable states: A new framework for neural computation based on perturbations, Neural Computation, 14(11): 2531-2560, 2002.
5. Brown T. H., Chattatji S.: Hebbian synaptic plasticity, In: Domany E. (ed.): Models of Neural Networks II, Springer-Verlag, New York, 1995.
6. Bower J. M., Beeman D.: The Book of GENESIS. Exploring Realistic Neural Models with the General Neural Simulation System, Springer-Verlag, New York 1995.
7. Hertz J., Krogh A., Palmer R. G.: Wstęp do teorii obliczeń neuronowych, WNT, Warszawa 1993.
8. Brost A, Theunissen F. E.: Information theory and neural coding, Nature, 2(11): 947-952, 1999.
9. Amit D. J., Brunel N.: Dynamics of a recurrent network of spiking neurons before and following learning, Network: Computation in Neural Systems, 8: 373-404, 1997.
10. Maass W., Markram H.: Synapses as dynamic memory buffers, Neural Networks, 15:1 55-161, 2002.
11. Jaeger H.: The Echo State approach to analyzing and training recurrent neural network, GMD Technical Report 152, German National Institute for Computer Science, 2001.
12. Gunay C.: Realization of a spiking neuroidal model: Symbol processing with biologically constrained connectionist networks, Technical Report TR-2003-2-1, University of Louisiana, Lafayette 2002.
13. Goldenholtz D.: Liquid computing: A real effect, Technical Report BE707, Boston University, Department of Biomedical Engineering, Boston 2002.
14. Wilson E. C.: Parallel implementation of a large scale biologically realistic neocortical neural network simulator, PhD Thesis, University of Nevada, Reno 2002.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 31-36.

ALGORYTMY EWOLUCYJNE W REALIZACJI SYSTEMÓW UCZĄCYCH SIĘ

EVOLUTIONARY ALGORITHMS IN LEARNING SYSTEMS

LESZEK NOWAK*, MARIUSZ ŚWIĘCICKI**, PIOTR WAIS***

**Instytut Politechniczny, Państwowa Wyższa Szkoła Zawodowa w Krośnie, Krosno, meadow@poczta.onet.pl*
***Instytut Modelowania Komputerowego, Politechnika Krakowska, Kraków, mswiecic@pk.edu.pl*
****Instytut Politechniczny, Państwowa Wyższa Szkoła Zawodowa w Krośnie, Krosno, waisp@poczta.onet.pl*

Streszczenie. Sztuczna inteligencja jest szeroką i szybko rozwijającą się dziedziną informatyki. Często do budowy inteligentnych systemów uczących się wykorzystywane są mechanizmy, które funkcjonują w naturze od bardzo dawna. Przykładem tutaj mogą być sztuczne sieci neuronowe, wzorowane na budowie i sposobie funkcjonowania ludzkiego mózgu oraz algorytmy genetyczne opierające swoje działanie na teorii ewolucji, sformułowanej w 1859 r. przez Karola Darwina. W ostatnich latach zainteresowanie także zaczęły wzbudzać sztuczne sieci immunologiczne, których główną ideą jest odwzorowanie działania ludzkiego systemu odpornościowego. Takie mechanizmy uczenia się pozwalają na rozwiązywanie złożonych problemów obliczeniowych z zakresu przetwarzania danych, sztucznej inteligencji, klasyfikacji danych wielowymiarowych, optymalizacji funkcji, ekstrakcji wiedzy i innych. Artykuł ten zawiera zastosowanie algorytmów ewolucyjnych w rozwiązywaniu przykładowego problemu.

Słowa kluczowe: sieci immunologiczne, algorytmy genetyczne, systemy uczące się

Abstract. Artificial intelligence is a fast developing discipline of computer science. Frequently, while creating learning systems, we use solutions that have been known to nature for ages. A good example is artificial neural network that is designed to work in a way similar to human brain. Another commonly used idea is genetic algorithm based on Charles Darwin's theory of evolution. Recently artificial immune systems are gaining popularity; their basic idea is to imitate the human immune system. Those learning mechanisms give possibility to solve complex problems of data processing, data mining, data classification, optimization, machine learning, pattern detection, and many more. This paper contains evolutionary algorithm designed to solve a simple machine learning problem.

Key words: artificial immune system, genetic algorithms, learning systems

1. Mechanizmy uczenia się

Algorytmy genetyczne

Przeniesienie procesu ewolucji w świat komputerów zaowocowało powstaniem algorytmów ewolucyjnych (Rys. 1), które pozwalają w łatwy, ale uproszczony sposób zwiększyć efektywność wykonywanego procesu. Algorytmy genetyczne są cennym narzędziem używanym do budowy systemów pozwalających na przeszukiwanie rozległych przestrzeni możliwych rozwiązań. Mechanizm ten daje możliwość znalezienia pewnej populacji różnych osobników, które będą odzwierciedlały wysoce optymalne, ale niekoniecznie idealne rozwiązania danego problemu. Osobnikiem w takim przypadku może być dowolna struktura danych przedstawiająca rozwiązanie problemu. Często osobnik reprezentowany jest w postaci ciągu binarnego o określonej długości, gdzie jeden lub kilka bitów stanowi jeden gen.

Tab. 1. Przykładowy osobnik zbudowany z 8 genów

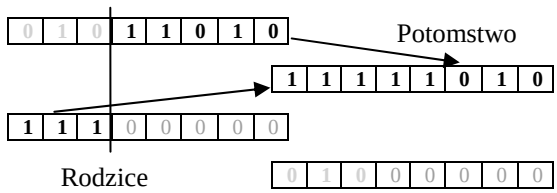
000	11	0	100	11100	00	101	11010011

Rys. 1. Ogólny schemat blokowy algorytmu genetycznego

Operatory genetyczne

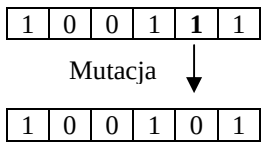
W algorytmach genetycznych stosuje się dwa podstawowe operatory: **krzyżowanie** osobników oraz **mutację**. Zadaniem tych operatorów jest tworzenie nowych osobników, i o ile krzyżowanie ma na celu tworzenie nowych populacji, to mutacja ma zapobiegać upodobnianiu się do siebie powstających osobników, dzięki czemu mamy pewność, że zostanie przeszukana w miarę duża przestrzeń możliwych rozwiązań.

Krzyżowanie jest operacją polegającą na łączeniu dwóch osobników rodzicielskich wylosowanych w procesie selekcji i stworzeniu z nich dwóch nowych, potencjalnie lepszych osobników potomnych (Rys. 2). Najczęściej w operacji selekcji stosuje się metodę ruletki, przedstawioną w [2].



Rys. 2. Przykład krzyżowania dwóch ośmiobitowych ciągów

Kolejnym etapem powinno być przeprowadzenie mutacji w nowo powstałej populacji. Proces ten wykonywany jest z pewnym niewielkim prawdopodobieństwem, aby uniknąć sytuacji, w której bardzo dobre rozwiązania są zbyt często mutowanie, co w efekcie prowadzi do zmniejszenia efektywności całego procesu. Mutacja jest prostym procesem polegającym na zamianie jednego losowo wybranego genu w chromosomie.



Rys. 3. Przykład mutacji ośmiobitowego ciągu

Po operacji mutacji powstaje nowa populacja, która powinna zawierać średnio lepsze rozwiązania niż poprzednia.

Systemy klasyfikujące

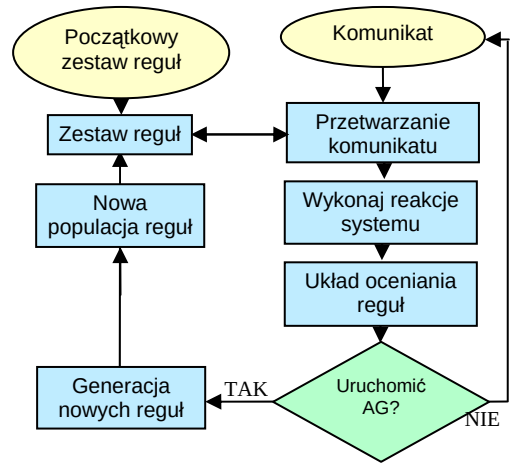
System klasyfikujący to bardzo dobre narzędzie do symulacji mechanizmów poznawczych żywych organizmów. System taki ma odzwierciedlać proste reakcje na sygnały płynące z otoczenia.

Działanie systemu klasyfikującego opiera się na uczeniu się prostych reguł (klasyfikatorów) i wykonywaniu prostych reakcji (efektorów). System klasyfikujący zbudowany jest z trzech elementów składowych:

1) układu przetwarzającego komunikaty;
2) układu oceniającego;
3) algorytmu genetycznego.

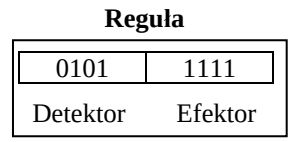
Układ przetwarzania komunikatu ma za zadanie odbieranie komunikatu z otoczenia, wyszukiwanie reguł, do których ten komunikat pasuje i odpalanie odpowiednich reguł. Reguły zbudowane są w podobny sposób jak funkcja IF w językach programowania, czyli jeżeli spełniony jest warunek, to podejmowana jest akcja. Zapis reguły wyglądać może tak:

jeżeli: <WARUNEK> to: <AKCJA>



Rys. 4. Ogólny schemat blokowy systemu klasyfikującego wykorzystującego algorytm genetyczny

Na regułę składa się detektor, który jest wzorcem sygnału ze środowiska i efektor, który jest odpowiedzią systemu na ten sygnał. Detektory mogą mieć dowolną strukturę pozwalającą na odbieranie sygnałów otoczenia. Struktura efektora powinna być odpowiednia do reakcji, jaką ma podjąć.



Rys. 5. Przykład prostej reguły

System klasyfikujący jest typowym przykładem uczenia się ze wzmocnieniem, za co odpowiada układ oceniający, którego zadaniem jest nagradzanie poszczególnych reguł za udział w rozpoznawaniu komunikatu, bądź też całych zespołów reguł, które dawały długoterminowe efekty. Reguły, które były często nagradzane, zwiększają swoją siłę, dzięki czemu wiadomo, że dawały dobre efekty i powinny być częściej używane.

Ostatnim elementem systemu klasyfikującego jest wcześniej omówiony algorytm genetyczny, którego zadaniem jest dostarczanie systemowi nowych reguł. Populację dla algorytmu genetycznego stanowi tu losowy zestaw reguł, gdzie jedna reguła odpowiada jednemu osobnikowi. W procesie selekcji do krzyżowania wybierane są najsilniejsze reguły i następnie osobniki potomne zastępują najsłabsze reguły w zestawie. Ważne jest, żeby wymianie ulegał jedynie niewielki procent zestawu reguł, w przeciwnym wypadku system szybko może utracić swoją sprawność.

Sztuczne sieci immunologiczne

Istnieje duże podobieństwo pomiędzy siecią immunologiczną a systemem klasyfikującym (Tab. 2).

Tab. 2. Podobne elementy systemu klasyfikującego i sieci immunologicznej

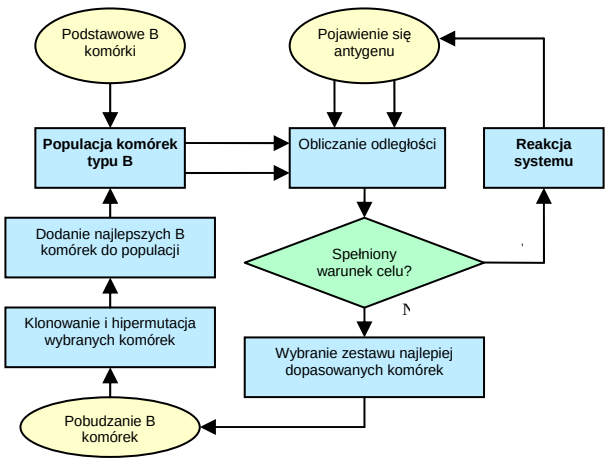
System klasyfikujący	Sieć immunologiczna
Klasyfikator	Przeciwciało
Detektor	Receptor
Efektor	Reakcja

System klasyfikujący	Sieć immunologiczna
Komunikat środowiska	Antygen
Dopasowanie	Odległość
Lista komunikatów	Sieć antygenów
Funkcja dopasowania	Stopień dopasowania antygenów
Generacja reguł	Klonowanie, hipermutacja
Siła klasyfikatora	Pobudzenie, dojrzałość
Odpalanie reguł	Negatywna selekcja

Uwagi: Tabela nie zawiera wszystkich podobieństw, jakie można zauważyć pomiędzy obydwoma mechanizmami.

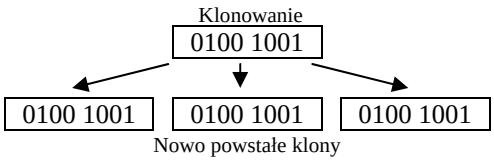
Dla sieci immunologicznych przeciwciało lub antygen mają postać ciągu binarnego lub wektora liczb rzeczywistych, co pozwala na obliczanie podobieństwa między nimi, nazywanego także odległością liczoną ze wzoru:

$$D = \sqrt{\sum_{i=1}^{L} \left(Ab_i - Ag_i\right)^2} \qquad (1)$$



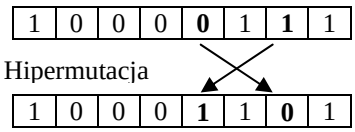
Rys. 6. Ogólny schemat blokowy sztucznej sieci immunologicznej

Pomimo podobieństwa sieci immunologicznej do systemu klasyfikującego działanie sieci jest znacząco odmienne i konieczne jest na tym etapie omówienie operacji klonowania i hipermutacji, jako że są to jedyne mechanizmy generacji nowych przeciwciał (B komórek). W sytuacji, gdy pojawienie się antygenu w systemie spowoduje pobudzenie odpowiednich przeciwciał, wtedy odpowiednio pobudzona B komórka podlega klonowaniu, czyli jest kopiowana z pewnym odpowiednim do wartości pobudzenia tej komórki prawdopodobieństwem.



Rys. 7. Przykład klonowania pobudzonego przeciwciała

W ten sposób powstałe kopie powinny ulegać procesowi hipermutacji, który odpowiada za powstawanie nowych odmiennych B komórek.



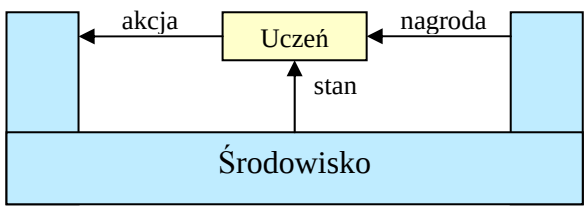
Rys. 8. Przykład hipermutacji

Odpowiednio dla nowych przeciwciał obliczana jest ze wzoru (1) odległość, co pozwala sprawdzić, które z powstałych komórek typu B są z punktu widzenia przydatności dla sieci dobre. Nieodpowiednie klony wymierają i są usuwane z sieci.

Podobnie jak miało to miejsce w systemach klasyfikujących, sieci immunologiczne są przykładem uczenia się ze wzmocnieniem. Za nagrodę tutaj możemy wziąć pobudzanie i dojrzewanie B komórek. Komórki dojrzałe i pobudzane stanowią posiadają wysoką przydatność dla systemu.

Uczenie się ze wzmocnieniem

Sprzężenie zwrotne, jakie istnieje pomiędzy systemem a środowiskiem pozwala na zastosowanie mechanizmu nagradzania, dzięki któremu możliwe jest przyznawanie systemowi „nagród" za każdym razem, gdy podejmuje on poprawną decyzję. Zastosowanie takie wymaga od systemu obecności czegoś, co pozwoliłoby na ocenienie, czy system podjął odpowiednią decyzję. Dlatego system musi zawierać funkcję celu, funkcję dostosowania lub jakąś inną technikę oceniania. W przedstawionych wcześniej mechanizmach uczenia się były to funkcje dostosowania, które pozwalały nam sprawdzić, czy reguła lub przeciwciało wpływały pozytywnie na efektywne funkcjonowanie systemu.



Rys. 9. System uczący się ze wzmocnieniem i jego środowisko [1]

1: **powtarzaj** do osiągnięcia oczekiwanego efektu;
2: pobierz nowy stan środowiska *s*;
3: wybierz akcję *a* odpowiednią dla stanu *s*;
4: wykonaj akcję *a*;
5: obserwuj wzmocnienie *w* i następny stan *s+1*;
6: ucz się na podstawie (*a, s, r, s+1*);
7: **koniec**

Rys. 10. Algorytm uczenia się ze wzmocnieniem [1]

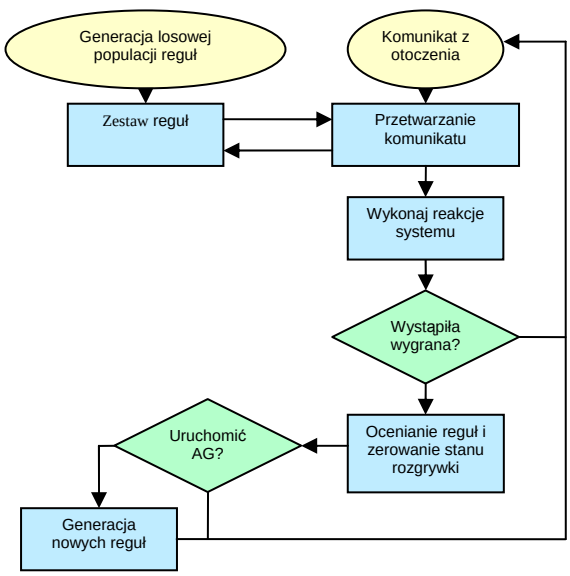
Istnienie uczenia się ze wzmocnieniem wymaga istnienia matematycznej reprezentacji stanów środowiska.

2. Przykłady uczących się systemów

Do przedstawienia możliwości systemów uczących się wybrany został problem, jaki stawia gra „kółko i krzyżyk" na planszy 30x30. Dla tego problemu możliwe jest ręczne napisanie algorytmu odpowiadającego za strategię gry, ale jak się okazuje, zajmuje to więcej czasu niż zaprojektowanie i zbudowanie systemu, który sam będzie uczył się strategii rozgrywki.

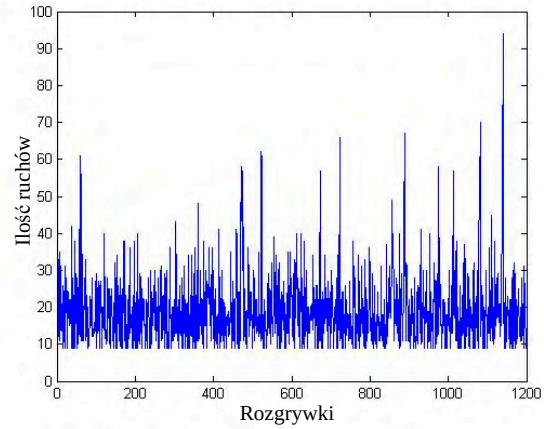
Osiągi systemu klasyfikującego

Podczas projektowania systemu podjęta została decyzja o ustaleniu rozmiaru komunikatu z otoczenia tak, aby umożliwiał on wykrycie sytuacji, w której osiągnięte zostanie zwycięstwo. Komunikat przyjął strukturę macierzy 5x5, czyli najmniejszą pozwalającą na utworzenie reguły zwycięstwa. Jak się potem okazało, rozmiar macierzy ma duży wpływ na możliwości formułowania strategii systemu. Zwiększenie rozmiaru komunikatu pozwala na wytworzenie bardziej skomplikowanych strategii, a co za tym idzie – zwiększenie skuteczności systemu. Problem, jaki się jednak pojawia, jest taki, że wraz ze wzrostem rozmiaru komunikatu należy zwiększyć liczbę reguł, na jakich system będzie operował. Ilość wszystkich możliwych kombinacji reguły dla przyjętego komunikatu wynosi 4^{25}, dla komunikatu 6x6 liczba rozwiązań zwiększyłaby się do 4^{36}. Dla porównania można powiedzieć, że przy rozmiarach 5x5 i zestawie 500 reguł procesor 2Ghz potrzebuje 1870 godzin na przeprowadzenie 2000 rozgrywek i dopiero wtedy zaczyna osiągać oczekiwane efekty. Ostatecznie liczba reguł została ograniczona do 50 i zastosowano reguły obrotowe.



Rys. 11. Ogólny schemat blokowy systemu testowego.

Projekt systemu opiera się w dużym stopniu na standardowym systemie klasyfikacyjnym, jednak zmianie uległ mechanizm przetwarzania komunikatu. Zmianę tę wymusiło zastosowanie obrotowych reguł, gdzie jedna reguła mogła pasować idealnie do 8 różnych komunikatów. Modyfikacja ta poprawiła w znaczący sposób efektywność systemu, zmniejszając wielokrotnie czas jego uczenia się.



Rys. 12. Wyniki systemu dla uczenia się bez wzmocnienia i bez użycia algorytmu genetycznego

Uczenie się tego systemu jest typowym przykładem uczenia ze wzmocnieniem. Nagradzaniu tutaj podlegają całe grupy reguł prowadzące system do osiągnięcia zwycięstwa. Nagradzanie reguł wyrażone jest wzorem:

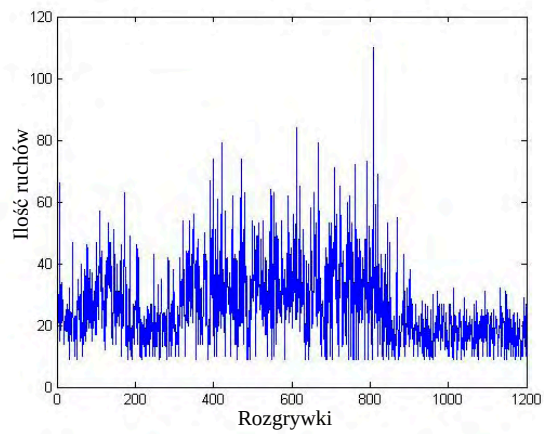
$$R = Tb * (A/25) \tag{2}$$

Gdzie *R* oznacza nagrodę dla klasyfikatora, *Tb* sumę złożonych ofert w jednej rozgrywce, *A* oznacza współczynnik dopasowania.

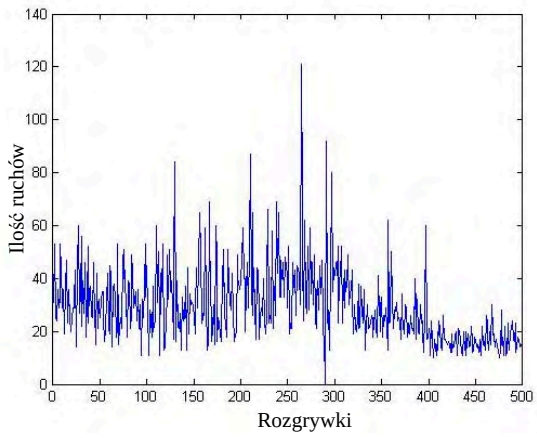
Zastosowano też karanie w sytuacji, gdy zestaw reguł prowadził do przegranej. Operacja ta jest wyrażona wzorem (3). Pozwoliło to uniknąć nieograniczonego wzrastania siły klasyfikatorów.

$$P = S * (A/25) * 0.2 \tag{3}$$

Zastosowanie algorytmu genetycznego było elementem, bez którego nie można by było osiągnąć żadnych efektów. Problem stanowi jednak dobranie parametrów algorytmu tak, aby nie miał on negatywnego wpływu na efektywność systemu, co miało dość często miejsce, a szczególnie przez operację mutacji, która często doprowadzała do sytuacji, w której idealne reguły były po prostu niszczone – można to zauważyć na rys. 13 w okolicach 300 rozgrywki, gdzie efektywność systemu spadła dużo ponad połowę.



Rys. 13. Wyniki systemu dla uczenia się ze wzmocnieniem oraz z algorytmem genetycznym

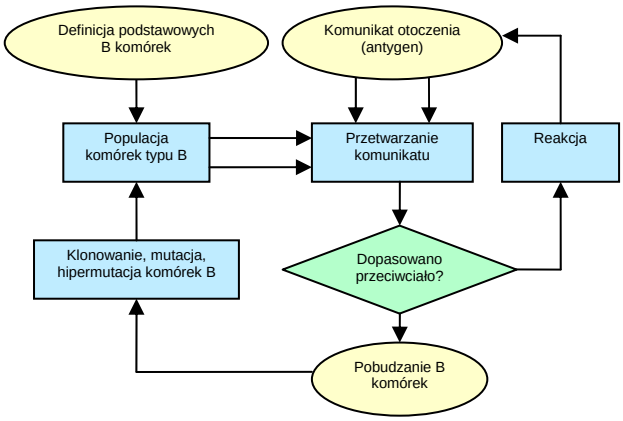


Rys. 14. Wyniki systemu dla zestawu reguł wyposażonego w przykładowe reguły idealne

Mogła się również zdarzyć sytuacja, w której system nie potrafił zareagować na najprostsze komunikaty. Dzieje się tak z powodu charakterystyki algorytmu genetycznego, który stara się wyeliminować najsłabsze i najmniej dokładne reguły. Usunięcie tego problemu nie jest możliwe bez całkowitej zmiany struktury zestawu reguł.

Osiągi systemu wyposażonego w sztuczną sieć immunologiczną

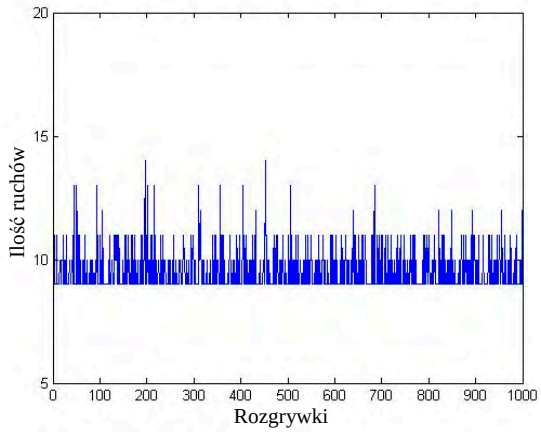
System opierający się o sieć immunologiczną korzysta z wielu komponentów podobnych do wcześniej przedstawionego systemu, tak więc mechanizm przetwarzania reguł nie uległ większym zmianom, podstawowe struktury danych są identyczne. W projekcie sieci immunologicznej konieczne jednak było wprowadzenie kilku istotnych zmian. Najważniejszą było dodanie operatora mutacji podczas tworzenia nowych przeciwciał, ponieważ nie byłoby możliwe stworzenie B komórek pasujących do nowych komunikatów, używając wyłącznie klonowania i hipermutacji.



Rys. 15. Ogólny schemat blokowy systemu testowego opartego o sieć immunologiczną

Charakterystyka sztucznej sieci immunologicznej spowodowała, że dla rozważanego problemu system potrafi wyjątkowo sprawnie i szybko działać, niestety nie jest on w stanie wytworzyć efektywnej strategii jeżeli zaczyna działanie przy minimalnej wiedzy początkowej. Taka sytuacja istnieje z powodu ograniczeń związanych z rozwiązywaniem problemu. Jak wiadomo, w złożonych grach możliwych dróg prowadzących do zwycięstwa jest wiele i każda rozgrywka jest inna, dlatego też w trakcie nauki system musi mieć przeciwnika, ponieważ niemożliwe jest, by mógł być nim człowiek, zastosowano sprzężenie zwrotne i system musi stawiać czoła samemu sobie.

Gdyby rozgrywka przebiegała zawsze identycznie, system w kilka chwil znalazłby sposób na osiągnięcie zwycięstwa. Jednak tak nie jest i o ile w przypadku systemu klasyfikującego możliwe jest, by jedna ze stron nie zwiększała swoich możliwości, ale działała poprawnie, to w sieci immunologicznej nie można tak postąpić, bo założeniem tutaj jest, że sieć ma szukać rozwiązań nowych problemów. Więc zablokowanie uczenia się jednej ze stron zatrzymałoby pracę systemu. Rozwiązaniem jest tutaj znalezienie dla sieci innego przeciwnika, ale jeżeli ten przeciwnik będzie rozpoczynał rozgrywkę z dużą przewagą, to sieć nigdy nie będzie mogła zwyciężyć i co za tym idzie – nie otrzyma nagrody od otoczenia, przez to też nie zwiększy swojej skuteczności.



Rys. 16. Wyniki sieci immunologicznej w rozgrywce z drugą siecią immunologiczną

W sytuacji, gdy jednak pozwolimy sieci uczyć się ze sprzężeniem zwrotnym, to jedna ze stron zawsze będzie wygrywać i będzie to ta strona, która pierwsza wytworzy sobie strategie zwycięstwa.

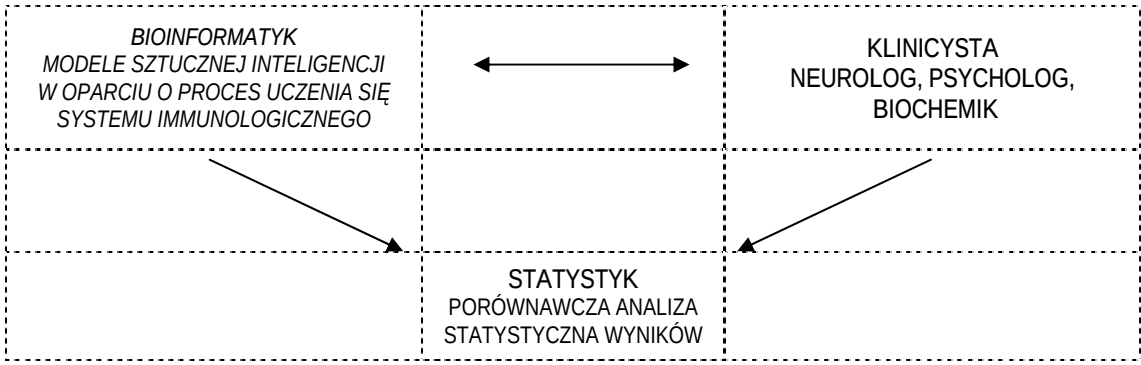
Nie da się zaprzeczyć jednak, że sieć potrafi szybko reagować na nowe sytuacje zaistniałe w środowisku. W przypadku rozgrywki z człowiekiem nie potrafi ona blokować jego działań, ale za to uparcie dąży do uzyskania zwycięstwa i jeśli tylko będzie się próbowało uniemożliwić sieci osiągnięcie przewagi, to prędzej czy później sieć zwycięży.

3. Podsumowanie

Z praktycznego punktu widzenia przedstawione w dokumencie mechanizmy uczące się stanowią potężne narzędzie, które może do wiele różnych zastosowań. Jak można się było przekonać, efekty osiągnięte dzięki tym narzędziom mogą być obiecujące. Dzięki cechom, jakie posiadają sieci immunologiczne, możliwe jest rozwiązanie problemów wymagających wnioskowania w oparciu o zgromadzoną wiedzę. A zdolność do samoorganizacji wiedzy, którą posiadają systemy klasyfikujące, pozwala na rozwiązywanie złożonych problemów poszukiwania. Niewątpliwie połączenie cech obydwu systemów mogłoby dać ciekawe efekty.

Literatura cytowana

1. Cichosz P.: Systemy uczące się, Wydawnictwo Naukowo-Techniczne, Warszawa 2000.
2. Goldberg D. E.: Algorytmy genetyczne i ich zastosowania, Wydawnictwo Naukowo-Techniczne, Warszawa 2003.
3. Wierzchoń S. T.: Sztuczne systemy immunologiczne – teoria i zastosowanie, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2001.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 37-44.

REINFORCEMENT LEARNING ALGORITHM BASED ON PSYCHOLOGICAL MODEL OF AUTONOMOUS SYSTEM

ALGORYTM UCZENIE ZE WZMOCNIENIEM BAZUJĄCY NA PSYCHOLOGICZNYM MODELU SYSTEMU AUTONOMICZNEGO

MARIUSZ ŚWIĘCICKI[*], PIOTR WAIS[**], WIESŁAW WAJS[***]

[*]*Instytut Modelowania Komputerowego, Politechnika Krakowska, Kraków, mswiecic@pk.edu.pl*
[**]*Instytut Politechniczny, Państwowa Wyższa Szkoła Zawodowa w Krośnie, Krosno, meadow@poczta.onet.pl*
[***]*Katedra Automatyki Akademii Górniczo-Hutniczej, Kraków, wwa@ia.agh.edu.pl*

Abstract. In this paper we describe the mechanism of a learning system, which is based on psychological model of autonomous system. The control autonomous agent has recently been the subject of intense research. A variety of approaches from engineering and artificial intelligence have been used to study different aspects of robot control. Many of these efforts are aimed at the developing of simple intelligent behavior. The paper presents autonomous agent as mobile robot controllers based on conception of autonomous agent as a cybernetic model of an autonomous system. It is based on psychological and biological phenomena which may be observed in living organisms. Our autonomous system is characterized by robust and reliable self-adaptation to the characteristic of the environment without external supervision or control. In this paper we describe implementation of an autonomous system to control mobile robots and preliminary results of experiment.

Key words: reinforcement learning, correlator, effectors, alimentators, homeostat, the state of functional equilibrium of the autonomous

Streszczenie. W pracy tej model autonomicznego agenta przedstawiono jako cybernetyczny model systemu autonomicznego. Model ten był wzorowany na zjawiskach psychicznych oraz biologicznych, które występują w żywych organizmach. W opisie systemu zostały wyodrębnione trzy podstawowe bloki funkcjonalne – korelator, homeostat, akumulator – wraz z dokładnym omówieniem ich funkcji. Na proces uczenia tego typu modelu istotny wpływ mają interakcje pomiędzy torem informacyjnym i torem energetycznym, co zostało pokazane w tej pracy. W tym artykule została przedstawiona i omówiona przykładowa realizacja systemu.

Słowa kluczowe: uczenie ze wzmocnieniem, korelator, efektor, alimentator, homeostat, stan równowagi funkcjonalnej systemu autonomicznego

1. Introduction

Autonomous biological agents are characterized by robust and reliable self-adaptation to characteristic of the environment without external supervision or control. This adaptation process takes place while agents operate in own environment. In several real world situations it would be desirable to employ robots that have some features of autonomous system, i.e. are capable of developing new behaviors or adapting existing strategies according to the unpredictable characteristics of real world.

As a reaction to the partial failure of classical AI approach to develop robust control systems for robots that need to operate autonomously in real situations. Whereas classic AI is more concerned with a high level definition of the environment and of the knowledge required by the system, behavior-based robotics stresses the importance of continuous interaction between the robot and its own environment for the dynamic development of the control system.

The control of mobile robots has recently been the subject of intense research. A variety of approaches from the fields of engineering and artificial intelligence have been used to study different aspects of robot control and navigation. Many of these efforts are aimed at the development of simple intelligent behavior, especially when the robot is operating in an unknown environment. In recent years, we have witnessed an increased interest in the application of neural networks and genetic algorithms to control mobile robots.

In the present article we will present a model of an autonomous agent based on the theory of the autonomous system as elaborated by Professor M. Mazur [2], [3]. Subsequently, we will proceed to a description of the implementation of individual blocks of the system, which were implemented by applying neuron networks. Using the process of finding

a solution for the problem of homing navigation with the application of artificial field signs as an example, we will show that a system of this type can constitute an element of equipment of mobile robots responsible for their autonomy.

2. Idea of operation of autonomous system

This chapter presents a model of an autonomous system related to the autonomous agent discussed in a paper by Prof. Marian [2], [3]. Consequently separate elements of this autonomous system as well as their interactions will be discussed.

The notion of an agent should be identified either with a physical being or with an abstract model. In the case of a physical being, the agent could be understood as a robot which has its own list of tasks to be performed. It should subordinate the implementation of the tasks to the environment, in which it exists as well as to the internal state. This internal state can be combined with power resources (e.g. battery condition) of the robot, which are necessary for active existence in this environment. In the case of an abstract being, we have in mind a program which should play a role of an expert system. A controllable system which is able to counteract the loss of controllability, can be called an autonomous system [2], [3], [5], [6]. Such a system should be equipped with following elements:

- Effectors, i.e. elements used to exert an influence on the environment. In the case of a robot these can be its manipulators. When considering the expert system – it is an answer to the question asked. The effectors should be provided with an information, whose possible influence on the environment should occur within a given period. In order to fulfill this condition, the system should include an information path which consists of following functional blocks: a correlator and receptors.
- Receptors are the elements used for taking information from the environment.
- Correlator is a sub-system responsible for data processing and storing. The presented system can be compared to organisms living in the natural environment. An equivalent of the correlator block would probably be the cognitive system block, responsible for the learning process in an organism.

The effectors should also be supplied with energy required for execution of labor necessary to exert the influence on the environment. Therefore another path – the so called power path – is required; it consists of alimentators and an accumulator.

- Alimentators are segments used for taking energy from the environment. In the case of a robot, the role of this element is rather obvious, while in the case of an expert system, it needs requires an explanation. As already stated, as far as an expert system is concerned, an abstract being is taken into consideration. Therefore it can be assumed that no power is required for its existence. When discussing the expert system, the notion the "power" or "energy" could be interpreted together with information of more elementary character in relation to the information received by the receptors of the system.
- Accumulator block is responsible for keeping and storing the energy.

The autonomous system should be equipped with a block which provides a coupling between the information path (receptors and a correlator) and the power path (an alimentator, an accumulator) of the system. When searching for close analogies in natural environment, such a block can be compared with a rewarding system which exists in all the living organisms. This rewarding system is responsible for distinguishing what is "right" and what is "wrong" for a given organism. The operation of this system can be manifested e.g. in finding pleasure (an awarded prize) or feeling pain (a penalty). Such a block will be called hereinafter a homeostat and it should perform following tasks [2], [3], [5], [6]:

- Homeostat is an element used to counteract the flow of information and energy, which reduces the ability to exert the influence of the system on the environment. This element is responsible for maintaining the state of functional equilibrium of the autonomous system [2], [3], [5], [6].

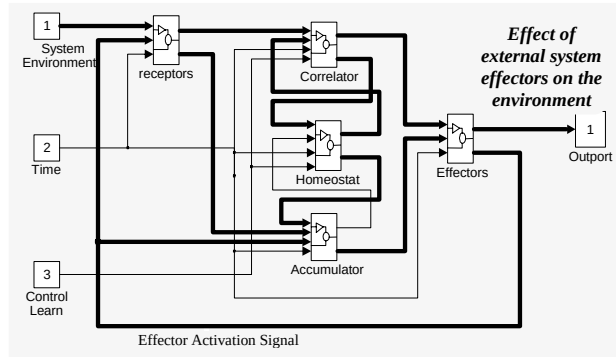


Fig. 1. Model of the autonomous system by M. Mazur

A visualization of such type of an autonomous system is presented in Fig. 1. The receptors, alimentators and effectors are functional blocks providing the contact of the system with its environment. The receptors and alimentators are system inputs, while the effectors are system outputs. The correlator, accumulator and homeostat are internal organs of the autonomous system. Gaining information from the environment consists in distinguishing the differences [2], [3], [5], [6] in action of individual elements of the environment on the receptors. In order to detect a defined action, a special receptor sensible for this particular type of action and neutral to otwers is required. A defined type of influence to be exerted on the environment requires a special effector. The greater the number of effectors the broader is the possible interaction between the system and its environment. As the information is saved in the correlator and the energy is kept in the accumulator, they can be used independently once these data had been gathered.

3. Homeostat

The self-control of the autonomous system in the environment is based upon the feedback. As a result, the changes occur not only in the environment but also in the autonomous system itself. The changes introduced into the autonomous system are limited to values that do not cause the loss of system controllability. A danger in loosing the controllability is

minimal if the state of the autonomous system is far away fromeither boundary. Such a state is called ***functional equilibrium*** [2], [3], [5], [6]. The main task of the homeostat is to keep the autonomous system in this state.

The greater the change occurring in the environment the more intensive will be its influence on the autonomous system via the receptors and alimentators. Similarly the more intensive will be the action of the accumulator and the correlator on the homeostat. Thus the disturbance of functional equilibrium results indirectly from the changes of the system environment. The homeostat has no contact with the environment. It means twat therefore everything twat happens at its input should be treated as its disturbance. The disturbance can be eliminated by the homeostat by means of its direct influence exerted on the accumulator and the correlator. For example if the correlator's and accumulator's action on the homeostat is increasing, then the counteraction of the homeostat on the correlator and the accumulator shall be reduced (and conversely). As a result of such actions, the functional equilibrium of the system is restored. It follows therefore that a negative feedback occurs between the homeostat and correlator as well as between the correlator and homeostat. Taking into consideration that many processes can run simultaneously, the homeostat should be an element consisting of a lot of mutually combined feedback circuits. Generation of positive feedbacks in certain circuits is balanced by negative feedbacks in other ones.

4. Information path of the system

The information path of the autonomous system includes the receptors, a correlator, the effectors as well as the correlator-homeostat feedback. As it follows from the scheme presented above, the correlator is equipped with two groups of inputs and two groups of outputs:

1. A receptor signal (Vr) is entered by a receptor into the correlator [2], [3], [5], [6],
2. An effector signal (Ve) is entered by the correlator into an [2], [3], [5], [6],
3. A perturbation signal (Vp) is entered by the correlator into the homeostat [2], [3], [5], [6],
4. A homeostatic signal (Vh) is entered by the homeostat into the correlator [2], [3], [5], [6].

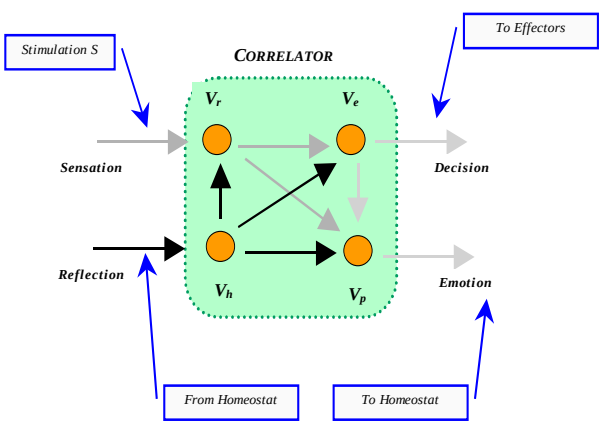


Fig. 2. General chart of realations ocurring

Regardless of the level of its complication, each correlator should perform following functions:

A. Gaining information,
B. Storing information,
C. Processing the information stored.

The first two functions are rather obvious, however the third one needs an explanation of the point of data processing in the correlator. Creation of an interaction between the correlation elements is called association [2], [3], [5], [6] of these elements and – indirectly – of the values of signals appearing at these elements. Taking into consideration all four basic types of correlator signals, following eight types of associations occurring in the correlator can be distinguished:

Input – input associations

1. If a sensation (Vr) of a stimulus is sufficient to generate a decision and a reaction then an association with another (V`r) stimulus may appear sufficient to generate the same decision as well. As an example a classic experiment carried out be Pavlov, as described in next chapter can be referred to.
2. If a sensation (Vr) of a stimulus is sufficient to generate a decision and a reaction then also the reflexion (Vh) leading to the imagination of such stimulus may appear sufficient to generate the same decision and the same reaction as (Vr).

Input – output association

3. A sensation (Vr) of a stimulus causes an associated decision (Ve) and reaction.
4. A reflexion (Vh) generates an associated decision (Ve) and reaction.
5. A sensation (Vr) of a stimulus causes an associated emotion (Vp).
6. A reflexion (Vh) generates an associated emotion (Vp).

Output – output association

7. If the sensation of a stimulus is sufficient to generate a decision (Ve) and a reaction then it may also be sufficient to generate an associated decision (V`e).
8. If the sensation of a stimulus is sufficient to generate a decision (Ve) and a reaction then it may also be sufficient to generate an associated emotion (Vp).

The Fig. 3 presents all the possible association types which can take place in the correlator. As a recapitulation of the above presented discussion on processing the information in the correlator, it could be stated that – generally speaking – the information processing consists in interactions between inputs and outputs of the system.

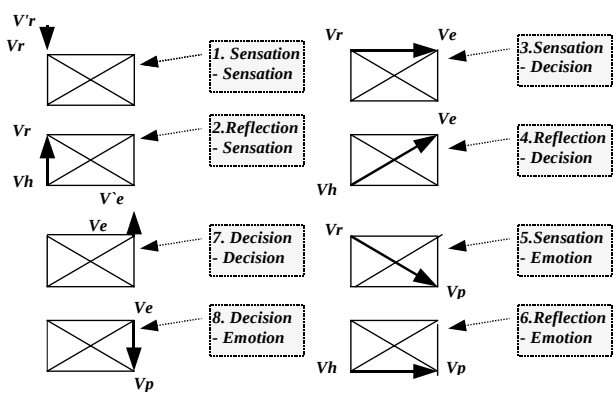


Fig. 3. Types of association

As it is visible, in the case of a correlator containing one receptor and one effector only, there exist eight possible types of associations which give dozens possible combinations of several associations occurring simultaneously. Furthermore, taking into consideration greater number of correlation elements, i.e. receptors and effectors, respectively, the total number of possible associations might be very large. Following terminology defining the relations between correlator's input and output signals has been introduced to describe this model:

- *Emotion* is an action of the correlator on the homeostat [2], [3], [5], [6]. In particular two basic cases of emotion can be distinguished:
  1. Aversion, i.e. emotion consisting in an amplification of the disturbing signal (Vp) [2], [3], [5], [6],
  2. Attraction, i.e. emotion consisting in a reduction of the disturbing signal (Vp) [2], [3], [5], [6],
- *Reflexion* is an action of the homeostat on the correlator [2], [3], [5], [6]; following options occur:
  1. Disapproval, i.e. reflexion consisting in reduction of the homeostatic signal (Vh) [2], [3], [5], [6],
  2. Approval, i.e. reflexion consisting in amplification of the homeostatic signal (Vh) [2], [3], [5], [6].

5. Interactions between the power path and information path of the system

After presentation of the information and power paths as well the interactions occurring within them, we are able to discuss the influence of information processes on power ones as well as the influence of the power processes on information ones, i.e. in general we are going to discuss the behavior of the autonomous system. The fact, that a homeostat is a common organ for both these areas causes following consequences. First of all the information processes exert an effect on power processes through the homeostat and also the power processes influence on the information processes (Fig. 1). In particular the following cases can be distinguished [2], [3], [5], [6]:

1. The reflexion depends not only on emotion, but also on the signal put into the homeostat by the accumulator, i.e. on the load;
2. The stress depends not only on the load but also on the information processes run in the correlator and in particular – on the perturbation signal (Vp), i.e. on emotion;
3. The emotion exerts an influence not only on reflexion but also on the stress;
4. The load exerts an influence not only on the stress but also on the reflexion.

The mutual interaction of information and power processes is determined by the rules of homeostat's cooperation with the correlator and the accumulator.

1. The homeostat reverses the action. If the input potential of the homeostat increases, then its output potential decreases. If the input potential of the homeostat decreases, then its output potential increases.
2. The correlator and the accumulator transfer the actions. If the input potential increases then also the output potential shall increase.
3. The homeostat influences on the correlator and the accumulator in the same way, i.e. it introduces the increasing or decreasing signals to both these blocks.

Basing upon the presented description of interactions, following basic four types of these interactions between the information and power system can be distinguished [2], [3], [5], [6]:

1. Attraction (inf.) causes approval (inf.) and acceleration of decision (inf.) as well as compression (en.) and intensification of the reaction (en.).
2. Aversion (inf.) causes disapproval (inf.) and postponement of decision (inf.) as well as decompression (en.) and weakening of the reaction (en.).
3. Decompression (en.) causes compression (en.) and intensification of the reaction (en.) as well as approval (inf.) and acceleration of decision (inf.).
4. Overload (en.) causes decompression (en.) and weakening of the reaction (en.) as well as disapproval (inf.) and postponement of decision (inf.).

6. Experimental

In the example, the results of which will be discussed hereunder, the fundamental objective of the autonomous robot consists in finding a way to a point with the greatest concentration of 'energy' (i.e. a source of light or sound), this to be detected by means of the robot's energy-sensitive receptors. The robot will proceed through environments containing obstacles. The robot should attempt to realise the objective in the shortest time possible and at the same time avoid bumping into any surrounding objects.

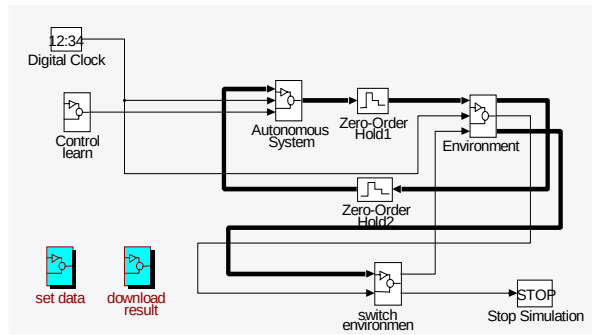


Fig. 4. System used for teaching the autonomous system

During the execution of these tasks, the agent will learn in real time the correctness of behaviour required in its respective environments. Firstly, therefore, it must identify the section of the environment containing the 'energy' source; secondly, it must learn how to identify objects present in the environment, while thirdly it must familiarise itself with the 'characteristics' of bumping into these objects.

The entire robot teaching process will take place in real time. In

Fig. 5 we have presented the environments which will be used for teaching the autonomous system. The system will have to execute the same tasks in both environments, which differ only as regards the arrangement of objects that the agent will have to avoid when realising its objectives. Therein, the system will be taught alternately, i.e. once it has executed its task in *environment 1*, it will proceed to that set in *environment 2*, while subsequently it will return to *environment 1*. At the end of the teaching process we will check whether the

robot will be able to generalise experience gathered during learning for a different environment. The environment within which we will examine the ability of the robot to generalise experienced gathered by the system has been presented in Fig. 6. Figure 7 in turn, shows the implementation of the environments of the autonomous system.

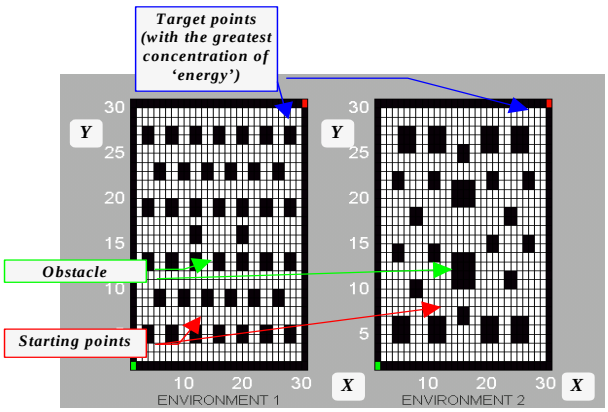


Fig. 5. Environments used for teaching the system (Environment 1, Environment 2)

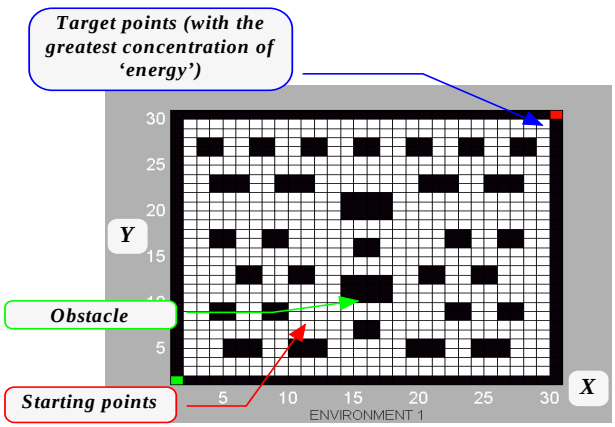


Fig. 6. Environment used to verify the ability of the system to generalise

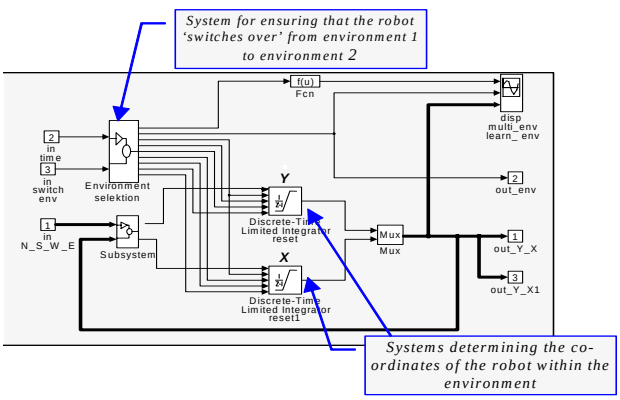


Fig. 7. Implementation of the environment of the autonomous system

Evaluation criteria

It has been mentioned that the objective of the agent consists in locating the point in the environment which has the greatest concentration of 'power' and in proceeding thereto so as to avoid objects present in the environment. We will evaluate the process of learning applying the following criteria.

1. The first will be the time necessary for the autonomous robot to reach the target point in a given environment; we will call this the task execution time and designate it hereinafter with the symbol T_{rz}.
2. The second criterion applied in order to evaluate the quality of the robot's learning process will inform us whether the robot has learned (a) to identify correctly objects present in the environment and (b) the appropriate reactions connected with such objects (i.e. whether it has learned to avoid them). We will define this criterion as follows:

$Error_{Bumps}$ = *(Number of bumps)/(Number of movements) [s]*

where:

Number of bumps – the number of times the robot bumps into objects present in the environment, counted from the initial point up to the attainment of the target point in a given environment by the system.

Number of movements – the number of movements which the system had to execute in order to proceed from the starting point to the target point in a given environment.

Example 1

As was mentioned earlier, the objective of the robot consisted in reaching the point in a given environment with the greatest concentration of energy. Figure 7 shows a diagram presenting the values of $Error_{Bump}$ and T_{rz}, while Fig. 9 shows the initial and final trajectories of the robot for environment 1.

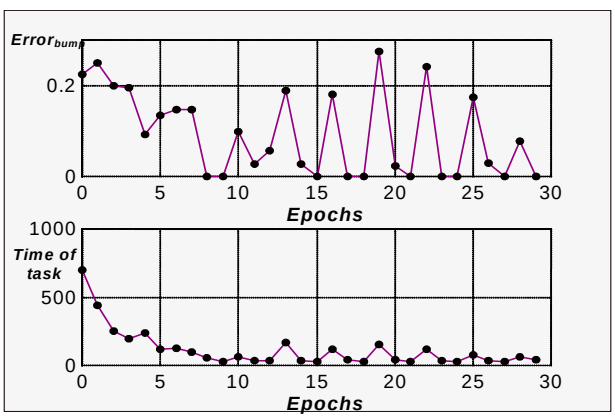


Fig. 8. Chart of $Error_{Bump}$ and T_{rz} (task execution time) values

Looking at Fig. 9, we can observe that the final trajectory of the robot is more regular in comparison with its initial trajectory. This indicates that the system has learned how to identify correctly the point of the environment with the greatest concentration of energy, and also that in the course of the teaching process it attempted to optimise the path leading

thereto. This is brought out by the fall in the value of T_{rz} (task execution time) as shown in Figure 8.

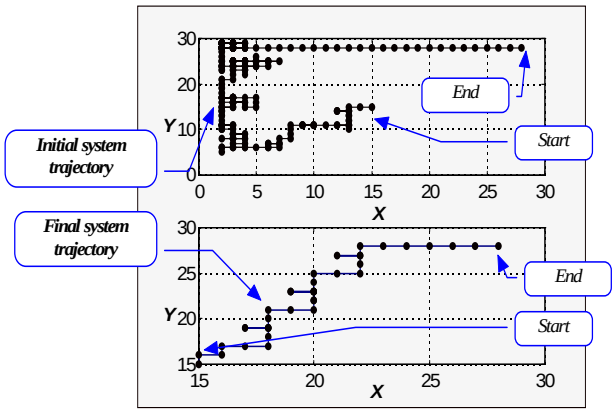


Fig. 9 Trajectories of the robot for environment 1

The same phenomenon may be observed in the case of environment 2 (cf. Fig. 10). Again, the final trajectory of the robot is regular in comparison with its initial trajectory, and the path leading to the fixed point of the environment has been optimised; this is reflected directly by the task execution time. During the teaching process, the execution time fell from an initial value of approximately 750 [s] to 80 [s].

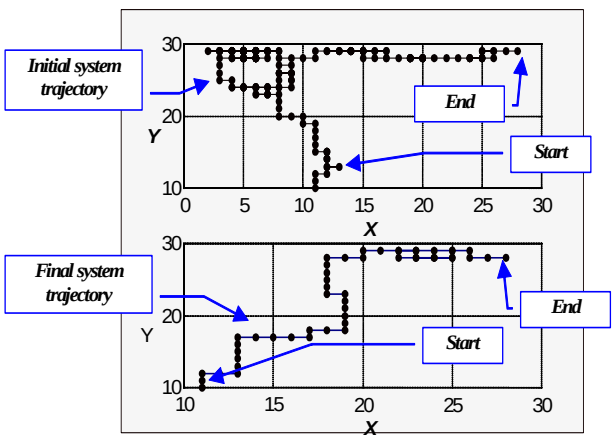


Fig. 10. Trajectories of the robot for environment 2

However, despite the fact that the system optimised the path to fixed point, the value of $Error_{Bump}$ remained relatively high and oscillations were present. As a result of the teaching process, the final error value had decreased only by a half in comparison with the initial error value. This clearly shows that the system does not always identify correctly objects present in the environment, and thus bumps into some of them. This is confirmed by the shape of the final system trajectory, which at certain points is made up of a number of parallel lines (Fig.9, Fig.10).

Recapitulating the results of this part of the experiment, we may state that the system has to a considerable degree learned how to react – with greater or lesser correctness – to objects, and also managed to optimise the path to the target point of an environment, but the knowledge gained thereby about objects in one environment is not adequate to be generalised for objects present in the second environment.

Example 2

In this case the rapidity of the system teaching process has been decreased by lowering the value of the teaching coefficient for all system neuron networks. The results of this experiment have been presented in Fig. 11, Fig. 12, and Fig. 13. As we can see, in this case the value of $Error_{Bump}$ at the end of the teaching process has reached zero, which fact means that the system has learned how to identify faultlessly objects present both in environment 1 and in environment 2.

The zero error value has led to the regular shape of the final trajectories for individual environments. Furthermore, we may observe that the time taken by the robot to reach the fixed point in individual environments has attained a minimal value, and also that towards the end of the teaching process the path of the system has been shortened in comparison with the initial path. We can observe this phenomenon for both environment 1 and environment 2.

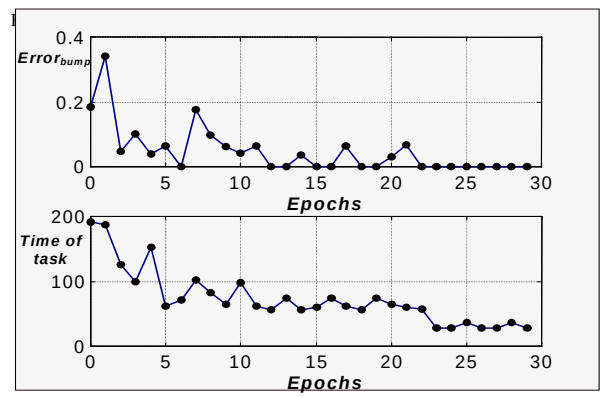


Fig. 11. Chart of $Error_{Bump}$ and T_{rz} (task execution time) values

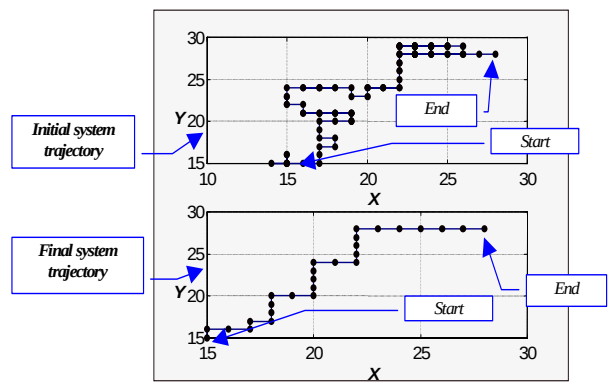


Fig. 12. Trajectories of the robot for environment 1

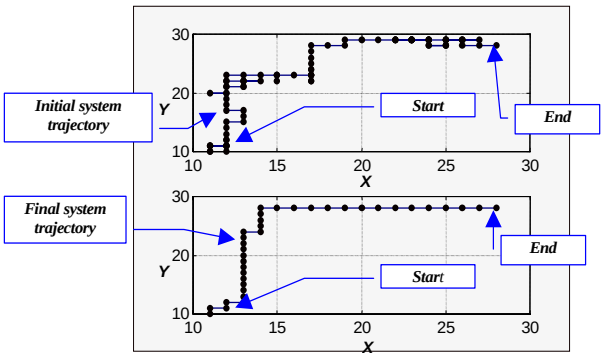


Fig. 13. Trajectories of the robot for environment 2

Finally, we will proceed to the last phase of the experiment, which will consist in examining whether the robot is able to generalise experience gathered in the course of the teaching process conducted in

Fig. 5 environments for a hitherto unvisited environment (Fig. 6).

Example 3

At this point the robot that was subjected to the teaching process in

Fig. 5 environments will be placed in the Fig. 6 environment and, as was the case previously, we will thereby attempt to assess whether the robot will be able to generalise its experience for objects present in the Fig. 6 environment.

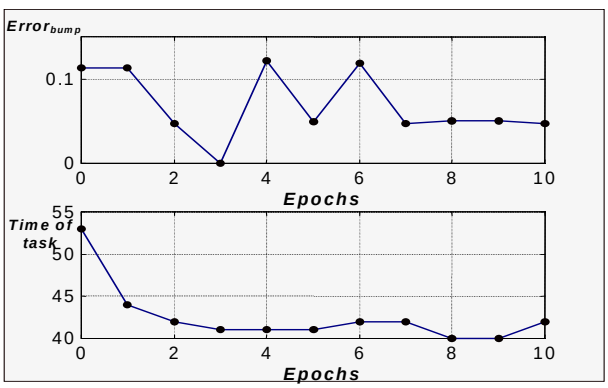


Fig. 14. Chart of $Error_{Bump}$ and T_{rz} (task execution time) values

Figure 14 shows that during the first run the $Error_{Bump}$ value increased in comparison with its final value in Fig. 11. This means that at the beginning of the robot's 'existence' in this environment there appeared problems with the identification of objects. However, due to the fact that the system teaching process was conducted in real on-line time, the system subsequently learnt how to correctly react to objects present in its new environment. As previously, the final trajectory of the robot was optimised in terms of length in comparison to the initial path of the robot in this environment, which in turn caused the T_{rz} (task execution time) to decrease.

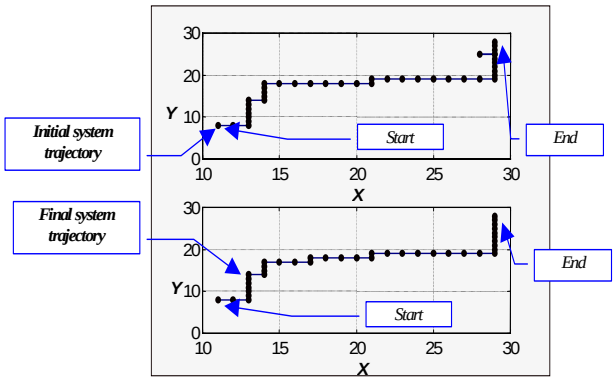


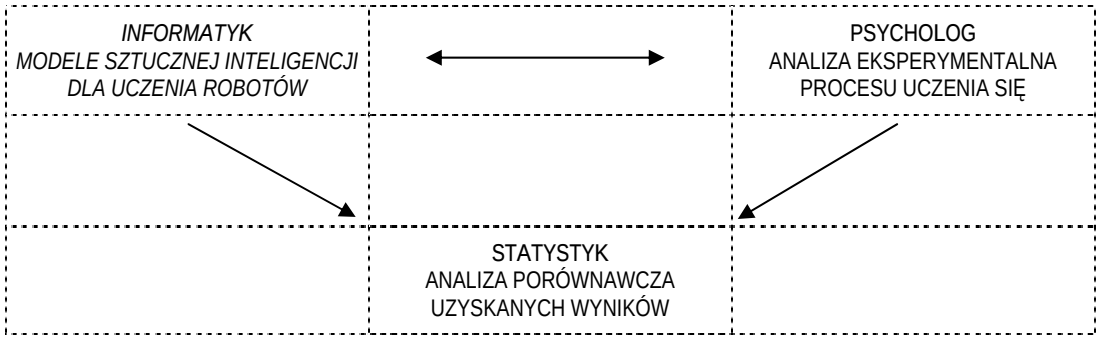
Fig. 15 Trajectories of the robot for the Fig. 6 environment

7. Conclusion

In the present article we have presented a model of an autonomous system that may be made responsible for autonomy in an actually existing mobile robot. In addition, we have presented the fundamental subsystems of the model, and also an exemplary implementation of individual subsystems. It has been shown that with the application of artificial field signs a system of this type is able to find independently a solution for the homing navigation problem, which forms one of the key problems in the construction of mobile robots, because the entire system teaching process takes place in real time and without any external control. From a practical point of view, it is also of interest that a system of this type can to a certain degree generalise experience gathered in one environment for other environments of the same class.

References

1. Maruichi T, Tokoro M.: „Modeling Autonomous Agents and Their Groups". In Demazeu Y. (Ed.) Decentralized AI. Proceedings of The First European Workshop on Modelling Autonomous Agents in a Multi-Agent World, Cambridge, 1989 England.
2. Mazur M.: Cybernetyczna teoria układów samodzielnych, Państwowe Wydawnictwo Naukowe 1966.
3. Mazur M.: Cybernetyka i charakter, AULA 1996.
4. Tadeusiewicz R: Sieci Neuronowe, Akademicka Oficyna Wydawnicza RM 1993.
5. Wajs W, Święcicki M.: Neuron model of an autonomous agent, Proceedings of the Third Conference Neural Networks and Their applications and Summer School on Neural Networks Applications to Signal Processing, KULE 97, Poland.
6. Wajs W, Święcicki M.: Neural Network Model of Autonomous Agent for Decision Support System, International Congress on Modeiling and Simulation MODSIM 97, Australia.
7. Wajs W, Święcicki M.: Static Neural Network Based Self Learning Autonomous System. Pre-proceedings of the First International Conference on Philosophy and Computer Science, Kraków 1998, Poland.


INFORMATYK
MODELE SZTUCZNEJ INTELIGENCJI
DLA UCZENIA ROBOTÓW
PSYCHOLOG
ANALIZA EKSPERYMENTALNA
PROCESU UCZENIA SIĘ
STATYSTYK
ANALIZA PORÓWNAWCZA
UZYSKANYCH WYNIKÓW

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 45-50.

ZASTOSOWANIE STANDARDU HL7 DLA AUTOMATYCZNEGO POZYSKIWANIA WIEDZY I PRZETWARZANIA JEJ W DRODZE ANALIZY ANALOGII PRZYCZYNOWO-SKUTKOWYCH

USING HL7 STANDARD IN AUTOMATED KNOWLEDGE MINING AND CAUSE-RESULT ANALOGIES ANALYSIS

MICHAŁ TUREK

Katedra Automatyki Akademii Górniczo-Hutniczej, Kraków,
al. Mickiewicza 30, 30-059 Kraków, mitu@agh.edu.pl

Streszczenie: Nie ulega wątpliwości, iż w rozproszonych systemach przetwarzania wiedzy medycznej kluczowym elementem jest sprawna i uniwersalna komunikacja. Powstało już wiele standardów wymiany danych medycznych oferujących stosy protokołów zaadaptowanych do najróżniejszych celów. Spora cześć to standardy uniwersalne, operujące w wielu dziedzinach działalności medycznej jednocześnie – od diagnostyki, przez zaopatrzenie ośrodków po kwestie rozliczeń usług medycznych czy ubezpieczeń. Rzadko jednak spotyka się standardy przewidujące w swoich założeniach obsługę systemów przetwarzania wiedzy zdefiniowanej formalnie – czy to na poziomie migracji posiadanej przez system wiedzy, czy też udostępniania usług takiego systemu. Istnieje jednak możliwość rozwijania istniejących już opracowań o podprotokoły wspierające taką komunikację. Dobrą podstawą do tych modyfikacji jest standard Health Level 7 (HL7) [2], stosowany już obecnie w wielu krajach jako obowiązujący standard komunikacji w systemach medycznych.

Artykuł ma na celu prezentację metody zastosowania standardu komunikacji systemów medycznych HL7 do wsparcia komunikacji w rozproszonym systemie pozyskiwania wiedzy typu health-care. Jednocześnie proponuje technikę przetwarzania możliwej do pobrania za pośrednictwem HL7 wiedzy. Celem jest zaproponowanie rozwiązania umożliwiającego wykrywanie analogii przyczynowo-skutkowych pomiędzy przypadkami klinicznymi (health-care). Analogie te postrzegane będą na podstawie dostarczanego przez otoczenie systemu zdarzeń. Na bazie pobieranych informacji metoda będzie zmierzać do proponowania decyzji lub ustalania przyczyn analogicznych do zaobserwowanych wcześniej sytuacji. Zdefiniowany zostanie schemat kodowania wiedzy, dopuszczający możliwość automatycznej rozbudowy bazy wiedzy w oparciu o postrzegane zdarzenia. Proponowana struktura będzie także rozpoznawalna dla eksperta dziedzinowego i łatwa do prezentacji graficznej, co umożliwi dalszą rozbudowę i wykorzystanie w ekspertyzach jednolitej bazy wiedzy pochodzącej zarówno z procesu samoistnego uczenia się (automatyczna analiza przypadków), jak i zdefiniowanej przez eksperta w dziedzinie.

Przedstawiane rozwiązanie zakłada, że postrzegane zdarzenia dotyczące zarówno pomiarów, jak i podejmowanych przez aktora decyzji są statystycznie poprawne. Taka właśnie informacja jest naturalnie produkowana w wyniku funkcjonowania rzeczywistych instytucji, w których podejmuje się określone czynności bazujące na wcześniejszych pomiarach. Problemem jest automatyzowanie procedur doradczych i decyzyjnych realizowanych przez czynnik ludzki, które mogłoby polegać na uogólnianiu zarejestrowanych sytuacji i decyzji faktycznie kiedyś już podjętych w analogicznych sytuacjach. Dzięki analizie analogii przypadków występujących masowo rozwiązanie będzie w stanie zniwelować niewielki odsetek informacji błędnej, nie prezentując jej w wynikach późniejszych ekspertyz. Prostym przykładem zastosowania omawianego rozwiązania będzie wspomaganie decyzji podejmowanych wobec określonych obiektów na podstawie historii zdarzeń oraz stanów tych obiektów. Metoda umożliwia śledzenie serii zdarzeń, kodowanie ich w postaci reguł zestawiających te zdarzenia oraz analizę wpływu danych czynności na zmiany stanów obiektów. Obiektami takimi mogą być pacjenci kliniki medycznej powiązani ze zdarzeniami typu:

- diagnostyka (określanie wartości danego parametru zdrowotnego w danej chwili),
- wykonywanie zabiegów, aplikacja leków (wywołane w wyniku prawidłowych decyzji czynnika ludzkiego).

Najlepsze wyniki metoda osiągnie w analizie obiektów występujących masowo posiadających zadaną pulę określonych skalarnie właściwości i stanów, oraz przechodzących w skutek odpowiednich czynności w nowe stany wynikowe.

Słowa kluczowe: HL7, health-care, wspomaganie decyzji, inżynieria wiedzy

Abstract: Some useful knowledge about object behavior can be stored just in a set of object's attribute values and actions performed on it. This paper shows a way of tracing connections between some registered actions and results of these actions at affected objects and using it to predict a factor correlation on other objects. Proposed solution analyses incoming information stored in simple facts, consisting of object name, timestamp, factor name, and factor amount only. A special knowledge format, similar to rule-based knowledge structure is defined. Knowledge usage process, thanks to special knowledge format, will be able to use the data in two ways. They are very similar in implementation, but quite different in application. The first will produce decision support around the selected object and situation. The second will try to explain reasons of the selected object condition. Discussed solution will use a popular HL7 standard for knowledge migration and service providing communication.

Key words: HL7, health-care, decision support systems, knowledge engineering

1. HL7 w zastosowaniu do przekazu wiedzy i zapytań health-care

HL7 standaryzuje struktury komunikatów, reguły kodowania wiadomości do transmisji oraz zdarzenia wyzwalające przesłanie komunikatu. Komunikat HL7 stanowi jednostkę danych przesyłaną pomiędzy elementami rozproszonego systemu informatycznego. Jest on złożony z grupy segmentów upakowanych w określonej kolejności. Każdy komunikat zawiera nagłówek specyfikujący typ wiadomości, a co za tym idzie jej przeznaczenie. Na przykład komunikaty ORM używane są do przesyłania zleceń, a komunikaty ORU zawierają wyniki obserwacji. Segment jest logiczną grupą pól, z których każde zależnie od systemu może być opcjonalne (niewypełnione) lub obowiązkowe. Format pól segmentu może być charakteryzowany przez typ standardowy albo wskazywać kolejne komponenty lub kolekcje komponentów zawarte wewnątrz segmentu. Każdy segment jest jednoznacznie identyfikowany trzyliterowym oznaczeniem zwanym identyfikatorem segmentu.

Nawiązując do problemu health-care konieczne będzie zanalizowanie struktury zawężonej puli komunikatów HL7, w szczególności dotyczących informacji o pacjentach. Śledząc nagłówki prowadzące do kolejnych wewnętrznych segmentów komunikatu, w kręgu zainteresowań będą kolejno leżeć następujące identyfikatory treści HL7 [5]:

MSH jest to nagłówek transportowy komunikatu, definiujący jednostkę źródłową i docelową dla komunikatu. Funkcjonalność HL7 zakłada także transport informacji o migracji pacjentów [1], a jednostka komunikująca się na bazie HL7 w rozproszonym systemie jest przypisana do konkretnego fizycznego miejsca (np. kliniki). Ponieważ przetwarza ona informacje dotyczące lokalnego pacjenta, możliwe jest śledzenie lokalizacji pacjentów, o których wiedza będzie wymieniana – przy użyciu dalszych segmentów HL7. Tym samym, analizując powyższe pola jesteśmy w stanie śledzić adresy źródłowe jednostek, w których znajdują się pacjenci leżący w kręgu zainteresowań systemu pozyskującego wiedzę i na wstępie eliminować komunikaty nieistotne dla tego systemu. W innym przypadku, jeśli przykładowo przetwarzamy w bazie wiedzy informacje o wynikach diagnostycznych pacjentów, można zawęzić śledzenie komunikatów HL7 do tych, które pochodzą tylko od jednostek przeprowadzających konkretne badania analityczne.

Kolejnym segmentem ulokowanym wewnątrz pakietu MSH jest identyfikowany nagłówkiem PID segment, zawierający ogólne dane identyfikacyjne i demograficzne pacjenta. Konieczność stosowania danych tego nagłówka jest oczywista – gwarantuje podstawową identyfikację pacjenta, czy to w przypadku migracji informacji o pacjencie czy wymiany zapytań i odpowiedzi na zapytania pomiędzy składnikami projektowanego systemu. Informacje zawarte w PID rozbudowuje segment PV1, przenoszący dane dotyczące hospitalizacji pacjenta.

Kolejnymi bardzo istotnymi dla pokrycia funkcjonalności wymaganej przez systemy wspomagania decyzji będą segmenty umożliwiające nawiązanie interakcji koniecznej do obsługi procesu analizy bazy wiedzy z przekazaniem wyników do klienta systemu. Tutaj możliwe będzie adaptowanie gotowego mechanizmu HL7, zapewniającego tzw. transakcje zleceń, a zdefiniowanego wokół segmentów ORC, ORU czy ORM. ORC to segment przygotowawczy zlecenia określonych usług, przenoszący informacje służące do konfigurowania trybu realizacji zlecenia po stronie usługodawcy. ORM to segment faktycznej transakcji zlecenia, definiujący wymogi dotyczące jego realizacji, zapytania o wyniki i inne. ORU to segment wyników przesyłanych do zleceniobiorcy. W segmentach transakcji zlecenia zawarte są kolejne wewnętrzne segmenty, występujące w kolekcjach i definiujące zarówno zapytania (OBR), jak i wyniki obserwacji (OBX). W przypadku zamawiania ekspertyz rekord OBX może być użyty zarówno do przekazania rezultatów ekspertyzy, jak i precyzowania zapytań (przesyłanie tzw. tendencji faktów, opisanych w następnych rozdziałach).

Przykładowe zapytanie o ekspertyzę ORC można zaproponować następująco (poszczególne segmenty i ich pola zaadaptowano zgodnie ze specyfikacją HL7 [3]):

```
MSH|^~\&|Klient_Expertyzy|KE1|Modul_Ekspertyzy
    |KE2|20050115101502||
PID|1|78010112345^PESEL|846123||JAN^KOWALSKI
    ||19811001|M||||||||||||
ORM^O01|EXPERT1012|KE1|2.3|AL||POL|||
ORC|NW|PYTANIE^25123413|
OBR|1|PYTANIE^25123413||12^7231^ DOPRZODU^
    Ob1|||20050115150000|||ANDRZEJ^ LEW^lek^med|||||
OBX|1|FT|Val3||0^30^120^0,6||||
OBX|2|FT|Val2||123^10^14^3,7||||
OBX|3|FT|Val4||80^40^135^1,3||||
```

Gdzie kolejno mamy:

- w MSH: separator dalszych pól (pierwszy znak)| znaki specjalne| nazwa nadawcy| ID procesu nadawcy| nazwa odbiorcy| ID procesu odbiorcy| czas|,
- w PID: ID segmentu PID| ID pacjenta (PESEL)| ID pacjenta w systemie szpitalnym| imiona i nazwisko| data urodzenia| płeć (M – male)|,
- w ORM: typ zdefiniowany na zlecenie (ORM – Order Message)| ID transakcji| ID procesu| wersja HL7 (2.3)| potwierdzenie (AL – zawsze)|| ID kraju|||,
- w ORC: typ zlecenia (NW – nowe)| ID bieżącego zapytania,
- w OBR: ID segmentu OBR| ID bieżącego zapytania (jak w polu ORC:2)|| ID usługi: zapytanie dotyczy podania (proponowanej) konsekwencji sytuacji przedstawionej w rekordach OBX, dane liczbowe parametryzują ekspertyzę (parametry zostały opisane w części dot. Algorytmu ekspertyzy||| spodziewana data zwrotu wyników||| dane zleceniodawcy||| dodatkowe fakty o pacjencie||,
- w OBX: ID rekordu OBX| typ wyniku: FT = tekst formatowany| identyfikator wyniku – ID czynnika, którego zmiany precyzuje tendencja faktu|| dane dotyczące przebiegu zmian czynnika (start^ długość trwania^ wartość początkowa^ współczynnik zmiany)|||. Rekord OBX występuje w kilku egzemplarzach, przekazujących łącznie w zapytaniu tzw. regułę wpływu, definiującą zdanie początkowe dla ekspertyzy.

Zgodna ze specyfikacją HL7 przykładowa odpowiedź w oparciu o powyższy opis jest już łatwa do sformułowania. Strukturalnie będzie się różnić jedynie brakiem segmentu informacji o pacjencie (zapytanie jest etykietowane identyfikatorami sesji) oraz występowaniem innych segmentów OBX. Kilka z przedstawionych pól na potrzeby systemu naturalnie można pozostawić w stanie niewypełnionym (przykładowo, dla wymogów ekspertyz wystarczająca informacja o personaliach pacjenta to jego ID w systemie).

Przedstawiona pula komunikatów umożliwi realizację dialogu związanego z zapytaniami do projektowanego modułu prowadzącego ekspertyzy. Kwestię opisanego w następnych rozdziałach gromadzenia wiedzy na rzecz ekspertyz należałoby zorientować wokół komunikatów dotyczących obsługi zdarzeń (grupa EVT). Komunikaty zdarzeń generowane są

w wielu przypadkach, również wtedy, gdy nastąpiła zmiana stanu pacjenta (czyli obserwowanego przez system obiektu). Użytkowany podtyp komunikatów z puli EVT [4], jak i ewentualne użycie innych komunikatów nie musi być w chwili obecnej jasno ustalone. Ważne jest jedynie, aby komunikaty te przenosiły informację o jednoznacznie identyfikowalnych parametrach śledzonych obiektów wraz z wartościami skalarnymi tych parametrów. W połączeniu ze znacznikami czasu zawartymi w komunikatach enkapsulujących, stanowić one będą wystarczające źródło informacji dla prezentowanego procesu automatycznego pozyskiwania wiedzy.

2. Automatyczne pozyskiwanie wiedzy – czynnik wiodący

Całość procesu kolekcjonowania wiedzy z otoczenia, jak i późniejszego jej wykorzystania będzie przebiegać pod kontrolą jednego wyróżnionego parametru, stanowiącego podstawowy składnik każdej informacji w bazie wiedzy. Parametr ten będzie reprezentowany w postaci skalarnej. Ponadto zbiór wszystkich jego możliwych wartości, jakie w danej chwili występują gdziekolwiek w bazie wiedzy będzie spójny, co daje możliwość dokonywania obliczeń i porównań z jego użyciem. W dalszej części rozważań parametr ten będziemy nazywać *czynnikiem wiodącym*.

Czynnik wiodący w prezentowanym modelu wiedzy służy jako podstawowa jednostka porównawcza, na bazie której określa się związki przyczynowo-skutkowe pomiędzy pobranymi przez system ze środowiska faktami, tworzy się, analizuje i rozbudowuje reguły spinające pobrane fakty. Typ czynnika wiodącego jest ustalany raz dla każdej bazy wiedzy (z chwilą jej tworzenia) i wykorzystywany jest do kodowania dalszej informacji w bazie. W klasycznej postaci związków przyczynowo-skutkowych, dotyczących np. parametrów danego obiektu w środowisku fizycznym, czynnikiem wiodącym będzie czas.

3. Fakty-zdarzenia

Wejściem dla procesu pozyskiwania wiedzy jest informacja pochodząca bezpośrednio z otoczenia, przetworzona tak, aby dotyczyła jednoznacznie identyfikowanych obiektów z tego otoczenia, jednoznacznie identyfikowanych cech lub czynników działających na obiekty, ustalonej wartości takich czynników oraz umiejscowienia tej informacji na skali umożliwiającej wytwarzanie relacji typu przyczyna-skutek z innymi wejściowymi danymi – czyli na skali czynnika wiodącego.

Jednorazową porcją informacji pobieraną z otoczenia jest więc zespolenie w formie faktu informacji powiązanej ze zdarzeniem, które miało miejsce w związku z określonym obiektem. W dalszej części rozważań taką jednostkę będziemy określali terminem *fakt-zdarzenie*. Każdy fakt-zdarzenie posiada przypisaną sobie wartość czynnika wiodącego, na skali którego zdarzenie wystąpiło. Posiada symboliczny identyfikator obiektu, próbkowany parametr obiektu oraz wartość tego parametru.

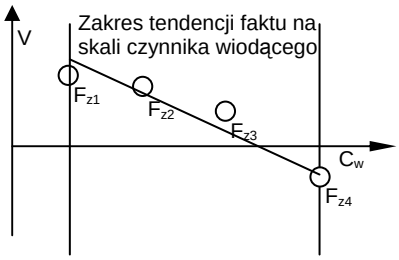
4. Tendencje faktów

Fakty-zdarzenia skoncentrowane wokół danego obiektu mogą stanowić podstawę do ustalenia informacji zbiorczych charakteryzujących dany obiekt. Jeżeli zarejestrowano (w postaci faktów) kilka wartości nazwanych, różniących się wielkością i wartością charakteryzującego je czynnika wiodącego, to możliwe jest zdefiniowanie uogólnienia do pewnej relacji w dziedzinie czynnika wiodącego. Funkcja taka powstanie w wyniku uśrednienia punktowych odczytów wartości nazwanej w faktach-zdarzeniach. Funkcją taką będzie (dla uproszczenia procesu) wyłącznie odwzorowanie liniowe czynnika wiodącego w rozważane wartości, dające się zakodować w tendencji faktu przy pomocy dwóch współczynników numerycznych. Uproszczenie to ma na celu ułatwienie obliczeń na etapie późniejszej analizy tendencji faktów. Kolejną informacją pobieraną z wartości maksymalnych i minimalnych czynnika wiodącego, charakteryzujących zaangażowane fakty-zdarzenia, które dostarczyły informacji. Wartości te są zdefiniowane skalarnie i będą w powstającej konstrukcji stanowiły odpowiednik ograniczenia dziedziny definiowanego odwzorowania (tendencji).

Tak zakodowaną informację o zmianie określonej skojarzonej z obiektem wartości będziemy dalej nazywali *tendencją faktu*. Widać tu, że tendencja faktu będzie zawierała następujące informacje:

- Kierunek zmian i stan początkowy danej wartości charakteryzującej obiekt (np. wartość rosnąca z prędkością zadaną współczynnikiem),
- Wartość czynnika wiodącego dla chwili początkowej zaobserwowanego zjawiska,
- Wartość czynnika wiodącego dla chwili zakończenia obserwacji.

Najbardziej istotną informacją przenoszoną przez tendencję faktu nie jest sama wartość określającego obiekt parametru, a kierunek jej zmian na przestrzeni czynnika wiodącego. Dzięki takiej notacji możliwe jest łatwe zestawianie tendencji faktów i wykrywanie korelacji pomiędzy nimi zarówno w interpretacji jednego obiektu jak i globalnie, na bazie analogii zjawisk występujących w wielu obiektach.



Rys. 1. Tendencja faktu: C_w – czynnik wiodący, V – wartość parametru przedstawianego przez fakty-zdarzenia, F_Z – fakt-zdarzenie

Tab. 1. Tendencja Faktu: T_0, T_1 – wartości czynnika wiodącego dla chwili początkowej i końcowej zaobserwowanego zjawiska

T_0	T_1	OID	VID	V_0	F

Uwagi: OID – symbol obiektu, którego dotyczyły fakty składające się na tendencję faktu, VID – symbol identyfikujący zmieniający się parametr obiektu, V0 – wartość początkowa parametru, F – współczynnik kierunku zmian.

5. Jednolita interpretacja danych wejściowych

Ponieważ celem planowanych ekspertyz na zdefiniowanym formacie wiedzy jest jedynie wykrycie analogii zjawisk występujących wokół wielu rozważanych obiektów (a tym samym związków pomiędzy tymi zjawiskami), możliwe jest ujednolicenie charakteru informacji wejściowej. Ujednolicenie to będzie polegało na jednakowym traktowaniu danych związanych z odczytem stanu obiektów, jaki i danych związanych

z zewnętrznie stymulowanymi akcjami, podejmowanymi wobec obiektu i mającymi wprowadzić – w zamierzeniu eksperta – kolejne zmiany stanu obiektu. Dlatego zarówno odczyty stanów, jak i obserwowane akcje na obiekcie są dla bazy wiedzy takimi samymi faktami-zdarzeniami, kodowanymi za pomocą nazwanej i określonej skalarnie wielkości (np. odczyt skalarnego stanu A obiektu o wartości A1, wywołanie wpływu na obiekt czynnikiem B o sile B1) oraz charakteryzowanymi dodatkowo jedynie określoną wartością czynnika wiodącego.

6. Reguły wpływu

Reguły wpływu konstruowane są na podstawie wiedzy postrzeganej w środowisku zewnętrznym i dotyczą pojedynczego obiektu. Nie stanowią jeszcze podstawy dla ekspertyz. Są tworem pośrednim, stanowiącym źródło dla definiowania i rozbudowy innych reguł, kompatybilnych ze sztucznie zdefiniowaną wiedzą ekspercką i kodujących korelacje zjawisk występujących zbiorczo w wielu obiektach.

Aby możliwe było zdefiniowanie najprostszej reguły wpływu potrzebne są przynajmniej dwie tendencje faktu dotyczące tego samego obiektu. Reguła wpływu nie koduje wartości czynników wiodących tendencji-faktów, które spina a jedynie przesunięcia pomiędzy tymi wartościami nadal wyrażone skalarnie. Jest to wystarczające do zapisania zależności pomiędzy tendencjami faktów na jego skali. Tendencja faktu, cechowana najmniejszą początkową wartością czynnika wiodącego, będzie zerowym punktem odniesienia, względem którego obliczone zostanie kolejne przesunięcie początku następnej, zgodnie z wartością czynnika wiodącego tendencji faktu. Reguła wpływu powstaje na podstawie tendencji faktów dotyczących tego samego obiektu. Poszczególne tendencje faktów są w regule numerowane. Numerację rozpoczyna reguła poruszając się zgodnie z tą numeracją w górę, będzie można poszukiwać skutków (w sensie czynnika wiodącego) danego stanu obiektu, w dół – przyczyn.

Tab. 2. Reguła wpływu (na dole) powstała na bazie kilku tendencji faktów (u góry)

T_0	T_1	OID	VID	V_0	F
80	110	Ob1	Val3	120	0,6
230	245	Ob1	Val7	2,4	1,3
100	130	Ob1	Val1	10	1,25

OID:	Ob1				
No	T_{dist}	T_{len}	VID	V_0	F
1	0	30	Val3	120	0,6
3	150	15	Val7	2,4	1,3
2	20	30	Val1	10	1,25

Uwagi: N_o – numer tendencji, T_{dist} – przesunięcie względem poprzedniej tendencji faktu, [T_{dist} .. T_{len}] – zakres względnej wartości czynnika wiodącego, na której tendencja faktu obowiązuje.

7. Reguły wynikowe – proces wzmacniania reguł

Reguły wynikowe przechowują zbiorczą informację o analogiach zdarzeń rejestrowanych na obiektach. *Wzmocnione* reguły wynikowe są tworzone na bazie reguł wpływu i poprzednich reguł wynikowych w momencie, gdy wobec przynajmniej dwóch różnych obiektów można zaobserwować analogiczną korelacje, tendencji faktów – czyli gdy istnieją analogiczne reguły wpływu na tych obiektach.

Znaczenie słowa „analogiczny" precyzuje szereg warunków i parametrów charakteryzujących proces wzmacniania reguł. Wśród nich można wyróżnić:

- maksymalną dopuszczalną różnicę wartości T_{dist} pomniejszoną o odpowiadające mu T_{len} w przypadku każdej możliwej pary tendencji faktów VID_i i VID_j zawartej w regule wpływu i wzmacnianej regule wynikowej,
- maksymalną dopuszczalną różnicę wartości F – analogicznie pomiędzy wyżej określonymi parami,
- (opcjonalnie) maksymalną dopuszczalną różnicę wartości V_0 (tę wartość można będzie interpretować podczas realizacji dalszych eksperymentów z metodą – nie przedstawia ona danych istotnych dla określania tendencji faktu, jednak w przypadku niektórych baz wiedzy może być istotna).

Wartości tych parametrów bezpośrednio wpływają na proces kształtowania bazy wiedzy o relacjach przyczynowo-skutkowych zachodzących pomiędzy własnościami obiektów. Ustawienie ich w sposób tolerancyjny zminimalizuje wielkość bazy wiedzy, ale doprowadzi do większego uogólniania zależności przyczynowo-skutkowych. Spowodować to może zgubienie podczas ekspertyzy niektórych relacji.

Proces wzmacniania reguł wynikowych może zachodzić nie tylko w przypadku pojawienia się nowej reguły wpływu, lecz także przy korekcie już istniejącej reguły wpływu. W przypadku, gdy reguła wpływu została rozbudowana o nowe tendencje faktów, a nowa tendencja faktu nie znajduje swojego analogicznego odpowiednika w regule wynikowej, reguła wpływu może zostać także wycofana z reguły wynikowej. Następuje wówczas wycofanie wzmocnienia (*osłabienie*) reguły, zaś nie pasująca reguła wpływu jest ponownie rozpatrywana jako źródło dla wzmocnienia lub utworzenia innych reguł wynikowych.

Podstawową decyzją wypływającą z parametryzowanego porównywania obecnego stanu bazy wiedzy z nową regułą wpływu jest to, czy reguła wpływu stanowi potwierdzenie reguły wynikowej i tym samym przeprowadzić należy procedurę wzmocnienia reguły wynikowej. Wzmocnienie może nastąpić wtedy, gdy wszystkie tendencje faktów zawarte w regule wpływu występują także we wzmacnianej regule wynikowej i wszystkie pary tendencji faktów są analogiczne – czyli rozbieżności ich cech mieszczą się w tolerancjach zadanych parametrami i opisanych wyżej. Możliwe jest wzmocnienie regułą wpływu posiadającą mniej tendencji faktu niż reguła wynikowa, nadal jednak z zachowaniem ich zgodności. Proces wzmocnienia reguły wpływu przebiega następująco:

Dla każdej tendencji faktu zapisanej w regule wynikowej zdefiniowana jest wartość całkowita (ozn. *STR*) reprezentująca siłę tej tendencji w regule, czyli opisująca to, ile razy tendencja faktu występowała już we wzmacniających bieżącą regułę wynikową regułach wpływu. Procedura wzmocnienia spowoduje zwiększenie o 1 tych wartości, które odpowiadają tendencjom faktu występującym we wzmacniającej regule wpływu. Procedura osłabiania zmniejszy te wartości, a po osiągnięciu 0 usunie z reguły odpowiednią tendencję faktu. W przypadku tworzenia nowej reguły wynikowej (gdy nie doszło do żadnego dopasowania) wartości inicjowane są na 1.

Tab. 3. Reguła wynikowa

No	T_{dist}	T_{len}	STR	VID	F
1	0	30	11	V3	0,6
3	150	15	21	V7	1,3
2	20	30	23	V1	1,25

Uwagi: STR – siła tendencji faktu w regule wpływu.

Wygenerowana reguła nie posiada jednoznacznie zdefiniowanych tez i przesłanek, definiując jedynie możliwe konsekwencje zaobserwowanych zjawisk (tendencji faktów) na skali wartości czynnika wiodącego. Co zostanie w regule potraktowane jako przesłanka, a co jako teza – będzie miało znaczenie dopiero podczas konkretnej ekspertyzy.

8. Eliminacja nadmiarowości danych

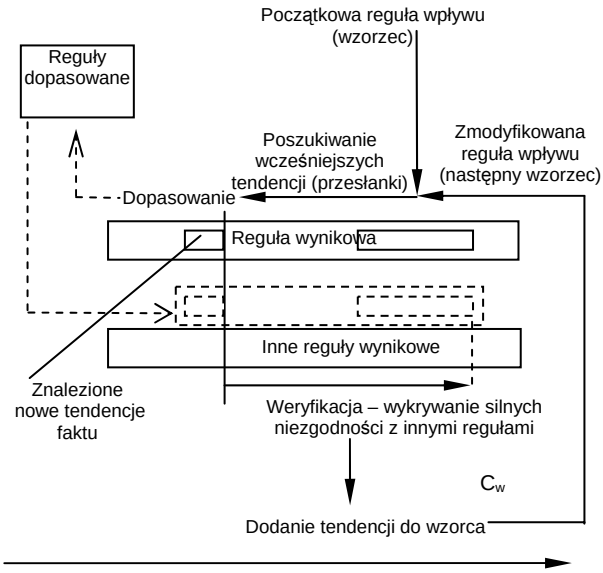
Liczebność zarejestrowanych faktów-zdarzeń dotyczących tego samego czynnika i tego samego obiektu jest ograniczana odgórnie zadanym parametrem. Parametr ten z jednej strony ogranicza precyzję postrzegania zmienności czynnika w czasie, z drugiej utrzymuje wielkość bazy fatów-zdarzeń w rozsądnych z punku widzenia możliwości przetwarzania rozmiarach. Nowo postrzegane fakty-zdarzenia są wykorzystywanie do tworzenia lub korygowania tendencji faktów, jednak z uwagi na znaczną utratę informacji, powodowaną przez procedurę interpolującą, nie ma powodu utrzymywania dowolnie dużej puli wartości interpolowanych. Nowe fakty-zdarzenia po liczebnym przekroczeniu wartości granicznej redukowane są w następujący sposób:

Nowo utworzony fakt-zdarzenie, gdy poszerza zakres wartości czynnika wiodącego, zastępuje fakt-zdarzenie poprzednio posiadający ekstremalną wartość czynnika wiodącego. W innym przypadku obliczana jest średnia arytmetyczna z nowo utworzonego faktu-zdarzenia i najbliższego w sensie wartości czynnika wiodącego faktu-zdarzenia zapisanego już w bazie wiedzy. Średnia ta zastępuje fakt-zdarzenie już istniejący w bazie. Zarejestrowanie nowego faktu-zdarzenia powoduje automatyczną korektę odpowiednich tendencji faktów i reguł wpływu, jeśli takie już istnieją.

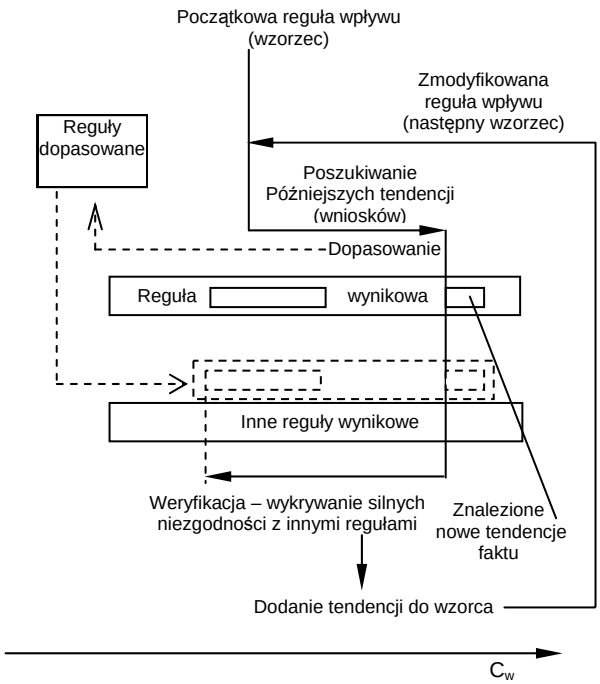
9. Ekspertyzy z użyciem wiedzy przyczynowo-skutkowej

Przetwarzanie wiedzy zebranej w wyniku opisanego w poprzednich paragrafach formatu umożliwia przeprowadzanie ekspertyz determinujących przyczyny lub skutki wystąpienia podanego do ekspertyzy zjawiska. Zjawisko to może być reprezentowane przez pojedynczą tendencję faktu, której okoliczności wystąpienia chcemy poznać. Możliwe jest także podanie na wejściu ekspertyzy kilku tendencji faktów, sprzężonych ze sobą w regule wpływu. Ta informacja wejściowa dla ekspertyzy w dalszej części będzie nazywana *wzorcem*. Ponieważ podstawowym miernikiem podanego zjawiska będzie parametr wiodący, możliwe jest zdefiniowanie dwóch wariantów przeprowadzanej ekspertyzy, posiadających bliźniacze algorytmy, a w zależności od wariantu – analizujących albo przyczyny albo skutki podanego na wejściu wzorca. Całość ekspertyzy będzie polegała na przemieszczaniu obliczeń po skali wartości czynnika wiodącego, wychodząc od wartości charakteryzujących podane zjawisko wejściowe i idąc albo w kierunku wartości C_w większych (poszukiwanie skutków, proponowanie rozwiązań) lub mniejszych (poszukiwanie przyczyn zjawiska). Warianty ekspertyzy można więc zdefiniować jako:

- *Ekspertyza w przód* – polega na ustalaniu konsekwencji przyczynowo-skutkowych wystąpienia podanej na wejściu sytuacji i może przykładowo służyć do sprecyzowania decyzji, jaką należy podjąć, aby właściwie zmienić zastaną sytuacją lub ciąg przyczynowo-skutkowy.
- *Ekspertyza wstecz* – polega na poszukiwaniu przesłanek przyczynowo-skutkowych wystąpienia podanej na wejściu sytuacji. Może zostać użyta do wyjaśnienia przyczyn zastanej sytuacji.



Rys. 2. Schemat procesu ekspertyzy dla wariantu: Ekspertyza wstecz



Rys. 3. Schemat procesu ekspertyzy dla wariantu: Ekspertyza w przód

Obydwie ekspertyzy realizowane są za pomocą bardzo podobnych do siebie algorytmów, różniących się jedynie kierunkiem przesuwania się po skali wartości czynnika wiodącego, zakodowanych w regułach przetwarzanej przez te algorytmy bazy wiedzy. Założenia dotyczące realizacji procesu ekspertyzy można ująć w następujących punktach:

Proces pobiera na wejściu regułę wpływu kodującą poszukiwane (wejściowe) relacje pomiędzy tendencjami faktów na skali czynnika wiodącego. Reguła ta może być (w najprostszym przypadku) pojedynczą tendencją faktu. Wtedy ekspertyza dotyczyć będzie wyłącznie okoliczności wzrostu czy spadku wartości danego parametru obiektu. Tendencje faktu

zapisane w tej regule tworzą wzorzec, który będzie rozwijany o nowe tendencje faktu umieszczone na skali wartości czynnika wiodącego w kierunku przeprowadzania ekspertyzy (wstecz – wartości mniejsze, do przodu – większe).

Proces będzie przeszukiwał posiadaną bazę reguł wynikowych (powstałych w wyniku automatycznego pozyskiwania wiedzy lub wprowadzonych przez eksperta dziedzinowego) dokonując próby dopasowania zadanej reguły wpływu do każdej z reguł wynikowych. Dopasowanie nastąpi w momencie, gdy wyszukana reguła będzie spełniała następujące warunki:

- będzie posiadała przynajmniej jedną tendencję faktu zgodną z tą zawartą w już ustalonym wzorcu,
- nie będzie posiadała tendencji faktu kolidującej z którąkolwiek z pozostałych we wzorcu, czyli posiadającej takie samo VID, pokrywające się przynajmniej częściowo przedziały wartości czynnika wiodącego, lecz z kolei wartości współczynnika F o znacznie różniących się wartościach (zwłaszcza sugerujących proporcjonalność odwrotną). Tolerancja takiej różnicy jest ustalana parametrycznie,
- będzie posiadała przynajmniej jedną tendencję faktu nie występującą w obecnym wzorcu i charakteryzowaną mniejszymi wartościami C_w w przypadku ekspertyzy wstecz i większymi w przypadku ekspertyzy w przód niż którakolwiek tendencja faktu w ustalonym już wzorcu.

Z puli wyselekcjonowanych reguł dopasowanych zostanie następnie wybrana reguła, posiadająca najsilniejszą tendencję faktu (biorąc pod uwagę tylko te tendencje, które nie występowały we wzorcu). Tendencja ta poszerzy tymczasowo wzorzec. Pozostały wybór reguł zostanie zachowany i użyty po ewentualnym powrocie z bieżącego wątku ekspertyzy.

W kolejnym kroku wybrana reguła będzie weryfikowana. Weryfikacja polegać będzie na próbie ustalenia innych silnych reguł wynikowych zawierających tendencje faktów kolidujące z ustalonym już wzorcem, a znajdujących się na skali czynnika wiodącego w kierunku przeciwnym do kierunku ekspertyzy (jak pokazano na rysunku). Innymi słowy, w przypadku ekspertyzy wstecz weryfikacja będzie postępować na skali C_w w kierunku zwiększających się wartości, a w przypadku ekspertyzy do przodu – w kierunku zmniejszających się.

W przypadku niepowodzenia weryfikacji reguły nastąpi powrót do puli wyselekcjonowanych reguł i pobranie z niej kolejnej, zgodnie ze strategią wyboru. Ekspertyza zakłada, jedynie możliwość wykrycia sytuacji analogicznych do zaobserwowanych w fazie automatycznego uczenia, więc nie wymaga się od niej szukania wszystkich możliwych relacji pomiędzy faktami-zdarzeniami, które mogłyby zostać zdefiniowane na przykład na podstawie mniejszej liczby zaobserwowanych przypadków analogicznych. Naturalnie istnieje możliwość typowania i takich wyników – rozpatrując kolejne reguły z puli dopasowanych w przypadku powodzenia ich weryfikacji.

Zweryfikowana reguła posiadać będzie nowe tendencje faktów, które zostaną włączone do aktualnego wzorca. Reguły będą włączane kolejno zgodnie z wartościami czynnika wiodącego (więc również zależnie od kierunku ekspertyzy). Nastąpi to jednak tylko wówczas, gdy siła reguły wpływu jest przynajmniej taka sama, jak siła reguły rozpatrywanej poprzednio.

Po wyczerpaniu puli wyselekcjonowanych reguł nastąpi nawrót do poprzednio zdefiniowanego zestawu reguł i próba użycia (weryfikacji) kolejnej z nich.

10. Algorytm dla ekspertyzy

Algorytm realizujący ekspertyzę można zapisać w następujących punktach:

1. Zdefiniuj pulę reguł dopasowanych zgodnie z założeniami opisanymi wcześniej.
2. Jeżeli zbiór jest pusty – zakończ, zwracając wzorzec bieżący.
3. Wybierz regułę z najsilniejszą nową tendencją faktu.
4. Usuń wybraną regułę ze zbioru reguł dopasowanych.
5. Dołącz tymczasowo do wzorca tendencje faktów posiadające STR większe lub równe od STR którejkolwiek tendencji faktu już znajdującej się we wzorcu. Jeżeli takich w regule nie ma – nawróć do kolejnej reguły dopasowanej, przechodząc do 3.
6. Zweryfikuj wzorzec tymczasowy (wykrywanie kolizji tendencji faktów). Jeśli kolizje wystąpiły – nawróć do kolejnej reguły w wybranych, przechodząc do 3.
7. Przyjmij wzorzec tymczasowy jako bieżący i przejdź do 1.

11. Podsumowanie

W artykule został zaproponowany sposób gromadzenia wiedzy umożliwiający analizę relacji przyczynowo-skutkowych pomiędzy atrybutami obserwowanych obiektów. Obserwacja obiektów przebiega w pełni automatycznie dzięki śledzeniu informacji zawartej w komunikatach HL7, będących częścią oficjalnie przyjętego standardu komunikacji. Zastosowane podejście nie wymaga wprowadzania żadnych modyfikacji jednostek już produkujących komunikaty na inne potrzeby. W regułach z wiedzy pozyskanej opisaną metodą nie dysponujemy jednoznacznie określoną pulą przesłanek i płynących z nich konkluzji. Mamy natomiast zakodowany ciąg często występujących po sobie zjawisk, zapisany przy użyciu tendencji faktów. Tendencje te można traktować zamiennie (jako przesłanki lub konkluzje) w zależności od tego, która z nich będzie punktem odniesienia dla pozostałych. Samo wykorzystanie reprezentowanej wiedzy nieznacznie nawiązuje do przetwarzania klasycznej wiedzy regułowej. Stosowany algorytm jest nadal rozwijany i jego ostateczny kształt pozostaje oczywiście sprawą otwartą. Dalszymi krokami na polu testów i modyfikacji metody będą bez wątpienia kolejne eksperymenty z modyfikacjami procesu wzmacniania reguł i ekspertyz na bazie powstałej sieci połączeń w regułach wynikowych. Możliwe jest stworzenie przejrzystej prezentacji graficznej wyników ekspertyzy z przedstawieniem ścieżki przejść pomiędzy regułami wynikowymi i odfiltrowaniem relacji występujących sporadycznie. Dodatkowo możliwe jest łatwe rozbudowanie procesu ekspertyzy o dodanie przetwarzania ścieżek alternatywnych, realizowanych po słabszych regułach, co umożliwi przedstawienie kilku alternatywnych wyników wraz uzasadnieniami, pozostawiając użytkownikowi wybór ostatecznego rozwiązania.

Literatura cytowana

1. Dębicki M., Grudzień W.: Interface HL7 pomiędzy szpitalnym systemem informatycznym (HIS) a specjalizowanym modułem diagnostycznym, Lublin 2004.
2. Rudowski R.: Informatyka Medyczna, 2003.
3. http://www.medinfo.rochester.edu/hl7
4. http://www.chip.org/projects/bili/hl
5. http://www.hl7.au.org

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 51-56.

ZASTOSOWANIE SIECI IMMUNOLOGICZNEJ DO DRĄŻENIA DANYCH

DATA MINING WITH ARTIFICIAL IMMUNE SYSTEM

WIESŁAW WAJS[*], MARIUSZ ŚWIĘCICKI[**], PIOTR WAIS[***], HUBERT WOJTOWICZ[***], PAWEŁ JANIK[***]

[*]*Katedra Automatyki Akademii Górniczo-Hutniczej, Kraków, wwa@ia.agh.edu.pl*
[**]*Instytut Modelowania Komputerowego, Politechnika Krakowska, Kraków, mswiecic@pk.edu.pl*
[***]*Instytut Politechniczny, Państwowa Wyższa Szkoła Zawodowa w Krośnie, Krosno, waisp@poczta.onet.pl, hubwoj@krosno.pwsz.pl*

Streszczenie. W artykule zostanie przedstawione wykorzystanie sztucznej sieci immunologicznej do klasyfikacji wielowymiarowych danych medycznych.
Stworzono sztuczną sieć immunologiczną, której zadaniem jest klasyfikacja parametrów gazometrii krwi tętniczej (pH, pCOa, pO_2, HCOs), która została poddana procesowi uczenia.
Dane do procesu uczenia pochodzą z Oddziału Intensywnej Terapii Noworodka Polsko-Amerykańskiego Instytutu Pediatrii Collegium Medicum UJ w Krakowie.

Słowa kluczowe: sztuczna sieć immunologiczna, drążenie danych

Abstract: In the paper, application of artificial immune system for classification of multidimensional medical data is presented. An artificial immune network was created for the purpose of blood gasometry parameters classification. Training data originate from the Infant Intensive Care Ward of the Polish–American Institute of Pediatry, Collegium Medicum of the Jagiellonian University in Krakow.

Key words: artificial immune system, data mining

1. Sztuczne systemy immunologiczne

1.1. Reprezentacja przeciwciał i antygenów

W początkowym okresie stosowania algorytmów immunologicznych zakładano, że antygeny i przeciwciała były modelowane przy pomocy łańcuchów binarnych [1]. Obecnie raczej jest stosowana reprezentacja, w której antygen lub przeciwciało są reprezentowane za pomocą wektorów liczb rzeczywistych. W tej sytuacji, można będzie skorzystać z metod matematycznych. Zastosowanie tych metod spowoduje, że relacja podobieństwa antygenu do przeciwciała stanie się wartością numeryczną.

Można przyjąć, że kształt molekuły *m* będzie reprezentowany w przestrzeni liczb rzeczywistych o następujących współrzędnych:

$m = \langle m_1, m_2, ..., m_L \rangle$, który można traktować jako punkt w przestrzeni *L*-wymiarowej.

Podobieństwo pomiędzy antygenem a przeciwciałem można interpretować jako odległość pomiędzy dwoma punktami w przestrzeni. Stopień dopasowania (powinowactwa) może być obliczony jako odległość Euklidesowa wg wzoru (1).

Wówczas współrzędne przeciwciała przedstawia następujący wektor $\langle ab1, ab2, ..., abL \rangle$, natomiast wektor $\langle ag1, ag2, ..., agL \rangle$ określa współrzędne antygenu.

$$D = \sqrt{\sum_{i=1}^{L} \left(Ab_i - Ag_i \right)^2} \qquad (1)$$

1.2. Algorytm selekcji klonalnej

Jest to algorytm, który bazuje na mechanizmach występujących podczas odpowiedzi immunologicznej systemu w sytuacji kiedy antygen zostanie zidentyfikowany. Nazywany jest on również algorytmem selekcji pozytywnej. Algorytm ten związany jest z limfocytami typu *B* i wykorzystuje mechanizm odpowiedzi *B* komórek, które zostaną aktywowane obecnością antygenu. Jeżeli np. przeciwciało, które jest przyczepione do komórki zidentyfikuje antygen, wówczas nastąpi przekazanie sygnału do wnętrza komórki. Sygnał ten zapoczątkuje proces podziału i różnicowania się aktywowanej komórki. Początkowo inicjujemy zbiór *(M)*, który zawiera przeciwciała, natomiast zbiór *(P)* zawiera wzorce, a posługując się przyjętą wcześniej nomenklaturą – zbiór ten będzie zawierał antygeny. Następnym krokiem algorytmu jest prezentacja antygenu

wszystkim przeciwciałom i obliczenie wartości powinowactwa. W punkcie pierwszym prezentowanego algorytmu wybieramy ustaloną *N1* komórek o największym poziome stymulacji. Wybrane przeciwciała poddawane są procesowi klonowania. Podczas procesu klonowania obowiązuje zasada, że komórka, a właściwie przeciwciało, które zidentyfikowało dany antygen, będzie posiadało największą liczbę kopii w porównaniu do przeciwciał o mniejszym poziomie stymulacji. Punkt drugi algorytmu realizuje procesu mutacji nowo utworzonych kopii przeciwciał. Przeciwciała mutowane są w ten sposób, że przeciwciało o największej wartości powinowactwa będzie posiadało najniższą wartość współczynnika mutacji, czyli zostanie najmniej zmodyfikowane. W trzecim punkcie realizowana jest operacja dodawania nowoutworzonych przeciwciał do zbioru *(M)* zawierającego przeciwciała.

Algorytm jest zatrzymywany wtedy, kiedy zostaną spełnione kryteria zatrzymania, które mogą być związane z liczbą poprawnie rozpoznanych wzorców *(P)* lub z błędem.

- *Inicjacja populacji przeciwciał w sposób losowy (M);*
- *For each antygenu należącego do zbioru P, prezentuj przeciwciała należące do zbioru M oraz wyznacz wartość powinowactwa dla każdego elementu zbioru M;*

  1. *Wybierz N1 przeciwciał o największej wartości powinowactwa ze zbioru M i w utwórz ich kopie proporcjonalnie do wartości powinowactwa do antygenu. (Przeciwciało o największej wartości powinowactwa posiada największą liczbę kopii).*
  2. *Poddanie procesowi mutacji wszystkich nowo utworzonych kopii przeciwciał ze wskaźnikiem mutacji odwrotnie proporcjonalnym do wartości powinowactwa (Duża wartość powinowactwa przeciwciała do antygenu, niska wartość wskaźnika mutacji).*
  3. *Dodanie wszystkich zmodyfikowanych przeciwciał w poprzednim punkcie do zbioru.*
- *Powtarzanie kroków 1, 2, 3 do momentu spełnienia zadanych kryteriów.*

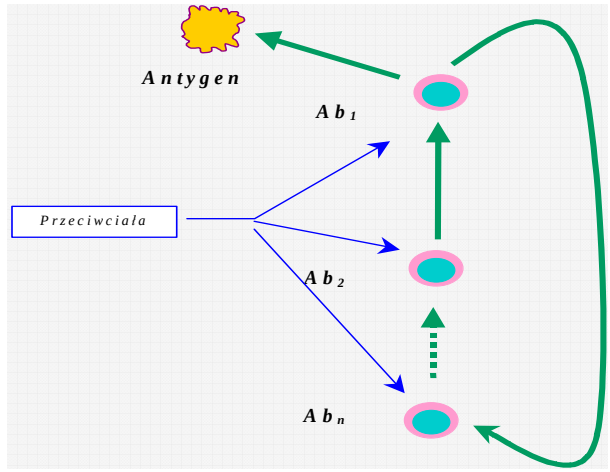
Jak można zauważyć, mechanizm funkcjonowania algorytmu selekcji klonalnej jest bardzo podobny do klasycznego algorytmu genetycznego. Jedną z różnic, która ma decydujące znaczenie dla jego własności jest to, że w przedstawionym algorytmie nie występuje operacja krzyżowania. Właśnie między innymi brak operacji krzyżowania powoduje, że populacja potencjalnych rozwiązań (*M*) ciągle zachowuje swoją różnorodność. Zachowanie różnorodności populacji ma decydujące znaczenie np. w przypadku optymalizacji funkcji wielomodalnych.

1.3. Sieć immunologiczna

Teoria dotycząca sieci immunologicznej została zaproponowana w 1974 roku [2]. Podstawowym założeniem, na którym bazowała i które częściowo wynikało z obserwacji systemu immunologicznego, była hipoteza dotycząca limfocytów. Zaobserwowano, że w sytuacji, kiedy system immunologiczny nie jest zainfekowany żadnym antygenem, limfocyty ciągle wykazują dużą aktywność.

Właśnie ta obserwacja skłoniła do postawienia hipotezy mówiącej, że podczas funkcjonowania systemu pewne przeciwciała są zdolne do rozpoznawania innych przeciwciał, tzn. tworzone są wiązania antyidiotypowe.

To założenie sugeruje, że system immunologiczny tworzy sieć komórek i zadaniem każdej z tych komórek jest kontrola aktywności komórek sąsiednich w taki sposób, jak to zostało przedstawione na rys. 1.



Rys. 1. Formowanie pętli antyidiotypowej

Sieć immunologiczną można zdefiniować jako komórki, które tworzą między sobą wiązania typu:

– przeciwciało-przeciwciało,
– wiązania antyidiotypowe,
– przeciwciało-antygen,
– wiązania idiotypowe.

Odpowiedź tak zdefiniowanego systemu immunologiczna może mieć charakter zarówno pozytywny, jak też negatywny, ponieważ system może rozpoznawać nie tylko antygen, ale także inne przeciwciała, które stanowią podstawę sieci immunologicznej.

1.4. Algorytm uczenia sztucznej sieci immunologicznej

Sztuczna sieć immunologiczna wykorzystuje mechanizm wiązań idiotypowych i antyidiotypowych i bazuje na interakcjach związanych z limfocytami typu *B*.

Pierwszym krokiem algorytmu, który tutaj został przedstawiony jest załadowanie populacji zbioru antygenów *(Ag)*. Przez zbiór antygenów należy rozumieć zbiór uczący, który zawiera wzorce.

Drugim równie istotnym krokiem jest inicjalizacja populacji *B* komórek, która w przypadku tego algorytmu została zredukowana do przeciwciał, które będą pełnić funkcje receptorów dla antygenów.

Po fazie inicjalizacji populacji komórek *B* zaczyna się etap uczenia sieci immunologicznej. Dla każdego antygenu, który należy do zbioru *(Ag)* przeprowadzamy operacje, które są opisane w punktach od 1 do 10 w przedstawionym poniżej algorytmie uczenia sieci idiotypowej.

For each Agj $\in$ Ag do

1. *Wyznaczenie powinowactwa pomiędzy j-tym antygenem Ag fi,j, i = 1,...,N, do wszystkich przeciwciał Abi. fi,j = 1/Di,j, i = 1,...,N;*

2. *Utworzenie zbioru A Ab{n} zawierającego n – przeciwciał o najwyższym powinowactwie;*

3. *n wybranych przeciwciał poddawane są procesowi klonowania proporcjonalnie do ich powinowactwa do antygenów fi,j oraz generowanie zbioru klonów C o najwyższym powinowactwie;*

4. *Elementy zbioru C poddawane są mutacji i generowany jest zbiór C*, prawdopodobieństwo mutacji jest odwrotnie proporcjonalne do powinowactwa przeciwciał rodzicielskich do j-tego antygenu Ag fi,j: Duże powinowactwo – mały wskaźnik mutacji;*

5. *Wyznaczenie powinowactwa dk,j = 1/Dk,j pomiędzy Agj dla każdego elementu zbioru C*;*

6. *Ze zbioru C* wybranie ζ% przeciwciał, które mają największy wskaźnik powinowactwa dk,j i utworzenie z nich komórek pamięciowych Mj;*

7. *Śmierć komórek: eliminacja wszystkich komórek pamięciowych ze zbioru Mj, dla których wartość powinowactwo spełnia relacje Dk,j >σd;*

8. *Wyznaczenie powinowactwa pomiędzy klonowanymi przeciwciałami si,k;*

9. *Suppresja: eliminacja tych klonów ze zbioru M, dla których spełniona jest relacja si,k < σs;*

10. *Połączenie wszystkich przeciwciał, które są zapisane w zbiorze Mj;*

- *Wyznaczenie powinowactwa pomiędzy wszystkimi przeciwciałami pamięciowymi Ab{m}:*
- *Supresja: eliminacja tych przeciwciał, które spełniają relacje: si,k <σs;*
- *Utworzenie zbioru przeciwciał*

  Ab ←[Ab{m};Ab{d}].

Gdzie:

Ab: dostępny zbiór przeciwciał (Ab ∈SN×L, Ab = Ab{d} ∪Ab{m});

Ab{m}: Zbiór przeciwciał które są komórkami pamięciowymi (Ab{m} ∈Sm×L, m ≤N);

Ab{d}: d nowo utworzone przeciwciała Ab (Ab{d} ∈Sd×L);

Ag: Populacja antygenów (Ag ∈SM×L);

fj: Wektor zawierający wartość powinowactwa dla wszystkich przeciwciał Abi (i = 1,...N) w relacji do j-tego antygenu Agj. W tym przypadku powinowactwo jest odwrotnie proporcjonalnie do odległości pomiędzy Ag-Ab;

S: macierz podobieństwa pomiędzy każdą z par przeciwciał Abi-Abj, si,j (i,j = 1,...,N);

C: Zbiór populacji N – klonów przeciwciał o największym powinowactwie; generacja na podstawie zbioru Ab (C∈ L Nc S ×);

C: populacja przeciwciał, które zostały utworzone z C i poddane procesowi mutacji;*

dj: Wektor zawierający wartości powinowactwa pomiędzy elementem zbioru C i każdym elementem zbioru Agj;*

ζ: parametr określający, jaki procent populacji przeciwciał będzie poddawany procesowi klonowania w procesie aktywacji antygenem;

Mj: Komórki pamięciowe dla antygenu Agj (pozostałe po procesie supresji);

Mj: Komórki pamięciowe utworzone z j-tego przeciwciała Agj;

σ_d: próg usuwania komórek;

σ_s: próg supresji.

2. Opis modelowanego problemu

Zjawiska biologiczne, a w szczególności procesy zachodzące w organizmie mają charakter dynamiczny. Taki też charakter ma diagnoza i leczenie. W dochodzeniu do rozpoznania i podejmowaniu decyzji terapeutycznych rozpatrujemy zjawiska zachodzące w czasie. Każdy lekarz wie, że bardzo rzadko stawiamy diagnozę i podejmujemy decyzję dotyczącą leczenia na podstawie jednej obserwacji klinicznej. Z reguły opieramy się na kilkakrotnym badaniu pacjenta, wykonaniu wielu, często powtarzanych pomiarów parametrów biofizycznych i biochemicznych, czy też badań obrazowych. Na pewnym etapie tych badań ustalana jest diagnoza wstępna, w dalszej kolejności ostateczna. W oparciu o ustalone na danym etapie obserwacji pacjenta rozpoznanie i ocenę dynamiki procesu chorobowego podejmujemy decyzję o sposobach leczenia.

W ciągu ostatnich dwudziestu lat w neonatologii dokonał się ogromny postęp. Był on możliwy dzięki rozwojowi nowoczesnych technik intensywnej opieki medycznej w zakresie badań diagnostycznych, jak i monitorowania czynności życiowych oraz postępu w dziedzinie nauk podstawowych. Badania diagnostyczne opierają się na mikrometodach. Nowoczesne techniki monitorowania – zarówno nieinwazyjne, jak i inwazyjne cechują się niewielką uciążliwością dla pacjenta, pozwalają dokładnie, w sposób ciągły monitorować wymianę gazową, mechanikę płuc i czynności układu krążenia. Poznano patofizjologię zespołu zaburzeń oddychania noworodków, rolę niedoboru surfaktantu w różnych formach niewydolności oddechowej. Poznano fizjologię i patofizjologię gospodarki wodno-elektorlitowej u noworodków donoszonych i urodzonych przedwcześnie.

Wstępna stabilizacja stanu dziecka jest zadaniem trudnym. Może trwać nawet kilka dni. Aby ją osiągnąć, lekarz wielokrotnie analizuje liczne parametry związane ze stanem zdrowia pacjenta. Są to dane z wywiadu okołoporodowego, odchylenia od stanu prawidłowego w badaniu fizykalnym, pomiary fizjologiczne (masa ciała, diureza godzinowa), wyniki badań dodatkowych (biochemicznych, mikrobiologicznych, obrazowych) oraz wskazania urządzeń monitorujących (pulsoksymetr, kardiomonitor, inwazyjny pomiar ciśnienia tętniczego, monitor mechaniki oddychania). Niejednokrotnie lekarz ocenia jednocześnie ponad 50 zmiennych. Analiza tak dużej ilości informacji jest trudna, wymaga doświadczenia, tym bardziej, że decyzję o sposobie postępowania często należy podjąć szybko. W wyniku przeprowadzonej analizy danych lekarz podejmuje decyzję o sposobach leczenia, spodziewając się korzystnych rezultatów, które wyrażają się pożądanymi zmianami w wynikach badań dodatkowych, wskazaniach urządzeń monitorujących, badaniu fizykalnym. Cały ten proces można sprowadzić do dokonywanego przez

lekarza modelowania przebiegu niewydolności oddechowej. Tworzenie tego modelu opiera się na wiedzy teoretycznej i empirycznej, zawartej w podręcznikach i artykułach naukowych oraz na doświadczeniu lekarza. Z konieczności tworzony model ma charakter subiektywny. Ponadto wynik analizy przeprowadzanej przez lekarza zawsze zależny jest od wielu czynników zewnętrznych – ostatnio przeczytanych doniesień, zmęczenia itd.

Nasilenie zaburzeń wymiany gazowej najlepiej odzwierciedlają parametry gazometrii krwi tętniczej, rozpatrywane w kontekście aktualnie stosowanych metod wspomagania wentylacji. Z tego względu jako wielkości wyjściowe wybrano cztery bezpośrednio mierzone parametry gazometrii krwi tętniczej: pH, pCO_2, pO_2, HCO_3 oraz wskaźnik tlenowy (pO_2/FiO_2). Pominięto niedobór zasad i wysycenie hemoglobiny tlenem jako parametry pochodne czterech uprzednio wymienionych wielkości.

2.1. Modelowane parametry

Poniżej przedstawiono krótką charakterystykę prognozowanych parametrów.

pH – Odczyn krwi jest parametrem zależnym od stanu układu oddechowego, ale także od czynności układu krążenia, nerek i metabolizmu ustroju. Jest to więc wskaźnik ogólny, który można traktować jako wskaźnik ogólnego stanu pacjenta. Wartości prawidłowe pH wahają się od 7.35 do 7.45.

pCO_2 – Ciśnienie parcjalne dwutlenku węgla ściśle koreluje ze stopniem wydolności układu oddechowego. Zależne jest przede wszystkim od wentylacji minutowej. Jej wzrost powoduje spadek pCO_2. Ze wzrostem ciśnienia parcjalnego dwutlenku węgla mamy głównie do czynienia w zaburzeniach oddychania przebiegających z obniżoną wentylacją. Jednak w ciężkich zaburzeniach oddychania, przebiegających z upośledzeniem dyfuzji gazów przez barierę pęcherzykowo-włośniczkową także dochodzi do upośledzenia wymiany dwutlenku węgla. Wzrost pCO_2 obniża pH krwi. U pacjentów wentylowanych mechanicznie pCO_2 zależy głównie od częstości oddechów (RR) i objętości oddechowej (TV). Wartości prawidłowe pCO_2 wahają się od 40 do 55 mHg.

pO_2 – Ciśnienie parcjalne tlenu także ściśle koreluje ze stopniem wydolności układu oddechowego. Prawidłowa wymiana tlenu w płucach zależy głównie od odpowiedniego stosunku wentylacji do perfuzji oraz od przepuszczalności bariery pęcherzykowo-włośniczkowej. Zaburzenie stosunku wentylacji do perfuzji obserwujemy głównie w przypadku niedodmy. Pogrubienie bariery pęcherzykowo-włośniczkowej w pierwszym rzędzie upośledza wymianę tlenu, dopiero dalsze ograniczenie jej przepuszczalności doprowadza do upośledzenia wymiany CO_2. U pacjentów, u których stosujemy wspomaganie oddychania, pO_2 zależy od stężenia tlenu w mieszaninie oddechowej (FiO_2) oraz od stopnia rozprężenia płuc. Ten z kolei u pacjentów wentylowanych mechanicznie uwarunkowany jest odpowiednim ciśnieniem w drogach oddechowych (PIP – ciśnienie szczytowe wdechu, MAP – średnie ciśnienie w drogach oddechowych). Niedobór tlenu w organizmie powoduje nasilenie metabolizmu beztlenowego, a więc także zwiększonej produkcji kwasu mlekowego i spadku pH krwi. U dzieci urodzonych przedwcześnie wartości prawidłowe pO_2 wynoszą od 40 do 50 mmHg.

HCO_3 – Stężenie wodorowęglanów we krwi zależy od metabolizmu ustroju oraz od funkcji nerek. W przypadku dysproporcji między dowozem tlenu do tkanek a potrzebami metabolicznymi ustroju dochodzi do spadku stężenia dwuwęglanów i kwasicy metabolicznej. Wielkość ta zatem pośrednio zależy od pO_2. Stężenie dwuwęglanów we krwi wyraża się w mmol/l, wartości prawidłowe mieszczą się w przedziale 21-25 mmol/l.

2.2. Źródło danych

Materiał badawczy wykorzystany w doświadczeniach do prezentowanej pracy został zgromadzony w trakcie kilkuletniej obserwacji i hospitalizacji pacjentów na Oddziale Intensywnej Terapii Noworodka Polsko-Amerykańskiego Instytutu Pediatrii Collegium Medicum Uniwersytetu Jagiellońskiego. Na tym oddziale wykorzystywana jest komputerowa baza danych – Neonatal Information System (NIS), służąca do przechowywania informacji o wszystkich hospitalizowanych dzieciach.

3. Przykład – klasyfikacja wielowymiarowych danych medycznych

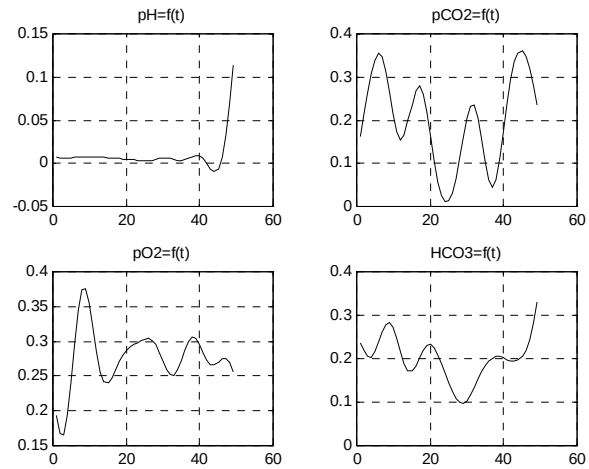
W danym przykładzie zostanie przedstawione wykorzystanie sztucznej sieci immunologicznej do klasyfikacji wielowymiarowych danych o charakterze medycznym.

W oparciu o bazę danych funkcjonującą od kilku lat na Oddziale Intensywnej Terapii Noworodka Polsko-Amerykańskiego Instytutu Pediatrii Collegium Medicum UJ w Krakowie stworzono sztuczną sieć immunologiczną, której zadaniem jest klasyfikacja parametrów gazometrii krwi tętniczej (pH, pCO_2, pO_2, HCO_3). W procesie uczenia sieci immunologicznej były wykorzystane dane wejściowe – uprzednie wartości gazometrii oraz nastawy respiratora i fakt podania surfaktantu.

Proces uczenia sieci immunologicznej został podzielony na dwie fazy. Pierwsza faza będzie związana z procesem trenowania sieci immunologicznej. Faza trenowania sieci będzie odbywać się według algorytmu przedstawionego wcześniej. Zbiór trenujący będzie zawierał przebiegi parametrów krwi (pH, pCO_2, pO_2, HCO_3) od czasu (t) - szeregi czasowe o takich samych długościach.

Druga faza będzie związana ze sprawdzeniem zdolności (testowania) do uogólniania sieci poprzez podanie na wejścia sieci przebiegów ze zbioru testującego.

Elementy zbioru uczącego – szeregi czasowe będą przez sieć interpretowane jako antygeny, które będą stymulowały odpowiednie przeciwciało tworzące sieć immunologiczną.



Rys. 2. Przykład wektora danych uczących

Zbiór uczący sieć immunologiczną zawiera wzorce przebiegów, które są reprezentowane w postaci wektorów. Z tego powodu każdy szereg czasowy zawarty w zbiorze uczącym można interpretować jak współrzędne punktu w przestrzeni o wymiarze zależnym od długości wzorca przebiegu.

Taka interpretacja elementów szeregu czasowego, pozwala na zastosowanie do obliczania powinowactwa relacji antygen *(Ag)* – przeciwciało *(Ab)* wzór (1).

Tab.1. Struktura zbioru uczącego.

Nr antygenu	Czas			
	t1,t2…t49	t1,t2…t49	t1,t2…t49	t1,t2…t49
1	pH	pCO_2	PO_2	HCO_3
2	pH	pCO_2	PO_2	HCO_3
:	pH	pCO_2	PO_2	HCO_3
200	pH	pCO_2	PO_2	HCO_3

3.1. Proces uczenia

W procesie uczenia, zbiór uczący zawierał 200 wzorców, a parametry związane z siecią immunologiczną miały następujące wartości:

$n = 8,\ \zeta = 30\%,\ \sigma d = 1.2,\ \sigma s = 0.2$

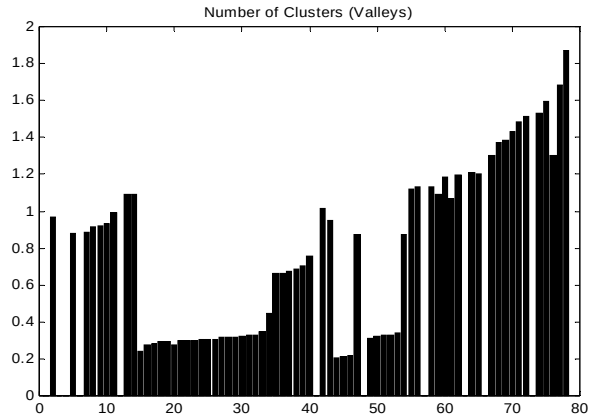
$n = 8,\ \zeta = 30\%,\ \sigma d = 1.2,\ \sigma s = 0.01$

Kryteria zatrzymania procesu uczenia były związane z liczbą generacji i w tym przypadku liczba ta wynosiła *Ngen = 200.*

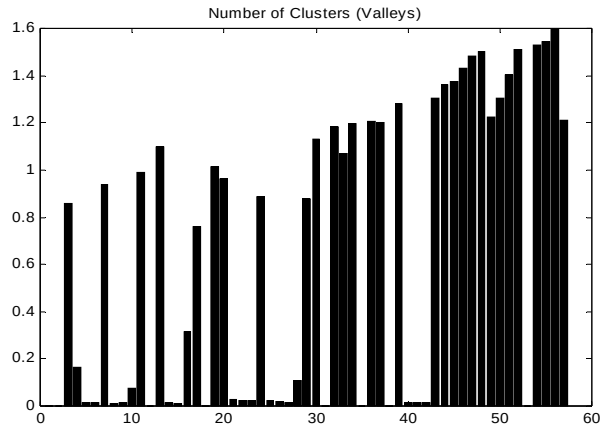
Jednym z ważniejszych parametrów, który ma decydujący wpływ na strukturę sieci immunologicznej powstającej w procesie uczenia, jest parametr (σs). Zmieniając wartość progu supresji (σs) wpływamy na plastyczność sieci immunologicznej. Zwiększenie wartości parametru (σs) spowoduje, że komórki, między którymi odległość jest mniejsza niż wartość progu supresji, będą w procesie uczenia sieci usuwane. Usunięcie tych komórek spowoduje, że sieć będzie bardziej plastyczna.

Na rys. 3 został przedstawiony rezultat procesu uczenia. Liczbę „dolin" na wykresie słupkowym należy skojarzyć z liczbą klas, na które zostały podzielone szeregi czasowe należące do zbioru uczącego.

Zmniejszenie wartości parametru (σs) będzie skutkowało, większą liczbą skupień – liczbą klas, na które zostaną podzielone dane należące do zbioru uczącego. Zmniejszenie wartości współczynnika supresji skutkuje tym, że sieć dokonuje podziału na coraz bardziej subtelne klasy, ale tym samym jest mniej plastyczna (rys. 4).

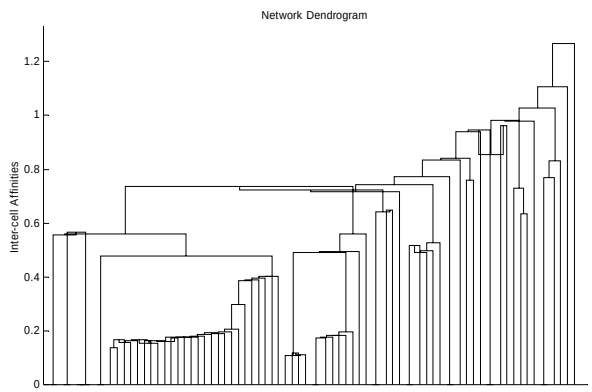


Rys. 3. Wykres słupkowy zbudowany w procesie uczenia dla parametru, $\sigma s = 0.2$

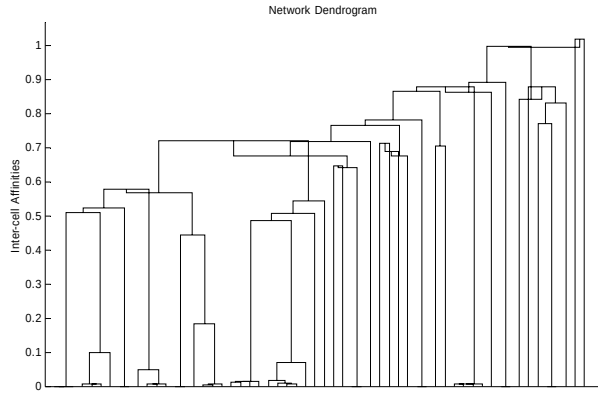


Rys. 4. Wykres słupkowy zbudowany w procesie uczenia dla parametru, $\sigma s = 0.01$

Na rys. 5 i rys. 6 przedstawiono dendogram po procesie uczenia sieci dla parametru $\sigma s = 0.2$ *i* dla parametru $\sigma s = 0.01$, ilustrujący podział na klasy dane należące do zbioru uczącego.



Rys. 5. Dendogram dla parametru, $\sigma s = 0.2$

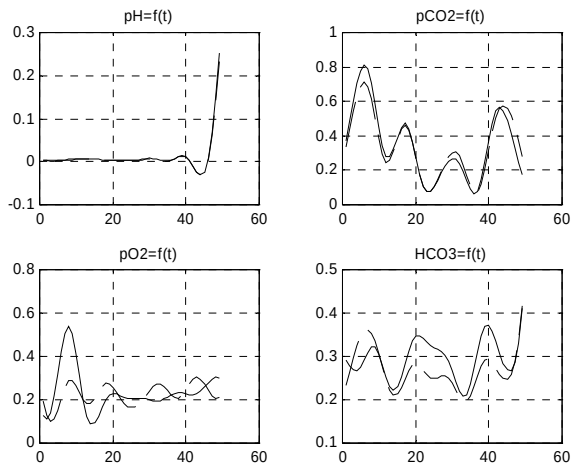


Rys. 6. Dendogram dla parametru, $\sigma s = 0.01$

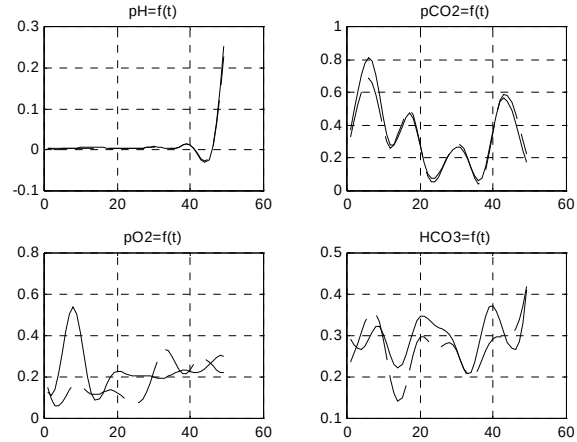
Po przeprowadzeniu procesu uczenia sieci immunologicznej można przejść do drugiej fazy eksperymentu, polegającej na sprawdzeniu zdolności sieci do rozpoznawania przebiegu (generalizacji). Na wejście nauczonej sieci będą podawane przebiegi, które nie należały do zbioru trenującego. W ten sposób będzie obserwowana odpowiedź sieci.

3.2. Proces testowania

Na rys.7 i rys.8 została przedstawiona odpowiedź sieci dla wybranego przebiegu ze zbioru testującego.



Rys. 7. Rezultat klasyfikacji sieci sygnału wejściowego dla parametru, $\sigma s = 0.2$



Rys. 8. Rezultat klasyfikacji sieci sygnału wejściowego dla parametru, $\sigma s = 0.01$

Linia przerywana na przedstawionym wykresie jest to najlepsza odpowiedź sieci na zadany przebieg wejściowy. Odpowiedź w tym sensie jest najlepsza, ponieważ sygnał wejściowy sieci traktowany jest przez sieć jako antygen. Powinowactwo sygnału wejściowego – antygenu jest obliczane dla każdego węzła sieci – przeciwciała. Wszystkie przeciwciała są w różnym stopniu aktywowane przez dany antygen.

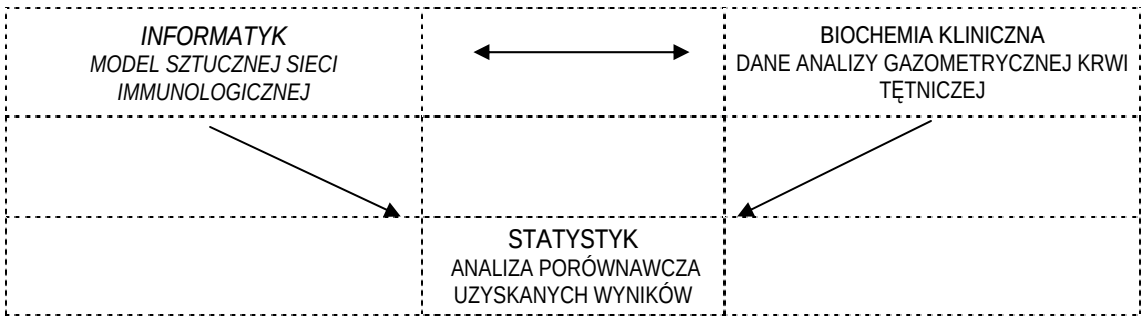
4. Podsumowanie

W artykule tym zostało przedstawione funkcjonowanie systemu immunologicznego człowieka, ze szczególnym uwzględnieniem mechanizmów, które obecnie są wykorzystywane w sztucznych systemach immunologicznych, tj. algorytmu uczenia sztucznej sieci immunologicznej.

Został przedstawiony również przykład zastosowania sztucznej sieci immunologicznej do klasyfikacji wielowymiarowych danych o charakterze medycznym.

Literatura cytowana

1. De Castro L. N., Von Zuben F. J.: An Evolutionary Immune Network for Data Clustering, Proc. of the IEEE SBRN, pp. 84-89, 2000.
2. De Castro L. N., Von Zuben F. J.: The Clonal Selection Algorithm with Engineering Applications, GECCO'00 – Workshop Proceedings, pp. 36-37, 2000.
3. De Castro L. N., Von Zuben F. J.: Artificial Immune Systems: Part I – Basic Theory and Applications, Technical Report – RT DCA 01/99, p. 95, 1999.
4. Wierzchoń S. T.: Sztuczne systemy immunologiczne – teoria i zastosowanie, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2001.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 57-60.

ABDUCTIVE LOGIC PROGRAMMING IN ALLERGY DIAGNOSIS

WNIOSKOWANIE ABDUKCYJNE W DIAGNOZIE ALERGII

PIOTR WALECKI[*], WOJCIECH LASOŃ[*], GRZEGORZ PORĘBSKI[**], KRYSTYNA OBTUŁOWICZ [**]

[*]*Zakład Bioinformatyki i Telemedycyny, Collegium Medicum Uniwersytetu Jagiellońskiego, 31-501 Kraków, ul. Kopernika 7e*
[**]*Katedra Toksykologii Klinicznej i Środowiskowej, Zakład Alergologii Klinicznej i Środowiskowej, Collegium Medicum Uniwersytetu Jagiellońskiego, 31-531 Kraków, ul. Śniadeckich 10*

Abstract. This paper presents an abductive model of diagnosis, which was designed for medical care (diagnosis and treatment) of certain types of allergy.
The reasoning is defined to have three forms: deduction, induction and abduction.
The American philosopher Charles S. Peirce was the first to notice abduction. It was characterized as the probative adoption of a hypothesis that explains observed facts (results) in accordance with known laws.
Abduction, as cognitive operation, creates the framework, which makes it possible to attribute a singular symptom to a disease. The interpretation of symptoms is always abductive. H.R. Fisher said that every medical diagnostic statement functions as an abductive inference.
The abductive schema will become a (nomologically-deductive) principle, if the first premise (the result) contains an unexplained surprising fact (perception, observation) – a symptom in medical diagnosis, which is then explained casually as the effect of a cause.
The abductive mode of interference involves two steps. In the first step a fact (a "result" in Peirce's terminology) to be explained or understood is presented, as the derived conclusion in classical schema. The second step introduces an available or newly constructed hypothesis (rule/law) by means of which the case is abducted. Knowledge becomes intelligible by way of its abductive incorporation into a coding system (i.e. medical classification) the logic, which the frame forms by means of new facts.

Key words: inference, abduction, medical diagnosis

Streszczenie: W pracy zaprezentowano model diagnozy lekarskiej bazujący na wnioskowaniu abdukcyjnym. Model ten został zaprojektowany w celu wspomagania diagnozy i terapii w pewnych przypadkach alergii.
Rozumowanie jest definiowane przez trzy podstawowe typy wnioskowania: dedukcję, indukcje i abdukcję.
Pierwszym, który wyróżnił i opisał wnioskowanie abdukcyjne był amerykański filozof Charles S. Peirce. Scharakteryzował je jako metodę dochodzenia do hipotezy, która najlepiej wyjaśnia obserwowane fakty w zgodzie ze znanymi prawami.
Abdukcja, jako wnioskowanie logiczne tworzy ramy odniesienia dzięki którym możliwe staje się przypisanie pojedynczego objawu do konkretnej jednostki chorobowej. H.R. Fisher stwierdził, że każda opinia diagnostyczna w medycynie bazuje na wnioskowaniu abdukcyjnym.
Jeśli początkowa przesłanka zawiera niewyjaśniony zaskakujący fakt – objaw w medycznej diagnozie, który spontanicznie tłumaczy się jako efekt określonej przyczyny, to wtedy wykorzystujemy schemat wnioskowania abdukcyjnego.
Abdukcyjny sposób wnioskowania składa się z dwóch etapów. W pierwszym etapie fakt („result" w terminologii Pierce'a), aby został wytłumaczony czy zrozumiany, zostaje potraktowany, jako konkluzja w klasycznym schemacie inferencji. Drugim etapem jest wprowadzenie dostępnych lub na nowo skonstruowanych hipotez (zasad/praw), za pomocą których dany przypadek jest abdukowany. Objawy wyjaśniane są poprzez ich abdukcyjne włączenie w kodujący system (np. klasyfikacja medyczna) logiki, której postać formuje się dzięki nowym faktom.

Słowa kluczowe: wnioskowanie, abdukcja, diagnoza medyczna

1. Introduction

Allergy is the most important health problem nowadays and unfortunately it continues to escalate. In the next few years almost fifty percent of population will be suffering from that disease. Many people are already sensitised for various reasons. To obtain a diagnosis, they should consult with their physician or an allergist/immunologist, a physician specially trained in this area. Numerous cases must be permanently monitored. This has to require huge expenditures, therefore there have been working out computer systems, which would aid diagnostic and therapeutic process.

Aspirin (acetylsalicylic acid) is one of the most broadly used drugs in the world. It is applied as analgesic, antipyretic, and as a coronary disease prevention. In patients, the aspirin causes hypersensitivity symptoms – asthmatic or angioedema and urticaria symptoms. In the face of wide-spreading of this drug it seems adequate to create freely available telemedical

system. To provide every interested party with information on the subject of hypersensitivity symptoms about aspirin and to perform self prediagnosis – is the aim of such system. The information about dermatological lesion hypersensitivity was colleted by conducting: 1) a search out in Medline and EMBASE medical database for key words “aspirin”, “urticaria” and “angioedema”; 2) information analysis current publications; 3) a consultation with specialist clinical centers [1].

A pilot telemedical system for information and diagnosis of skin hypersensitivity to aspirin was created, which on the day of insertion into the system allows (in autodiagnostic mode) differentiation of causes of the oedema, and later allows obtaining information about next management (low-salicylic diet, cross-reaction with nonsteroidal anti-inflammatory drugs, treatment of exacerbation, medical centers specialized in management of aspirin idiosyncrasy). The widespread using of the acetylsalicylic acid and the frequent angioedema episode in general population gives reasons creation common-accessible information sources for health professionals [2].

We projected low-cost Internet-based diagnosis support system for certain types of allergy. The telemedical system for information and diagnosis may be seen as a very sophisticated portal that allows access to data, and to allow its meaningful input into expert decisions. The proposed method was evaluated on a medical database, the experimental results of which show that abducted (Abductive Logic Programming) rules correctly represent experts’ decision processes. According to the rules acquired from medical experts, medical differential diagnosis is a focusing mechanism: first, medical experts focus on some general category of diseases, such as asthma or angioedema. After excluding the possibility of other categories, medical experts proceed into the further differential diagnosis between diseases within a general category. The above experiments show that rule abduction with this grouping generates rules, which are similar to medical experts’ rules and they suggest that our proposed method should capture medical experts’ reasoning. In order to determine a probability of successful diagnosis as a function of clinical data and experience in which a diagnosis is refined and updated based upon subsequent medical data obtained for a particular patient or subsequent additions and updates to the collection of clinical medical data, Abductive Logic Programming (ALP) based methods are utilized.

2. Abductive Logic Programming (ALP)

H. R. Fisher said: “Every diagnostic statement by a medical functions as an abductive inference” [3]. The American philosopher Charles S. Peirce was the first to notice abduction. It was characterized as the probative adoption of a hypothesis that explains observed facts (results) in accordance with known laws.

Peirce introduced the new kind of inference as reasoning *a posteriori*, thus setting it apart from deduction *a priori*. Abduction is “the process of forming hypothesis. It is only logical operation, which introduces any new idea; for induction does nothing but determine a value, and deduction merely evolves the necessary consequences of a pure hypothesis. Deduction proves that something must be; induction shows that something actually is operative; abduction merely suggests that something may be” [4]. The reasoning is defined to have three forms among which locate abduction. General rules applied to particular cases with the inference of a result feed analytic process of reasoning known as deduction. A form of synthetic process distinguished as induction infers the rule from the case and the result, and abduction – another form of synthetic inference, but of the case from a rule and a result [5].

- Deduction is an analytic process based on the application of general rules to particular cases, with the inference as a result.
- Abduction is synthetic reasoning which infers a case (or a fact) from the rules and the result.
- Induction is synthetic reasoning which infers a rule from the case and the result.

The Peirce’s conception of inference is presented in Fig. 1. First column contain a schema of the classic syllogism, and in round brackets are Peirce’s terms. The categorical syllogism relates three concepts, S (subject), P (predicate), and M (middle), in three statements (major premise, minor premise, conclusion) in order to examine their validity. Next column contain a schema of three distinguished kinds of inference. Boxes with continuous lines contain premises/hypothesis that are presupposed as given/true. Boxes with dotted lines contain hypotheses that are inferred. Numbers in round brackets are sequence of inferring. Deduction is analytically true (redundant) and was considered to be merely an “explanatory statement”. Deductive conclusion completely contained in the premises, therefore it is not synthetic (content-increasing), does not lead to new knowledge. In the case of induction the premises are observational statements, and an inferred conclusion is considered to be content-increasing but not truth-conserving, because it is only a hypothesis that cannot be proved with ultimate certainty [6]. In Peirce’s description: induction – the discovery of laws, abduction (hypothetic inference) – the discovery of causes, deduction – the prediction of effects.

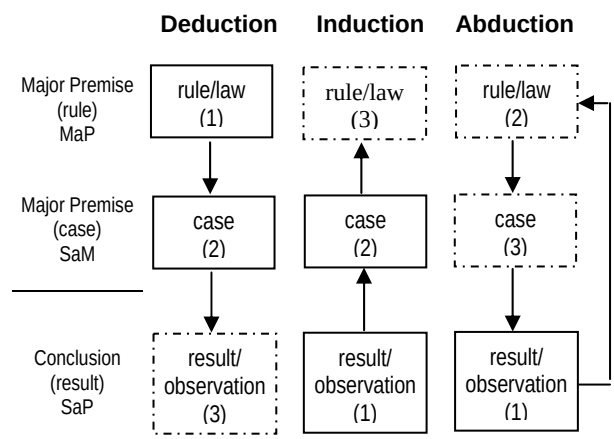


Fig. 1. Forms of inference [6].

The abductive schema will become a (nomologically-deductive) principle, if the first premise (the result) contains an unexplained surprising fact (perception, observation) – a symptom in medical diagnosis, which is then explained casually as the effect of a cause. The abductive mode of interference involves two steps. In the first step a fact (a

"result" in Peirce's terminology) to be explained or understood is presented, as the derived conclusion in classical schema. The second step introduces an available or newly constructed hypothesis (rule/law) by means of which the case is abducted. Knowledge becomes intelligible by way of its abductive incorporation into a coding system (i.e. medical classification) the logic, which the frame forms by means of new facts. Knowledge becomes intelligible by way of its abductive incorporation into a coding system (i.e. medical classification) the logic of with form the frame within witch the facts (phenomena) acquire meaning by virtue of having become signs [6].

Abductive inference in medical diagnosis is shown in Fig. 2. Abduction, as cognitive operation, creates the framework, which makes it possible to attribute singular symptom to disease. The interpretation of symptoms is always abductive. Eco wrote: "Abduction, therefore, is the experimental and risk-laden searching-out of a system of rules of signification which make it possible for a sign to gain its meaning(s)" [7].

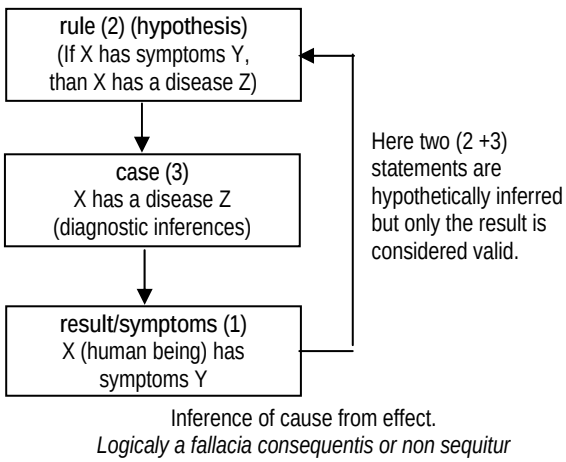


Fig. 2. Abductive inferences in diagnosis – an example of a medical diagnosis [6].

The abduction in its most general form sounds like this: is given a background theory T and an observation q to explain, to find an explanation theory Π such that $\Pi \cup T \models q$. Some additional conditions want normally to be put on Π, that it is consistent with T and that contains only those proportions called abducibles. For instance in prepositional logic, given a background theory T, a set A of assumptions or abducibles, and a proposition q, an explanation S of q is commonly defined to be a minimal set of literals over A such that $\Pi \cup T \models q$ and $\Pi \cup T$ is consistent [8].

A hypothesis concerning the presence or absence of disorders as a faults or abnormal processes solve a diagnostic problem in medicine. The research on formal aspects of diagnosis, casual knowledge represented by means of a binary casual relation, which is employed for computing diagnoses, have been undertaken with the help of logic as primary tool. In the logical theory of abductive diagnosis, identification of disease is formalized as reasoning from effects to causes, with casual knowledge implicated in the form: *causes* → *effects*, where causes are usually abnormalities or faults, or they may also include normal situations. This abductive type of reasoning is contrasted with deduction, which for implications of the form above and under certain conditions, like that given causes and effects are conjunctions of positive literals, would amount to reasoning from *causes* to *effects* [11]. The logical theory of abductive diagnosis links to set-covering theory, because casual relations are also use to find causes for certain observed findings, however at the same time, by possible explicit representation of various types of interaction, is more expressive [12].

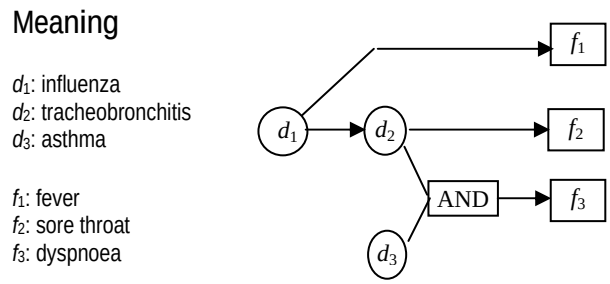


Fig. 3. Causal net [16].

Consider the following piece of medical knowledge: "influenza causes fever and infection of the trachea and bronchial tree (tracheobronchitis), which causes a sore throat; if the patient suffers from asthma, shortness of breath (dyspnoea) will occur as well". In Fig. 3, the directed-graph representation of the causal knowledge embodied in this medical description is depicted, where an arc denotes a cause-effect relationship. The medical meaning ascribed to the elements in the causal graph is indicated in Fig. 3. Using logic as our representation language, the figure may be assumed to correspond to a *causal specification* $C = (\Delta, \Phi, R)$, were $\Delta = \{d_1, d_2, d_3\}$ denotes to set of disorders, $\Phi = \{f_1, f_2, f_3\}$ denotes a set of observable findings, and R is a collection of rules in prepositional logic: $d_1 \rightarrow d_2$, $d_1 \rightarrow f_1$, $d_2 \rightarrow f_2$, $d_1 \wedge d_3 \rightarrow f_3$. Such a causal specification is typically used in medical diagnosis based on the ALP [15]. Note that the disorders d_1 and d_2 are causally related to each other; causal interaction is expressed by logical implication. A causal specification can be used for predicting observable findings. Assuming the presence of certain disorders, e.g., influenza (d_1), $R \cup \{d_1\} \models \{f_1, f_2\}$ expresses that a patient with influenza will have symptoms and signs of fever (f_1) and sore throat (f_2) via a causal mechanism, where $\models$ denotes standard logical entailment. Here, the interaction between disorders, and between disorders and observable findings, is monotonic, due to the monotonic nature of ordinary logical entailment: by assuming more disorders, more observable findings will be predicted. Knowledge of normal and abnormal functional behaviour can be effective for diagnosing device problems, where the behaviour of the device is described in terms of relationships between input and output signals. These relationships are obtained from knowledge of the behaviour of the device's components and of the way in which these components are interconnected, i.e., the structure of the device [16].

3. Conclusions

The telemedical diagnosis support is a knowledge-based system, developed by the Department of Bioinformatics and Telemedicine and Department of Allergy and Immunology at the Medical College Jagellonian University, which advises on the treatment of certain types of allergy [21-25]. An application of abduction in inference systems brings a new scope to the structure of our knowledge. This was first noticed by C. S. Peirce, but the importance of this conception for mental word-making rediscovered H. R. Fisher. He wrote: 'the (potential) confirmation of such hypotheses (logical inferences or theories) says nothing about reality (ontology) in the sense of a representation or mapping but only that the hypotheses are functioning. Therefore, the structures of our logic(s) or our theories do not mirror the structures of things nor are they derived (deduced) from them. With the primary act of the formulation of such a hypothesis a paradigm, a rule, a method of measurement is provisionally laid down by means of which we then *measure* or *compare* what we conceive of as *nature*. Only in a second act arise the notions of correct/incorrect, fitting/non-fitting, rule-conforming/rule-breaking etc. The actual constructive act consist in the a priori specification of a *measuring method* by the cognizing subject (the scientific community) because the world cannot determine for us directly what kind of measuring method or paradigm we must use. And just as the measuring method must be specified before measuring can take place, so induction must be directed/governed/controlled by hypothesis that has been constructed abductiviely beforehand' [18].

4. References

1. Porebski G, Obtulowicz K, Obtulowicz P, Walecki P, Roterman-Konieczna I, Czarnobilska E, Krzanowska K, Kwapińska A, Targosz K.: Internet as a potential source of information about allergy for patients, The XIXth World Allergy Organisation Congress, June 26 - July 1, 2005, Munich, Germany.
2. Porębski G., Walecki P., Obtułowicz K., Czarnobilska E., Krzanowska K., Kwapińska A., Targosz K., Częstość wykorzystywania różnych źródeł wiedzy o alergii przez chorych ze schorzeniami atopowymi, Allergy Asthma Immunology 7th Central and Eastern European Conference June 8-11, 2005 Łódź, Poland.
3. Fisher H. R.: Abductive Reasoning As A Way Of Worldmaking, in Foudations of Science, 6, Kluwer Academic Publishers, Brüssel, 2001, p.371.
4. Peirce C. S.: Collected Papers Of Charles Sanders Peirce, Harvard University Press, 1958-1966.
5. Cox P. T., Pietrzykowski T.: General diagnosis by abductive inference, In Proceedings IEEE Symposium on Logic Programming, 1987, pp.183-189.
6. Fisher H. R.: ibidem, pp.361-372.
7. Eco U.: Semiotics and the Philosophy of Language, Indiana University Press, 1986.
8. Lin F., You J.-H.: Abduction in logic programming, In Artificial Intelligence, 140, 2002, pp.175-205.
9. Kakas A., Kowalski R., Toni F.: The role of abduction in logic programming, In Handbook of Logic in Artificial Intelligence and Logic Programming, Oxford University, 1995.
10. Console L., Torraso P.: Integrating models of correct behavior into abductive diagnosis, In Proceedings ECAI-90, pp. 160-166.
11. Lucas P. J. F.: Structures in diagnosis: from theory to medical application. Amsterdam 1996.
12. Poole D.: A methodology for using a default and abductive reasoning system, In International Journal of Intelligent Systems, 5, 1990, pp.551-548.
13. Lucas P. J. F.: Logic engineering in medicine, In The Knowledge Engineering Review, 12 (2), 1995, pp.153-179.
14. Lucas P. J. F.: Analysis of notions of diagnosis, In Artificial Intelligence, 105, 1998, pp.295-343.
15. Poole D.: Representing diagnosing knowledge, In Ann. Math. Artificial Intelligence, 11, 1994, pp.33-50.
16. Lucas P. J. F.: Analysis of notions of diagnosis, In Artificial Intelligence, 105, 1998, pp.298-300.
17. Sun Z., Finnie G.: Weber K., Integration of abductive CBR and deductive CBR, In FUZZ-IEEE 2001, Melbourne, pp.1432-35.
18. Fisher H. R., ibidem, p.370-371.
19. Peng Y., Reggia J.A.: Abductive Inference Models for Diagnostic Problem Solving, Springer, New York, 1990.
20. Josephson J. R.: Josephson S.G., Abductive Inference. Computation, Philosophy, Technology, Cambridge University Press, 1996, p. 6.
21. Walecki P., Lasoń W., Wiśniowski Z., Roterman-Konieczna I, Abductive logic programming in medical diagnosis, Journal of Artificial Intelligence, Vol. 1, No 2(24): 107-115, 2004.
22. Sarapata K., Usarz M., Trąbka J., Walecki P., Pyrczak W., Roterman-Konieczna I, Security of net servers in biological approach, Journal of Artificial Intelligence, Vol. 1, No 1(23): 31-38, 2004.
23. Walecki P., Lasoń W., Porębski G., Obtułowicz K., Abductive logic programming and medical diagnosis. MSCB 2005, p. 76.
24. Walecki P., Nonmonotonic logic as a formalized model of human thinking. Associative and interpretative representational theory of mind, MSCB 2005, p. 44.
25. Walecki P., Lasoń W., Porębski G., Obtułowicz K., Skórna postać nadwrażliwości na aspirynę - pilotażowy telemedyczny system informacyjno-diagnostyczny, MSCB 2005, p. 37.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 61-64.

SYSTEMOWY ALGORYTM EWOLUCYJNY (SAE)

SYSTEM'S EVOLVING ALGORITHM

JERZY RUDOLF TCHÓRZEWSKI

Zakład Sztucznej Inteligencji, Instytut Informatyki, Wydział Nauk Ścisłych, Akademia Podlaska, jtchorzewski@interia.pl, jerzy.tchorzewski@ii3.ap.siedlce.pl

Streszczenie. Praca zawiera propozycję nowej metody z grupy algorytmów genetycznych i ewolucyjnych nazwanej systemowym algorytmem ewolucyjnym (SAE). Klasyczne już algorytmy genetyczne i ewolucyjne nie zawierają na przykład metody prowadzącej do otrzymywania pierwotnej populacji wprost z systemu oraz adekwatnego sposobu na otrzymywanie nowej populacji (odpowiednio zdefiniowanej funkcji przystosowania). Artykuł zawiera ponadto szereg definicji takich jak: strukturalne i parametryczne zmiany, kod informacyjny rozwoju systemu, rozbieżności systemowe pomiędzy systemem a otoczeniem (uogólnienie pojęcia funkcji przystosowania), mapy wiedzy o rozwoju systemu itp.

Słowa kluczowe: metody sztucznej inteligencji, algorytmy genetyczne, sztuczny kod genetyczny, funkcja przystosowania, mapy wiedzy

Abstract. This paper contains proposition of new method of evolving algorithm called evolving algorithm of system (SAE). Classical algorithms are not, for example, furnished with a method of getting population directly from systems and adequacy methods to qualify artificial genetic code to new populations. The article presents the way to produce genetic code directly from systems and also defines steps leading to new population. Besides this, the article presents definitions of structural and parametric changes, information code of systems, divergence between system and environment (special fitness function), knowledge maps on system development, and others.

Key words: artificial intelligence methods, genetic algorithms, artificial intelligence code, fitness function, knowledge maps

1. Definicje podstawowych pojęć

Na wstępie zostaną zdefiniowane podstawowe pojęcia dla potrzeb opracowania systemowego algorytmu ewolucyjnego. Jako bazę wyjściową przyjmuje się teorie sterowania oraz inżynierię systemów, w tym zwłaszcza inżynierię rozwoju systemów [3], przy czym systemem działania jest każdy przetwornik organizacyjny dokonujący przemiany strumieni wejściowych na strumienie wyjściowe. Na tej bazie można określić rozwój jako działanie, w którym występuje ewolucyjne sprzężenie zwrotne, a ściślej – ewoluujące sprzężenie zwrotne dodatnie. Taka definicja rozwoju wprowadza rzecz zupełnie nową do teorii algorytmów ewolucyjnych – przejście od opisu systemów w czasie krótkim do opisu systemów w czasie długim. A zatem przez rozwój rozumiane jest działanie dokonujące przemiany, a nawet transformacji systemu, co zwykle wiąże się ze zmianami nie tylko parametrów, lecz również struktury systemu. Należy tutaj dodatkowo zauważyć, ze sam rozwój może być systemem, dlatego można mówić także o systemach rozwoju obok rozważania aspektów rozwoju systemów. W przypadku rozpatrywania rozwoju systemów w kategoriach rozwoju i cybernetyki warto zauważyć, że rozwój wynika z jednej strony z natury systemu (z jego organizacji wewnętrznej, wewnętrznego uorganizowania systemu), a z drugiej strony z natury systemów oddziałujących na dany system. A zatem systemy rozwoju charakteryzują się pewną strukturą i parametrami. Zachodzi w nich proces przemiany strumieni organizacyjnych wejściowych na strumienie organizacyjne wyjściowe. Na wejściu systemu występuje potrzeba dopływu konkretnego systemu, a na jego wyjściu możliwość jego generowania. Potrzebami i możliwościami dysponują wszystkie systemy z danej sytuacji systemowej. Można przyjąć potrzebę systemu jako stan ujemny, a możliwość jako stan dodatni i w ślad za tym mówić o napięciu systemowym występującym pomiędzy odpowiednimi potrzebami i możliwościami systemów współpracujących ze sobą. W przypadku braku pokrycia potrzeb przez możliwości systemowe pojawiają się rozbieżności systemowe [3], które mogą dotyczyć stanu napięcia systemowego pomiędzy możliwościami systemu a potrzebami otoczenia oraz stanu napięcia pomiędzy potrzebami systemu a możliwościami otoczenia.

Jeżeli zmianom ulegają jedynie parametry strumieni organizacyjnych oraz parametry przetworników organizacyjnych, a liczba i typ strumieni oraz struktura i typy przetworników nie ulegają zmianom, to taki rozwój systemu jest rozwojem parametrycznym. A zatem przy rozwoju parametrycznym, sterując parametrami systemu dąży się do zmniejszenia rozbieżności systemowych, a tym samym do zmniejszenia napięć organizacyjnych.

Natomiast w przypadku pojawienia się w systemie nowego strumienia organizacyjnego lub nowego przetwornika organizacyjnego (a w konsekwencji innych elementów systemu, włącznie z podsystemami) lub też odpowiednio zanik, to wówczas występuje rozwój strukturalny. A zatem w przypadku rozwoju strukturalnego sterując strukturą systemu,dąży się do zmniejszenia rozbieżności systemowych, a tym samym dąży się do zmniejszenia napięć organizacyjnych.

2. Sztuczny kod genetyczny

Do modelu systemu prowadzą przynajmniej dwie drogi:

- pierwsza – poprzez opis systemu przy pomocy modelu matematycznego (gdy znane są prawa lub sposoby opisu obiektu przy pomocy zależności matematycznych),
- druga – poprzez identyfikację systemu (gdy nie w pełni lub w ogóle nie są znane sposoby opisu systemu zależnościami matematycznymi).

W obu sytuacjach dąży się do uzyskania sedna („jądra") wiedzy o systemie i ujęcia jej w postaci modelu. Model systemu może być przy tym zapisany w odpowiedniej postaci, a więc np. w postaci modelu ciągłego lub dyskretnego, modelu w dziedzinie czasu lub częstotliwości, w postaci np. transmitancyjnej, zero-biegunowej lub równań stanu itp.

A zatem model zawiera całą wiedzą o systemie, która jest w nim zapisana w postaci jawnej lub zagregowanej, a nawet ukrytej. Na podstawie wiedzy zakodowanej w modelu można przy tym poszukiwać swoistej informacji o systemie i jego rozwoju na podobieństwo informacji zawartej w kodzie genetycznym organizmów żywych. Kod informacyjny występujący w modelu systemu nazwano *sztucznym kodem genetycznym*, opracowując wygodną metodę jego wyprowadzania wprost z modelu systemu. Taki sztuczny kod genetyczny w terminologii teorii sterowania i systemów może zostać zdefiniowany jako wektor:

$$K_g = (g_1, g_2, ..., g_n), \qquad (1)$$

gdzie:

g_i – i-ty *sztuczny gen*, który może mieć dalej postać również sztucznego kodu genetycznego.

A zatem w systemowym algorytmie ewolucyjnym mogą występować *sztuczne kody genetyczne* w postaci wektorów, macierzy lub tablic wyższego wymiaru w postaci swoistego zanurzenia sztucznego kodu genetycznego w innym sztucznym kodzie genetycznym. Takie rozumowanie prowadzi do inicjacji populacji początkowej dla algorytmu ewolucyjnego nie w postaci losowej, lecz na podstawie wiedzy wydobywanej wprost z modelu. Mogą być przy tym wykorzystywane elementy występujące w liczniku oraz w mianowniku modelu zapisanego w postaci transmitancyjnej, mogą to być zera i bieguny wynikające z zapisu zero-biegunowego, mogą to być wreszcie macierze *A, B, C, D* lub zawarta w nich wiedza wynikająca z zapisu w postaci równań stanu, jak np. współczynniki równania charakterystycznego lub jego pierwiastki.

Przykład:

1) *Zapis sztucznego kodu genetycznego w postaci współczynnikowej:*

$$K_w = (v; a_1, a_2, ..., a_n; b_1, b_2, ...,b_m), \qquad (2)$$

gdzie:

v – stały współczynnik, od którego zależą zmiany parametryczne systemu,

a_i – współczynnik występujący przy i-tym wyrazie wielomianu licznika transmitancji operatorowej,

b_j – współczynnik występujący przy j-tym wyrazie wielomianu mianownika transmitancji operatorowej.

2) *Zapis sztucznego kodu genetycznego w postaci pierwiastkowej:*

$$K_p = (k; z_1, z_2, ..., z_n; p_1, p_2, ..., p_m), \qquad (3)$$

gdzie:

k – stały współczynnik, od którego zależą zmiany parametryczne,

z_i – i-ty pierwiastek licznika transmitancji operatorowej (i-te zero),

p_j – j-ty pierwiastek mianownika transmitancji operatorowej (j-ty biegun).

Uwaga:

Elementy sztucznego kodu genetycznego zapisane w postaci wartości (np. liczb) mogą być traktowane jako pierwotne sztuczne geny, a elementy sztucznego kodu genetycznego zapisane w postaci formuł matematycznych mogą być traktowane jako podchromosomy, mające podobna strukturę jak chromosom nadrzędny. W takiej sytuacji występuje zanurzenie jednego chromosomu w innym, a wartości chromosomów mogą być ustalane po wygenerowaniu populacji początkowej chromosomów.

Przykład liczbowy:

Niech sztuczny kod genetyczny posiada następującą postać:

$$K_w = (1, 0, 525, 35168, 28804) \qquad (4)$$

Przykład symboliczny:

$$\begin{aligned} &K = (ch(G), ch(L), ch(O)\ , \\ &ch(G) = (P, Q, U, Y), \qquad (5) \\ &ch(L) = (X_{ij}, Y_{ij}, X1_{ij}, Y1_{ij}), \\ &ch(O) = (P, Q, U, Y), \end{aligned}$$

gdzie: P, Q, U, Y, X_{ij}, Y_{ij}, $X1_{ij}$, $Y1_{ij}$ – odpowiednie parametry.

W takiej sytuacji pojedynczy chromosom będzie więc występował w następującej postaci macierzowej:

$$Ch = [P, Q, U, Y; X_{ij}, Y_{ij}, X1_{ij}, Y1_{ij}\ ; P, Q, U, Y] \qquad (6)$$

3. Systemowy algorytm ewolucyjny (SAE)

Zdefiniowany w p.1 sztuczny kod genetyczny można dalej wykorzystać do wygenerowania populacji początkowej chromosomów (oraz chromosomów z zanurzeniami), a następnie poddać populację początkową działaniu systemowego algorytmu ewolucyjnego [1], [5], [6], [7], [8], który, w uproszczeniu, można sprowadzić do następujących podstawowych kroków:

Krok pierwszy:

Inicjacja populacji początkowej. Należy przewidzieć możliwość wygenerowania wszystkich możliwych chromosomów wynikających ze sztucznego kodu genetycznego.

W przypadku chromosomów bez zanurzenia należy generować geny w zakresie dopuszczalnej tolerancji zmienności poszczególnych parametrów składających się na poszczególne współczynniki lub pierwiastki np. równania charakterystycznego, a w przypadku chromosomów z zanurzeniem należy najpierw umożliwić wygenerowanie różnych struktur chromosomów, a potem zbiorów ich wartości. Ten etap jest najbardziej istotny w algorytmie SAE, dlatego powinna wystąpić duża dokładność obliczeń.

Krok drugi:

Inicjacja tzw. rozbieżności systemowych, a więc niedopasowań w systemie występujących pomiędzy wyjściami systemu a oczekiwaniami otoczenia oraz pomiędzy wejściami systemu a możliwościami ich wygenerowania przez otoczenie. Rozbieżności będzie zatem tyle, ile wystąpi łącznie wejść i wyjść systemowych, należy więc stworzyć tablicę wyboru modelu systemu (1-1, 2-1, 2-2, 1-2,..., aż do n wejść oraz m wyjść, przy czym parametry n oraz m powinny być określane przez użytkownika aplikacji.

Ogólna zależność dotycząca rozbieżności:

$$\Delta S = \left| WE/WY\ (SYS) - WE/WY(OTO) \right|, \qquad (7)$$

przy czym w rozbieżnościach należy wyróżnić funkcję celu oraz funkcję przystosowania chromosomu. W przypadku wyjść systemowych funkcja celu wynika z otoczenia, a funkcja przystosowania z systemu. Natomiast w przypadku wejść jest odwrotnie. Funkcja celu wynika z systemu, a funkcja przystosowania z otoczenia. To determinuje potraktowanie otoczenia jako systemu i konieczność inicjacji populacji początkowej reprezentującej otoczenie na identycznej zasadzie, jak w przypadku systemu.

Zależności definiujące *funkcję przystosowania* oraz *funkcję celu* powinny więc wynikać wprost z modelu systemu oraz wprost z modelu otoczenia. W przypadku identyfikacji po stronie wyjścia *funkcją przystosowania* może być np. w przestrzeni stanów równanie wyjścia, np.:

$$y = C\,x + D\,u + K\,e, \qquad (8)$$

a *funkcją celu* może być np. funkcja potrzeb wyrażona w postaci sygnałów wyjściowych z konkretnego systemu. W praktycznych przypadkach może to być np. następująco wyrażona funkcja:

funkcja celu = wyjście y z modelu (*funkcja przystosowania*) $\pm$ błąd wynikający z identyfikacji systemu (różnica pomiędzy systemem a modelem).

Natomiast po stronie wejściowej systemu sytuacja będzie podobna tylko w odniesieniu nie do systemu, lecz do otoczenia.

Krok trzeci:

Zastosowanie zmodyfikowanych operatorów genetycznych, a więc m.in. krzyżowania (należy umożliwić przy tym wybór odpowiedniej metody krzyżowania, jak np. krzyżowanie wielopunktowe, krzyżowanie z maską, krzyżowanie segmentowe, krzyżowanie mieszające itp. oraz *zmodyfikowany operator krzyżowania*, a więc zmiana struktury modelu, czyli dopuszczenie pojawienia się nowego współczynnika lub pierwiastka itp.), zmodyfikowana metoda mutacji (należy umożliwić wybór odpowiedniej metody klasycznej np. zmiana jednego z dowolnych genów, wstawienie nowego genu itp. oraz zmodyfikowany operator mutacji np. wprowadzenie zmian parametrów w wybranym jednym lub kilku genach (w zakresie dopuszczalnych zmian wartości wynikających z realizowalności systemu np. w tym samym rzędzie wartości), itp. oraz zmodyfikowana metoda selekcji (umożliwić wybór odpowiedniej metody klasycznej selekcji oraz zmodyfikowanej selekcji wynikającej np. z warunków stabilności modelu, a także stabilności poszczególnych chromosomów w przypadku zanurzenia – tzn. stabilność podsystemów, zastosować odpowiednio zmodyfikowaną metodę procedury na zatrzymanie SAE np. wygenerowanie nowego chromosomu bardziej „krzepkiego" w stosunku do chromosomów z populacji początkowej itp.

Krok czwarty:

Przejście na drugi, trzeci i ewentualnie dalsze poziomy szczegółowości chromosomów, a więc kolejno każdy gen chromosomu głównego jest też chromosomem, a w nim znów jego geny są chromosomami (analogia do systemu i podsystemów, które są również systemami, zatem posiadają swoje podsystemy), przy czym kolejne poziomy może obsługiwać ten sam algorytm SAE.

Ogólny algorytm rozwoju systemu rynku energii elektrycznej można sprowadzić do następujących podstawowych faz, a mianowicie:

(1) identyfikacja struktury prakseologicznej istniejącego systemu rynku energii elektrycznej (IRE) ze względu na relację zabezpieczenia, w celu zbudowania modelu prakseologicznego rynku energii elektrycznej (MP),

(2) wybór przedziału systemowego $\{S_{i-1}, S_i, S_{i+1}\}$ oraz przeprowadzenie w nim identyfikacji rozbieżności systemowych we wszystkich możliwych więziach prakseologicznych w celu otrzymania systemowego modelu rozwoju rynku energii elektrycznej (MRE),

(3) określenie głębokości podsystemowania wzdłużno-poprzecznego (bądź wysokości nadsystemowania) w ujęciu J. Koniecznego oraz przeprowadzenie odpowiedniej operacji na systemowym modelu rozwoju rynku energii elektrycznej w celu wygenerowania modelu cybernetycznego (MC),

(4) przekształcenie modelu cybernetycznego w schemat blokowy zmiennych stanu przy odpowiednich założeniach systemowych (wybór strumieni organizacyjnych jako zmiennych stanu, określenie związków z otoczeniem itp.),

(5) zapis schematu blokowego zmiennych stanu w przestrzeni stanów przy pomocy równań stanu i wyjścia, a więc stworzenie modelu matematycznego rozwoju systemu rynku energii elektrycznej (MM),

(6) wyznaczenie macierzy uogólnionej systemu rynku energii elektrycznej, a więc i macierzy organizacji procesu, zawierającej w swych elementach konkretnie charakterystyki rozbieżności systemowych (ΔS_i),

(7) obliczenie wyznacznika det(s1-A) dla konkretnego wypełnienia charakterystyk rozbieżności systemowych, a więc ustalenie rodziny równań charakterystycznych rozwoju systemu rynku energii elektrycznej, a więc rodziny stopni i współczynników wielomianu charakterystycznego (M(1)), (C(1)),

(8) zbadanie stabilności rozwoju SRE przed zmianami parametrycznymi,

(9) inicjacja genotypu SRE jako zbioru kodów informacyjnych stanowiących populację rozwiązań początkowych (faza „s"),

(10) obliczenie wartości funkcji przystosowania dla każdego osobnika w populacji, a więc ocena przystosowania chromosomów w populacji (faza „i") i podjęcie decyzji czy wprowadzać zmiany parametryczne czy też zmiany strukturalne (P?S),

(11) wprowadzenie zmian parametrycznych, a więc np. wprowadzenie wyrazu wolnego w postaci np. $k=k_1k_2k_3$ (P),

(12) zbadanie stabilności rozwoju rynku energii elektrycznej po zmianach parametrycznych, a przed zmianami strukturalnymi,

(13) wprowadzenie zmian strukturalnych, a więc wprowadzenie nowego pierwiastka (ewentualnie współczynnika) do kodu informacyjnego (S_{n+1}, a_{n+1}),

(14) zbadanie stabilności rozwoju systemu energii elektrycznej po zmianach strukturalnych,

(15) obliczenie funkcji przystosowania dla każdego osobnika po zmianach strukturalnych, a więc ocena strukturalna przystosowania chromosomów w populacji („P") i podjęcie decyzji czy podejmować dalsze zmiany strukturalne (P?S),

(16) sprawdzenie warunku zatrzymania sztucznej ewolucji (np. osiągnięcie pożądanego kodu informacyjnego) i, gdy warunek jest spełniony, wyprowadzenie najlepszego sztucznego kodu genetycznego,

(17) gdy warunek zatrzymania sztucznej ewolucji nie jest spełniony to selekcja kodów informacyjnych do utworzenia par rodziców (faza „d"),

(18) zastosowanie operatorów genetycznych (krzyżowania, mutacji, inwersji itp.),

(19) wyznaczenie nowego kodu informacyjnego oraz nowej macierzy organizacyjnej A uwzględniającej zmiany strukturalne rozwoju rynku energii elektrycznej (faza „u'"),

(20) wyznaczenie wektora zapotrzebowania na rozwój systemu ESP oraz ustalenie metody likwidacji pojawiających się rozbieżności systemowych – tych dla których wyznaczono nowy kod informacyjny (faza „l"),

(21) badanie wrażliwości rozwoju dla systemu zbudowanego na nowym sztucznym kodzie genetycznym poprzez wyznaczenie odpowiedzi systemu $y(\theta)$ na nietypowe (w tym gwałtowne) zmiany parametrów i struktury (faza „w") ,

(22) opracowanie sposobu wdrażania zmian parametrycznych oraz zmian strukturalnych (faza „z"),

(23) realizacja zmian parametrycznych oraz zmian strukturalnych (faza „r"),

(24) podjęcie decyzji o przyjęciu modelu nowego rynku energii elektrycznej i czynności końcowe (faza „k"), czy też o powtórzenie sztucznej ewolucji i wprowadzenie nowych rozbieżności systemowych ΔS_i.

4. Uwagi implementacyjne

Systemowy algorytm ewolucyjny był implementowany dla różnych eksperymentów badawczych, dla systemów techniczno-ekonomicznych oraz technicznych. Wykorzystywano m.in. środowisko takich języków programowania, jak np. język Delphi, j. C++, a także takie środowiska programowe, jak np. środowisko MATLAB z odpowiednimi toolbox'ami takimi jak Simulink, System Identification Toolbox, Control System Toolbox, Neural Network Toolbox, Symbolic Math Toolbox czy też środowisko SPHINX.

Jednym z poważniejszych problemów napotykanych przy implementacji był problem właściwego ustalenia wszystkich parametrów procesu ewolucyjnego możliwych do określenia przez użytkownika, takich jak np.: wielkość populacji, liczba powtórzeń, współczynnik (prawdopodobieństwo) mutacji, współczynnik (prawdopodobieństwo) krzyżowania, współczynnik wymiany populacji itp., a także głębokość podsystemowania. Ustalenie prawidłowych wartości powinno być poprzedzone głęboką analizą problemu rozwoju odpowiedniego systemu.

5. Wizualizacja wiedzy o rozwoju systemu

Wizualizacja rozwoju może być przeprowadzona za pomocą np. linii pierwiastkowych Evansa, co szczegółowo opisano m.in. w pracach [2], [3]. Współcześnie obiecującym kierunkiem badawczym jest także kierunek wizualizacji wiedzy wiązany z sieciami neuronowymi samoorganizującymi się (SOM) [9], przy czym wizualizacja chromosomów może być przeprowadzona za pomocą linii, mapy czy też wykresów przestrzennych, a nawet za pomocą rzutowania wiedzy z wyższych wymiarów na odpowiednie wymiary wizualnie, możliwe do oceny przez decydentów systemowych [3], [4], [8].

A zatem w przypadku sieci SOM można np. przyjąć, że poszczególne sztuczne geny są w pewnym sensie składowymi wektorów wejściowych do sieci SOM, a na wyjściu można odpowiednio przyjąć np. dwuwymiarową siatkę heksagonalną. Badania w tym kierunku są kontynuowane.

Literatura cytowana

1. Arabas J.: Wykłady z algorytmów ewolucyjnych, WNT, Warszawa 2001.
2. Osowski St.: Sieci neuronowe do przetwarzania informacji, WNT, Warszawa 2000.
3. Tchórzewski J.: Systemowe wspomaganie procesu badania prawidłowości rozwoju systemu sterowania na przykładzie elektroenergetycznej sieci przesyłowej, AP, Siedlce 1999.
4. Zieliński J. [red. nauk]: Inteligentne systemy w zarządzaniu. Teoria i praktyka, PWN, Warszawa 2000.
5. Tadeusiewicz R.: Sieci neuronowe, AOW, Warszawa 1993.
6. Goldberg D. E.: Algorytmy genetyczne i ich zastosowania, WNT, Warszawa1995.
7. Michalewicz Z.: Algorytmy genetyczne + struktury danych = programy ewolucyjne, WNT, Warszawa1996.
8. Huk. M., Kwaśnicka H.: Nowe podejście do wykorzystania programowania genetycznego w ekstrakcji wiedzy z sieci neuronowych. Materiały na Konferencję zorganizowaną przez Katedrę Systemów Sztucznej Inteligencji Akademii Ekonomicznej we Wrocławiu. s. 88-95. Wyd. AE. Wrocław 2001.
9. Kłopotek M.: Inteligentne wyszukiwarki internetowe, EXIT, Warszawa 2001.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 65-71.

WIRTUALNY UKŁAD ODDECHOWY I JEGO WYKORZYSTANIE W TESTOWANIU METOD SZTUCZNEJ WENTYLACJI I WSPOMAGANIA ODDYCHANIA

VIRTUAL RESPIRATORY SYSTEM AND ITS USE IN TESTING OF ARTIFICIAL VENTILATION AND VENTILATORY SUPPORT METHODS

TOMASZ GÓLCZEWSKI, MAREK DAROWSKI

Instytut Biocybernetyki i Inżynierii Biomedycznej PAN, Warszawa, tgol@ibib.waw.pl

Streszczenie. Problemy techniczne, ekonomiczne oraz wzrastające znaczenie kwestii etycznych powodują, że wykorzystanie w badaniach naukowych wirtualnych organów może okazać się bardziej wygodne niż eksperymenty na zwierzętach czy ograniczone możliwości wykorzystania pacjentów w badaniach. W szczególności wirtualny układ oddechowy (WUO) może być użyteczny w testowaniu respiratorów i nowych metod wspomagania, w edukacji i treningu personelu, a także we wstępnym testowaniu hipotez naukowych. W pracy przedstawiono: (1) ideę WUO, (2) weryfikację za pomocą standardowej spirometrii (wyników symulacji nie można odróżnić od wyników otrzymanych dla realnych pacjentów), (3) przykłady wykorzystania WUO w: (a) porównaniu zagrożeń dla zdrowia podczas różnych trybów sztucznej wentylacji, (b) analizie konieczności stosowania niezależnej wentylacji płuc, (c) analizie skuteczności wspomagania oddychania metodą CPAP w obturacyjnej chorobie płuc.

Słowa kluczowe: CPAP, niezależna wentylacja, układ oddechowy, rzeczywistość wirtualna, sztuczna wentylacja

Abstract. Economic, technical and ethical problems may render the use of virtual organs in investigations more convenient than experiments on animals or limited investigations on patients. In particular, Virtual Respiratory System (VRS) may be useful for tasks such as respirator and support methods testing, education, staff (medical & technical) training, (initial) testing of scientific hypotheses, etc. The paper presents: (1) the VRS idea, (2) verification with standard spirometric examination (it is impossible to distinguish between results for real patients and VRS, which confirms VRS reliability), (3) examples of the use of VRS in: (a) comparison of health hazard in ventilation of different modes, (b) analysis of necessity of differential ventilation, (c) analysis of efficiency of support with CPAP in obstructive lung disease.

Key words: artificial ventilation, CPAP, differential ventilation, respiratory system, virtual reality

1. Wprowadzenie

Znaczenie badań z użyciem modeli zamiast zwierząt lub pacjentów zaczyna być doceniane nie tylko przez cybernetyków, ale też przez lekarzy, czego dowodem może być cytat z [1]: „…we may expect computer simulations to play an increasingly important role in future biomedical research". Rzeczywiście, badając modele:

- Możemy przebadać dowolną liczbę „pacjentów".
- Unikamy kwestii etycznych, tj. systematycznie rosnących zaostrzeń dotyczących doświadczeń na zwierzętach i eksperymentowania z pacjentami.
- Mamy dostęp do wielkości, których w organizmie albo nie da się w ogóle mierzyć, albo jest to bardzo trudne.
- Możemy dowolnie manipulować wybranymi parametrami organizmu (w rzeczywistym organizmie często w ogóle nie da się manipulować poszczególnym, pojedynczym parametrem, gdyż wszystkie one są wzajemnie powiązane sprzężeniami zwrotnymi: gdy zmieni się jeden, inne też ulegną zmianie). Możliwość takiego wybiórczego manipulowania pozwala na badanie wpływu jednego czynnika niezależnie od wpływu innych, tj. na analizę „krok po kroku".

W przypadku testowania, optymalizowania i porównywania metod sztucznej wentylacji oraz wspomagania oddychania natrafiamy szczególnie na następujące, poza etycznymi, problemy:

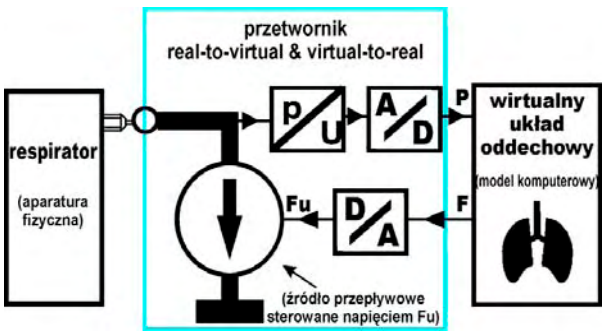
- pacjenci: ograniczone możliwości manipulowania parametrami wentylacji i praktycznie brak możliwości manipulowania parametrami układu oddechowego,
- zwierzęta: wysoki koszt badań, ograniczona modyfikowalność i obserwowalność modeli żywych,

- modele fizyczne: relatywnie zbyt mała złożoność, aby obserwować różnice w efektach różnych metod wspomagania, słaba modyfikowalność, wysokie koszty.

By uniknąć tych problemów, w IBIB PAN zbudowano na tyle złożony model układu oddechowego (UO), że – jak się wydaje – może on być traktowany jako wirtualny UO, który byłby w stanie zastąpić pacjentów, zwierzęta czy też modele fizyczne, przynajmniej we wstępnych testach. Rys. 1 pokazuje ideę łączenia fizycznego sprzętu z wirtualnym UO.

Wg naszej koncepcji:

- model komputerowy jako wirtualny UO zachowuje się tak, jak układ rzeczywisty odnośnie relacji ciśnienie – objętość – przepływ gazów,
- wartości parametrów modelu dla standardowego pacjenta otrzymano, analizując dane z literatury; ich modyfikacje „tworzą" analizowane patologie,
- dla tak zmodyfikowanych parametrów badana jest reakcja UO na testowany rodzaj aktywności respiratora (np. testowaną metodę wspomagania),
- dzięki temu, że wirtualny UO jest w pełni obserwowalny i łatwy do modyfikacji, można obserwować reakcje UO na zmiany wartości wybranych, pojedynczych parametrów (oddechowych i/lub układu), co jest nieosiągalne w przypadku modeli żywych.

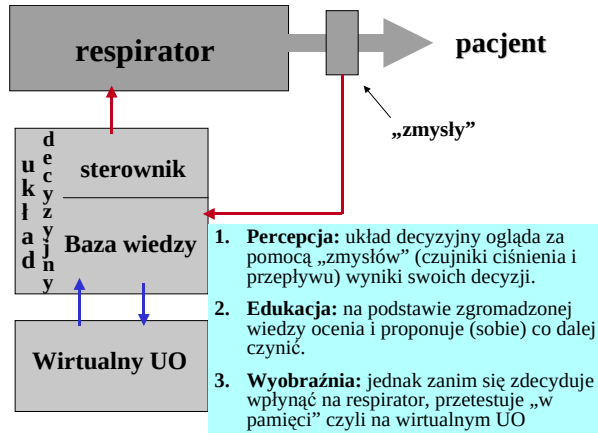


Rys. 1. Idea zastąpienia rzeczywistych układów oddechowych układem wirtualnym podczas testowania respiratorów [2]. Bez względu na typ respiratora i na rodzaj metody wspomagania faktycznym rezultatem działania respiratora jest zmiana/utrzymanie ciśnienia w rurce intubacyjnej. Ciśnienie to stanowi wejście do świata wirtualnego. Komputer oblicza, ile powietrza przeszłoby przez oskrzela w danej chwili, przy takim ciśnieniu. Obliczona wartość, po zamianie na napięcie Fu, steruje źródłem przepływowym, które realizuje ten obliczony przepływ powietrza. W efekcie, z punktu widzenia respiratora jest tak, jakby był on przyłączony do rzeczywistego układu oddechowego

Prócz pierwotnego zastosowania wirtualnego UO przewiduje się dodatkowe:

- w edukacji oraz treningu personelu (np. kursy spirometrii – patrz niżej: 3. Weryfikacja modelu),
- jako pomoc w weryfikacji hipotez naukowych (np. czynniki determinujące częstość oddychania – patrz niżej w: 4.3. Analiza skuteczności metody CPAP),
- a także – w bliżej nieokreślonej przyszłości – jako składnik inteligentnego respiratora (rys. 2).

W niniejszej pracy przedstawiono w skrócie czteroletnie doświadczenia w budowaniu wirtualnego UO i przykłady jego wykorzystania w testowaniu metod sztucznej wentylacji (dwa przykłady) oraz wspomagania oddychania spontanicznego (jeden). Część przedstawionych tu wyników została już opublikowana lub zaprezentowana na konferencjach krajowych i międzynarodowych.



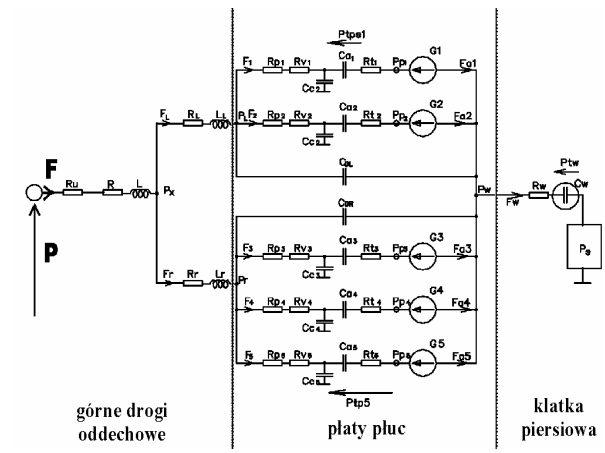
Rys. 2. Wirtualna rzeczywistość jako niezbędny składnik „prawdziwej" sztucznej inteligencji na przykładzie koncepcji inteligentnego respiratora

2. Opis modelu

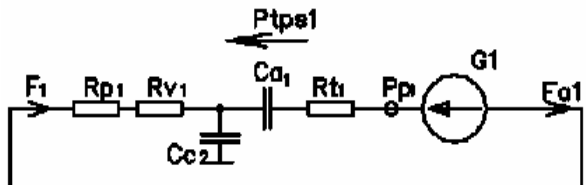
Ogólną strukturę modelu pokazuje rys. 3, natomiast szczegóły dotyczące płata płucnego przedstawia rys. 4. Głównymi cechami modelu są:

- rozdzielenie płuc i klatki piersiowej umożliwiające obserwację ciśnienia śródopłucnowego (które wpływa na opór dróg oddechowych oraz układ krążenia),
- podział płuc na pięć płatów (jak na razie),
- podział oporów dróg oddechowych (Raw) na:
  a) opory zależne od objętości płuc (będące częścią miąższu płuc) – Rv na rys. 4,
  b) opory zależne od ciśnienia transmuralnego, tj. opory oskrzeli średniego rzędu, które ulegają zapadaniu – R_p na rys.4;
  c) opory dużych oskrzeli, tchawicy, jamy ustnej.
- oraz wiele innych, jak: lepkość płuc i klatki piersiowej oraz nieliniowość ich podatności, wpływ grawitacji, sprężystość powietrza, przestrzeń martwa, itd.

Więcej szczegółów dotyczących modelu zawierają [2], [4].



Rys. 3. Generalna struktura modelu, tj. jego uproszczony „schemat" (np. bez pokazywania nieliniowości elementów)



Rys. 4. Fragment rys.3 (górny płat lewego płuca). F – przepływ powietrza, Fa – zmiana objętości płata, Ptps – statyczne ciśnienie transpulmonarne, G – wpływ grawitacji, Pp – ciśnienie przypłucne, Rt – lepkość tkanki płucnej, Ca – podatność płata, Cc – ściśliwość powietrza, Rv – opory dróg oddechowych zależne od objętości płata, Rp – opory dróg oddechowych zależne od ciśnienia transmuralnego

3. Weryfikacja modelu

Niektórzy różnicę między modelem (komputerowym) obiektu a obiektem wirtualnym definiują następująco:

- model odwzorowuje pewne właściwości, których analiza była powodem tworzenia modelu,
- wirtualny obiekt odwzorowuje sam obiekt na tyle wiernie, że w przypadku danej klasy działań, dla analizy których wirtualny obiekt powstał, nie powinno być możliwym odróżnienie, czy reakcje na tego typu działania pochodziły od obiektu realnego czy wirtualnego.

W naszym przypadku chodzi o reakcje UO na ciśnienia zadawane przez respirator (lub generowane przez mięśnie oddechowe, choć sama działalność mięśni nie jest na razie zadowalająco symulowana). Uznano, że przynajmniej w części model będzie pozytywnie zweryfikowany (jako wirtualny UO), jeśli po poddaniu go standardowym badaniom medycznym nie będzie można odróżnić, czy wyniki pochodzą od realnego pacjenta, czy też są rezultatem symulacji. Takim podstawowym badaniem UO jest spirometria, a szczególnie parametry i wykresy opisujące wytężony wydech.

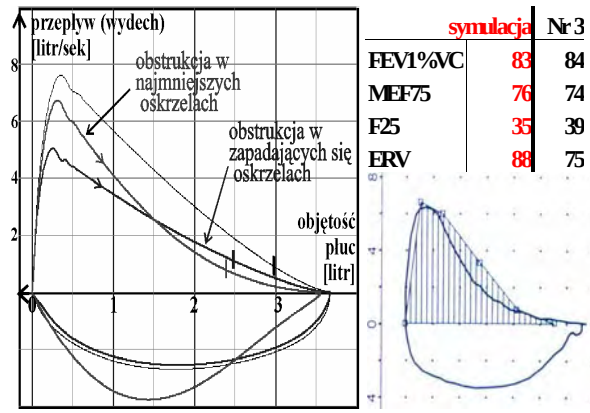
Choć spirometria jest podstawową i najłatwiejszą metodą diagnostyki UO, to bazuje ona na jednych z najbardziej złożonych własności UO, powodujących m.in. tzw. (auto)limitowanie przepływu w trakcie wytężonego wydechu. Generalnie, limitowanie polega na tym, że jeśli wydech jest wytężony, tj. odpowiednio gwałtowny, to natężenie przepływu nie zależy od sposobu wydechu. Natężenie jest funkcją jedynie objętości płuc, która to funkcja zależy tylko od własności mechanicznych UO. Ta specyfika UO jest niezwykle istotna w diagnostyce UO, bowiem eliminowany jest wpływ sposobu oddychania, będący w większy lub mniejszym stopniu indywidualną cechą pacjenta. Jest to też istotne dla weryfikacji modelu, bowiem:

- można sprawdzić, czy tak ważne zjawisko, jak limitowanie przepływu, w ogóle zachodzi,
- podobieństwo między symulowanym i rzeczywistym przebiegiem przepływu gazu z płuc nie byłoby żadnym dowodem wierności modelu, gdyby w przypadku każdego modelu – w tym również złego - można było dla dowolnego przebiegu przepływu dobrać odpowiedni kurs wydechu realizujący ten przebieg; co innego, gdy przebieg przepływu nie zależy od kursu wydechu.

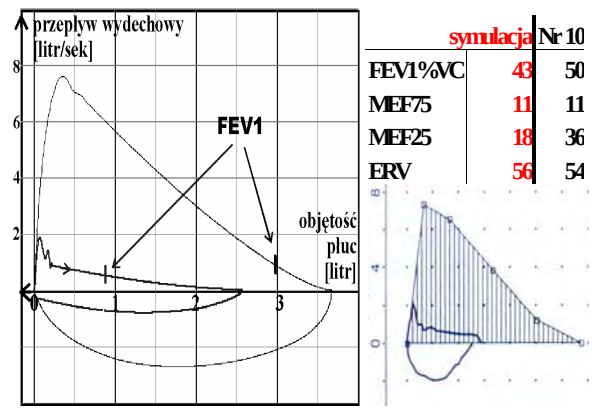
Rys.5 i 6 pokazuje przykładowe wyniki spirometrii wirtualnego UO oraz realnych pacjentów. Oczywiście, w przypadku wirtualnego UO nie trzeba było używać realnego spirometru, bowiem to, co mierzy spirometr, jest bezpośrednio dostępne obserwacji. Wyniki realnych pacjentów zaczerpnięto za zgodą autorów z [5].

Zarówno w przypadku zaprezentowanych tu przykładów, jak i pozostałych przeprowadzonych symulacji, nie można z samych wartości parametrów i kształtów krzywych odróżnić, które wyniki dotyczą realnych pacjentów, a które są rezultatem symulacji. I tak, na przykład na pokazanych tu rysunkach widać wklęsłość charakterystyczną dla obstrukcji w najmniejszych oskrzelikach (rys. 5), czy też pik przepływu na samym początku wydechu w poważnej chorobie obturacyjnej płuc (rys. 6) – pełna obserwowalność wirtualnego UO pozwoliła stwierdzić, że pik ten odpowiada opróżnianiu się przestrzeni martwej, tj. opróżnianiu się oskrzeli ściskanych w trakcie wydechu.

Tak więc wydaje się, że model spełnia, przynajmniej w części, kryteria bycia wirtualnym UO.



Rys. 5. Wyniki spirometrii wirtualnego UO dla dwóch łagodnych obstrukcji o różnej lokalizacji (po lewej) i przykładowy wynik dla rzeczywistego pacjenta (nr 3 z [5]) z obturacyjną chorobą płuc w najmniejszych oskrzelikach. Jak zwykle w tych badaniach, oś objętości dotyczy objętości powietrza wydalonego z płuc, a więc maksymalnej objętości płuc odpowiada zero na osi. W tabelce, wyrażone procentem normy, wartości najważniejszych parametrów: FEV1%VC – objętość wysilonego wydechu w pierwszej sekundzie w stosunku do pojemności życiowej VC, MEF75/25 – przepływ przy 75/25% VC, ERV – rezerwowa objętość wydechowa



Rys. 6. Jak na rys.5, dla obturacyjnej choroby płuc znacznego stopnia. Cienka linia na rys. po lewej (symulacje) i zakreskowane pole po prawej (realny pacjent) przedstawiają normę

4. Przykłady wykorzystania modelu

4.1. Porównanie trzech trybów sztucznej wentylacji w aspekcie potencjalnych zagrożeń dla zdrowia [3]

Generalnie, można wyróżnić dwie grupy metod sztucznej wentylacji:

VCV [*volume-controlled ventilation*] – takie tryby wentylacji, w których przebieg przepływu gazu (więc i objętość oddechową) zadaje lekarz, natomiast wartość ciśnienia zależy od własności UO.

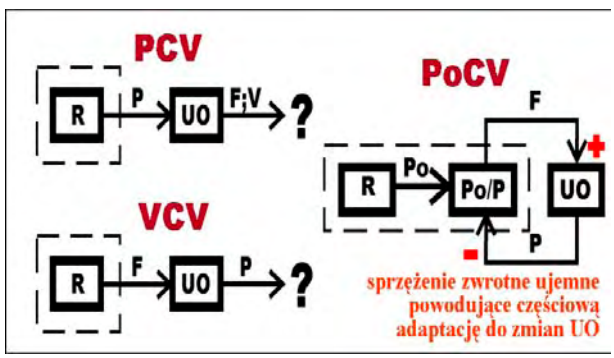
VCV zapewnia wymaganą przez lekarza wentylację pacjenta, jednak istnieje groźba:

- nadmiernego wzrostu ciśnienia [barotraumy] przy małej podatności UO,
- nadmiernej wentylacji zdrowej części płuc [wolutraumy] przy asymetrycznych zmianach oporów dróg oddechowych (co zostanie pokazane), itd.

PCV [*pressure-controlled ventilation*] – takie tryby wentylacji, w których wartość ciśnienia zadaje lekarz, natomiast przepływ gazu (więc i objętość oddechowa) zależy od własności UO. Co prawda, jeśli lekarza nie popełnił pomyłki, wartości ciśnienia nie zagrażają zdrowiu podczas wentylacji trybem PCV, jednak istnieje groźba:

- nadmiernej wentylacji płuc [wolutraumy], gdy podatność UO jest duża,
- hipowentylacji pacjenta, jeśli opór dróg oddechowych jest duży lub podatność płuc mała, itd.

By zmniejszyć ww. niebezpieczeństwa, Marek Darowski (IBIB PAN) zaproponował (1990 r.) tryb wentylacji, w którym wielkością zadaną jest moc, tj. iloczyn ciśnienia i przepływu, a nie ciśnienie czy przepływ z osobna. Tryb ten nazwano PoCV [*Power-Controlled Ventilation*] lub adaptacyjnym, bowiem parametry wentylacji po części adaptują się do stanu UO (rys. 7).



Rys. 7. Różnica między tradycyjnymi trybami wentylacji a PoCV. W przypadku PCV respirator generuje zadany przebieg ciśnienia P – brak wpływu na wielkość przepływu F, a więc i objętości płuc V (ryzyko wolutraumy lub hipowentylacji). Analogicznie dla VCV (brak wpływu na ciśnienie – ryzyko barotraumy). W przypadku PoCV zadawana jest moc, tj. iloczyn przepływu i ciśnienia. Od własności UO zależy "rozdział mocy" na ciśnienie i przepływ

Choć tryb PoCV wg założeń powinien być bezpieczniejszy, należałoby to sprawdzić w badaniach eksperymentalnych. Użycie w tym celu wirtualnego UO wydaje się być jak najbardziej na miejscu. Porównano PoCV, o stałej, w czasie wdechu – mocy – z PCV i VCV o stałym w czasie wdechu – odpowiednio ciśnieniu i przepływie. Przeprowadzono dwa eksperymenty:

(a) asymetryczna obstrukcja w jednym płucu istnieje przed określeniem nastawów respiratora,
(b) asymetryczna obstrukcja pojawia się w trakcie wentylacji (np. zatkanie oskrzela plwociną, co dość często zdarza się w praktyce).

W obu przypadkach założono, iż pacjent wymaga wentylacji minutowej równej 10 litrów/min przy częstości wentylowania równej 15/min.

W pierwszym eksperymencie badano wartość maksymalną i średnią ciśnienia w zdrowym płucu (ryzyko barotraumy, negatywny wpływ na układ krążenia), maksymalny przepływ (np. ryzyko cofania plwociny spowrotem do płuc) oraz zmiany w dystrybucji wentylacji.

W drugim eksperymencie badano zmiany wentylacji minutowej oraz objętości oddechowej po pojawieniu się obstrukcji.

Wyniki i wnioski

Wyniki podane w tab. 1 i tab. 2 potwierdzają, że biorąc pod uwagę niektóre parametry, tryb VCV jest lepszy od PCV, uwzględniając inne – PCV jest lepszy od VCV. Tryb PoCV dla wszystkich parametrów plasuje się w środku.

Tab. 1. Wartości ciśnień i przepływów dla badanych trybów wentylacji, dla zdrowych płuc i asymetrycznej obstrukcji w prawym płucu (2 i 4-krotny wzrost oporu dróg oddechowych)

prawe płuco Raw=	Ppeak tchawica [H_2O]	Ppeak lewe płuco [H_2O]	Pmean lewe płuco [H_2O]	Fpeak tchawica [l/s]
norma	6.64 C 7.15 A 7.35 F	6.67 C 6.92 A 7.00 F	3.61 F 4.08 A 5.81 C	0.50 F 0.74 A 2.02 C
2x norma	6.84 C 7.61 A 7.92 F	6.94 C 7.42 A 7.57 F	3.93 F 4.41 A 5.96 C	0.48 F 0.72 A 1.66 C
4x norma	7.55 C 8.55 A 9.04 F	7.65 C 8.33 A 8.63 F	4.40 F 4.93 A 6.51 C	0.49 F 0.72 A 1.59 C

Ppeak/Fpeak – największa wartość ciśnienia/przepływu w czasie wdechu, Pmean – średnie ciśnienie. Wentylacja z częstością 15/min, wentylacja minutowa 10 litrów/min. W tabeli poszczególne tryby wentylacji (F=VCV, A=PoCV, C=PVC) umieszczono w kolejności od najlepszych (najmniejsze ciśnienia i przepływ) do najgorszych.

Tab. 2. Stosunek wentylacji danego płuca przy obstrukcji w prawym płucu do wentylacji zdrowych płuc

	Raw=2x norm			Raw=4x norm		
	F	A	C	F	A	C
prawe	0.83	0.86	0.91	0.59	0.62	0.69
lewe	1.22	1.20	1.18	1.54	1.50	1.47

Wentylacja z częstością 15/min, wentylacja minutowa 10 litrów/min.

Tab. 3. Zmiany wentylacji minutowej i objętości oddechowych po nagłym pojawieniu się obstrukcji w prawym płucu

tryb	Vm	V_T prawe	V_T lewe
PCV	88	59	131
PoCV	89	54	136
VCV	97	57	150

Vm – wentylacja minutowa, V_T – objętość oddechowa. Zmiany wyrażone w % w stosunku do wentylacji przed pojawieniem się obstrukcji, kiedy to Vm=10 litrów/min.

W przypadku nagle pojawiającej się obstrukcji (tab. 3), tryb PoCV jest gorszy od trybów PCV i VCV, jedynie jeśli chodzi o hipowentylację płuca, w którym doszło do częściowego zatkania oskrzeli. Jednak PoCV należy sumarycznie uznać za lepszy od VCV w sytuacji nagle pojawiającej się przeszkody w oskrzelach, z powodu znacznie większego ryzyka wolutraumy w zdrowym płucu w przypadku VCV.

Tak więc z punktu widzenia żadnego parametru tryb PoCV nie jest najlepszy, ale też nie jest najgorszy. W przypadku jakości łańcucha decydujące znaczenie ma najsłabsze ogniwo, a nie najmocniejsze. Podobnie w terapii: nie powinien decydować największy zysk, a największa szkoda (*primum non nocere*), inaczej pacjenci będą kpić: operacja się udała, ale pacjent umarł.

Reasumując: Podczas wentylacji z zadaną mocą występuje częściowa auto-adaptacja wentylacji do zmian stanu układu oddechowego. Dzięki temu unika się skrajnych wartości istotnych dla zdrowia parametrów. Z powyższego powodu wentylację w tym trybie powinno uznać się za stwarzającą najmniejsze zagrożenie dla zdrowia pacjenta.

4.2. Niektóre aspekty niezależnej wentylacji płuc

Opisane wyżej niedoskonałości, np. ryzyko hiperwentylacji jednego z płuc, nie wystąpią, jeśli zastosowana zostanie niezależna wentylacja płuc (NWP), czyli wentylowanie każdego z płuc z osobna, choć w sposób synchroniczny (zazwyczaj).

Konieczność stosowania NWP wydaje się oczywista np. w przypadku uszkodzenia jednego z płuc lub jedno-stronnego urazu klatki piersiowej, czyli generalnie: w przypadkach asymetrycznych patologii UO.

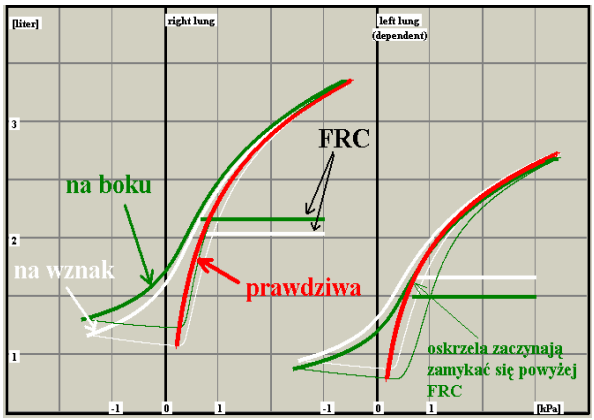
Nas zainteresowała kwestia wentylacji zdrowych płuc, ale pracujących w asymetrycznych warunkach, np. przy bocznym położeniu pacjenta.

Wiadomo, co również dotyczy naszego wirtualnego UO (rys. 8), że:

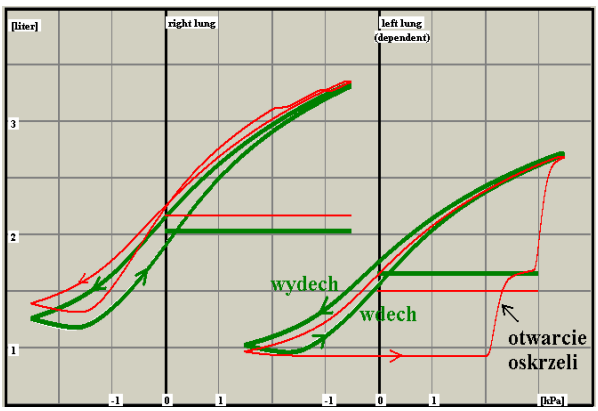
- oskrzela zamykają się przy małej objętości płuca, co uniemożliwia dalsze opróżnianie się płuca mimo coraz bardziej ujemnego ciśnienia; z punktu widzenia respiratora wygląda to na znaczny spadek podatności płuc (aż do zera),
- wpływ grawitacji powoduje, że w pozycji bocznej objętość górnego płuca w stanie spoczynku (FRC) jest większa niż w pozycji na wznak, natomiast FRC dolnego płuca jest mniejsze,
- w miarę starzenia się organizmu objętość płuca, przy której oskrzela ulegają zamknięciu, zwiększa się.

Celem symulacji było sprawdzenie, jak powyższe fakty wpływają na wentylację każdego z płuc przy zmianie pozycji z na wznak na pozycję boczną. Rys. 9 pokazuje, że mogą zajść – u osób zdrowych, lecz już w starszym wieku – dramatyczne zmiany wentylacji. Wentylacja może być całkiem „normalna" w pozycji na wznak. Po obróceniu pacjenta na bok, FRC dolnego płuca zmniejsza się i może być mniejsza od objętości, przy której oskrzela zamykają się. Wtedy ciśnienie generowane przez respirator przez dłuższy czas nie jest w stanie otworzyć tych oskrzeli i płuco nie jest wentylowane (rys. 9). Natomiast w chwili otwarcia następuje dość gwałtowne napełnianie się płuca, co może cofać plwocinę.

Jeśli więc nawet lekarz prawidłowo nastawił respirator, gdy pacjent leżał na wznak, ale potem przewrócono go na bok, to utlenowanie krwi może być zdecydowanie niewłaściwe: mimo że sumarycznie płuca są wentylowane odpowiednią Vm, to dolne płuco, przez które przepływa więcej krwi, jest krócej i gorzej wentylowane. Jedynym wyjściem jest niezależna wentylacja obu płuc: normalne wentylowanie górnego płuca oraz takie wentylowanie płuca dolnego, które utrzymywałoby dodatnie ciśnienie w czasie wydechu w celu niedopuszczenia do zamykania się oskrzeli.



Rys. 8. Wyniki symulacji dla prawego płuca (po lewej) i lewego (po prawej): prawdziwe zależności objętości od ciśnienia transpulmonarnego, tj. prawdziwe podatności płuc (krzywe czerwone) oraz podatności „widziane" przez respirator u pacjenta leżącego na wznak (białe) i na lewym boku (zielone). Oś pozioma – ciśnienie, oś pionowa – objętość płuca. Poziome kreski ilustrują zmiany FRC przy zmianie pozycji



Rys. 9. Pętle objętość-ciśnienie w tchawicy (wdech-wydech) dla prawego płuca (po lewej) i lewego w pozycji pacjenta na wznak (zielone) i na lewym boku (czerwone). Oś pozioma – ciśnienie. Pacjent zdrowy, ale w starszym wieku

4.3. Analiza warunków skuteczności wspomagania oddychania metodą CPAP w chorobie obturacyjnej

Jak wyżej pokazano, jeśli FRC jest mniejsze od objętości płuc, przy której płuca ulegają zamknięciu, potrzebne jest utrzymywanie dodatniego ciśnienia w czasie wydechu. Jeśli pacjent oddycha samodzielnie, wymaga on w takiej sytuacji wspomagania za pomocą CPAP [*Continuous Positive Airway Pressure*], tj. za pomocą utrzymywania cały czas stałego, dodatniego ciśnienia w drogach oddechowych (np. za pomocą maski). W tym wypadku mechanizm pozytywnego wpływu jest

oczywisty. Natomiast nie do końca był jasny mechanizm, a przede wszystkim warunki skuteczności metody CPAP w przypadku choroby obturacyjnej. Rys. 10 pokazuje plusy i minusy CPAP. Zanalizowano, kiedy dodatni wpływ przeważa nad ujemnym.

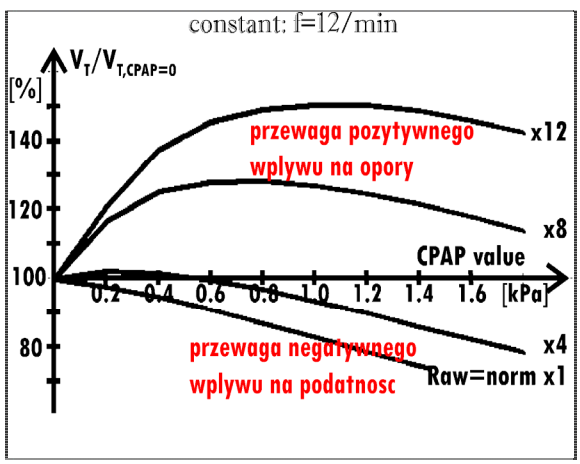
Rys. 10. Schematycznie przedstawiony dodatni i ujemny wpływ wspomagania metodą CPAP na prace UO

Metoda i wyniki

Celem oddychania jest zapewnienie odpowiedniej wentylacji minutowej Vm (odpowiedniego utlenowania krwi) przy możliwie najmniejszych kosztach własnych, tj. możliwie małym wysiłku oddechowym. Organizm sam „optymalizuje" ww. koszty, dobierając sobie odpowiednią częstość i głębokość oddychania. Wiadomo, że w chorobie obturacyjnej częstość oddechu spada, choć wymaga to głębszych, a więc o wiele bardziej energochłonnych wdechów.

Związek między pracą mechaniczną (dW=p•dV) możliwą do obliczenia, a pracą „biologiczną", której miarą jest zużycie tlenu, jest dość luźny. Dlatego nie badano, jak zmienia się trudny do oszacowania wysiłek oddechowy przy stałej, wymaganej przez organizm wentylacji minutowej Vm, a na odwrót – analizowano:

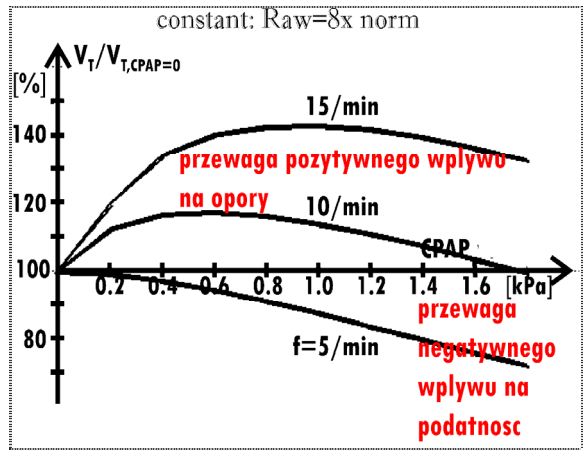
- jak zmienia się Vm i objętość oddechowa V_T=Vm/f,
- przy stałej sile rozwijanej przez mięśnie oddechowe,
- przy selektywnie zmienianych wartościach CPAP, Raw i częstości oddychania f.



Rys. 11. Zależność względnej zmiany objętości oddechowej V_T od wartości CPAP dla stałej częstości oddychania, dla czterech oporności dróg oddechowych: normalnej oraz jej wielokrotności

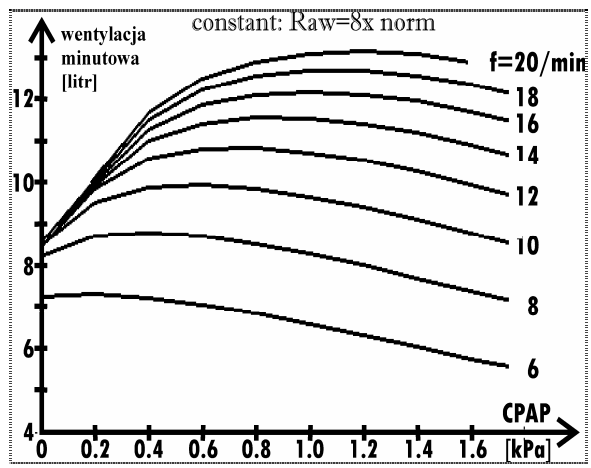
Rys. 11 pokazuje, że w zależności od wielkości Raw CPAP powoduje zwiększenie lub zmniejszenie V_T. W szczególności CPAP utrudnia oddychanie osobie zdrowej. Mamy tu więc do czynienia z rzadko uświadamianym, ale dość częstym zjawiskiem, że to, co pomaga choremu, szkodzi zdrowemu.

W symulacji pokazanej na rys. 11 zmieniano Raw i utrzymywano stałą częstotliwość oddychania. Na rys. 12 pokazano, co dzieje się, gdy zmienia się częstotliwość, a utrzymuje stały, zwiększony Raw. Symulacja pokazała, że gdy częstość jest niska, CPAP jest zbędny – tak jak przy normalnym Raw – gdyż spadek częstości rekompensuje wzrost Raw.



Rys. 12. Zależność względnego przyrostu V_T od wartości CPAP dla różnych częstości oddychania przy stałym zwiększonym oporze Raw=8x norma

Wyniki zaprezentowane na rys. 13 pokazują przede wszystkim to, dlaczego osoby z obturacyjną chorobą płuc oddychają wolniej: im po prostu nie opłaca się oddychać szybciej, gdyż i tak nie poprawi to wentylacji minutowej. Za to wysiłek oddechowy wzrósłby proporcjonalnie do wzrostu częstotliwości oddychania. Dopiero zastosowanie CPAP umożliwia oddychanie z normalną częstością.



Rys. 13. Wpływ CPAP na wentylację minutową dla różnych częstości oddychania przy podwyższonym oporze dróg oddechowych (dla przypomnienia: we wszystkich symulacjach siła rozwijana przez mięśnie oddechowe była taka sama – to postulowana we wprowadzeniu analiza „krok po kroku")

Dyskusja i wnioski

Z jaką częstotliwością powinno się oddychać? Zakładając wstępnie, że podatność płuc C nie zależy od ich objętości, praca W przeciw „oporom sprężystym" przy wentylacji V litrów w jednostce czasu równa jest:

$W_1 = E \cdot V^2/2$

w przypadku jednego głębokiego oddechu lub

$W_2 = 2 \cdot E \cdot (V/2)^2/2 = (1/2) \cdot E \cdot V^2/2 = 0.5 \cdot W_1$

dla dwóch, dwa razy płytszych oddechów.

[W liczono ze wzoru $dW = p \cdot dV = E \cdot V \cdot dV$, gdzie $E = 1/C$]

Tak więc skoro $W_2 = 0.5 \cdot W_1$, tj. dwa płytsze oddechy wymagają dwa razy mniejszego wysiłku niż jeden głębszy, opłaca się oddychać częściej, ale płyciej. Fakt, że w rzeczywistości podatność maleje w miarę wzrostu objętości płuc, czyni oddychanie częstsze ale płytsze jeszcze bardziej ekonomicznym. Głównym ograniczeniem płytkości oddechów jest to, że do płuc trafia najpierw „zużyte" powietrze pozostałe w drogach oddechowych, a dopiero potem powietrze zawierające tlen. Wdech nie może więc być płytszy niż objętość przestrzeni martwej.

Wyniki symulacji pozwalają na sformułowanie tezy, że CPAP zastępuje dobrze znany, fizjologiczny mechanizm obronny polegający na zmniejszeniu częstotliwości oddychania w przypadku choroby obturacyjnej płuc. Jednak CPAP nie ma wady tego mechanizmu, którą jest konieczność zwiększenia wysiłku oddechowego związanego z koniecznością pogłębieniem wdechów w celu utrzymania Vm na wymaganym poziomie przy mniejszej częstości oddychania. Dlatego lepiej w chorobie obturacyjnej stosować CPAP niż zdać się na naturalny mechanizm obronny.

Reasumując: Wydaje się, że łatwiejsze wdychanie powietrza poprzez zmniejszone opory dróg oddechowych ma marginalne znaczenie w obturacyjnej chorobie płuc. Skuteczność CPAP polega raczej na umożliwieniu efektywnego oddychania z częstością zbliżoną do normalnej, co czyni zbędnym naturalny mechanizm obronny, tj. powolne oddychanie z głębokimi, energochłonnymi wdechami.

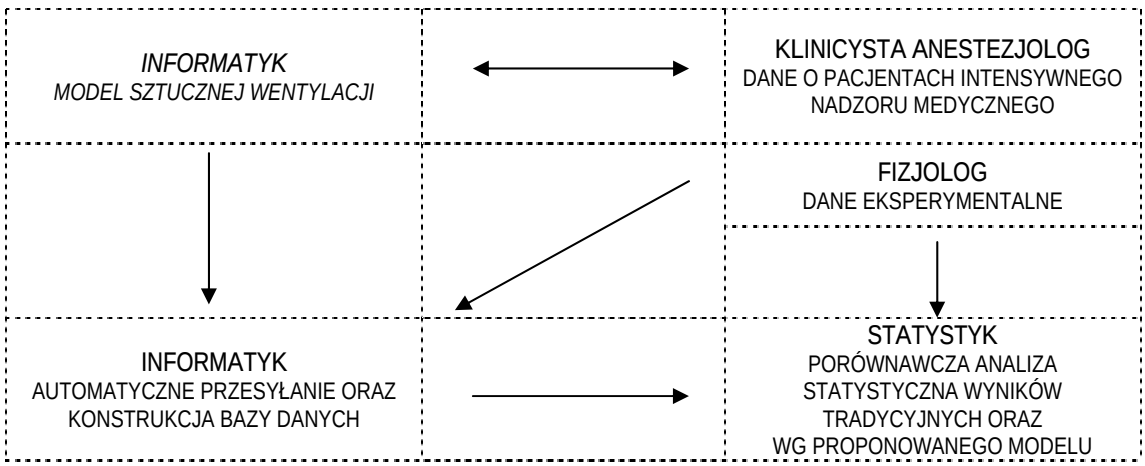
5. Podsumowanie

Jak się wydaje, opracowano wystarczająco wierny model układu oddechowego, by mógł być on używany w charakterze wirtualnego układu oddechowego. Co prawda wymaga on jeszcze wielu modyfikacji niezbędnych do tego, by można było go wykorzystać w problematyce, którą obecnie zajmujemy się w IBIB PAN. Ale już dziś można by go wykorzystać w edukacji czy np. jako pomoc w kursach spirometrii dla lekarzy (po opracowaniu interfejsu użytkownika wygodnego dla lekarzy).

Część prezentowanych prac wykonano w ramach grantu KBN nr 4T11E02924

Literatura cytowana

1. Schuessler T. F., Gottfried S. B., Bates J. H. B.: A model of the spontaneously breathing patient: applications to intrinsic PEEP and work of breathing, J. Appl. Physiol., 82(5): 1694–1703, 1997.
2. Gólczewski T., Kozarski M., Darowski M.: The respirator as a user of virtual lungs, Biocybernetics and Biomedical Engineering, 23(2): 57-66, 2003.
3. Gólczewski T., Darowski M.: Influence of ventilatory mode on respiration parameters – investigation on virtual lungs, Biocybernetics and Biomedical Engineering, 23(3): 63-72, 2003.
4. Kozarski M, Gólczewski T, Darowski M.: Model studies on respiratory parameters for different lung structures, Biocybernetics and Biomedical Engineering, 20:67-77, 2000.
5. Lubiński W., Targowski T.: Ćwiczenia ze spirometrii, Prace Wojskowego Instytutu Medycznego, CSK MON, Warszawa 2003.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 73-80.

NOWE SPOJRZENIE NA GENEZĘ PIERWOTNEGO NADCIŚNIENIA TĘTNICZEGO – HIPOTEZA CYBERNETYCZNA

A NEW APPROACH TO THE GENESIS OF PRIMARY ARTERIAL HYPERTENSION – A BIOCYBERNETIC HYPOTHESIS

TOMASZ GÓLCZEWSKI

Instytut Biocybernetyki i Inżynierii Biomedycznej PAN, Warszawa, tgol@ibib.waw.pl

Streszczenie. Nowa, biocybernetyczna teoria genezy pierwotnego nadciśnienia tętniczego została zaproponowana w 1992 r. Wyniki eksperymentalne ostatnich lat otrzymane przez innych autorów wypełniły ją biochemiczną treścią. Najważniejsze z tych wyników dotyczą wpływu pulsacyjności przepływu krwi na wydzielanie przez śródbłonek takich substancji wazoaktywnych, jak: wazodylatory prostacyklina i tlenek azotu (NO) oraz np. endotelina (ET-1) obkurczająca naczynia. W pracy pokazano, jak pętle ujemnego sprzężenia zwrotnego (SZ), które stabilizują ciśnienie tętnicze, są skomponowane z oporu obwodowego (PR), dynamiki zmian przepływu krwi (DA) oraz sekrecji NO lub ET-1. DA zależy od wielkości odbić fal przepływu determinowanej przez różnicę między PR i impedancją tętniczą. W normalnych warunkach mamy pętle: PR→+DA→+NO→-PR oraz PR→+DA→-ET-1→+PR (gdzie →+/- oznacza sprzężenie dodatnie/ujemne, np. NO→-PR znaczy, że wzrost ilości NO powoduje spadek PR). Jednakże, jeśli ze wzrostem ciśnienia impedancja zacznie rosnąć szybciej niż PR (np. z powodu starzenia), to dodatnie sprzężenie PR→+DA zamieni się w ujemne PR→-DA. W efekcie ujemne SZ zamienią się w dodatnie i tylko nieliniowości układu krążenia (np. ograniczona wydolność serca) limitują zmiany PR i P. Te dodatnie SZ powodują albo nadciśnienie, albo hipotonię, co zależy od znaku wyjściowego uchybu, który ulega wzmocnieniu przez dodatnie SZ. Prezentowana teoria dotyczy tego typu pierwotnego nadciśnienia, które jest związane z uczuciem zmęczenia, a nie pobudzenia. Teoria tłumaczy negatywny lub pozytywny wpływ większości czynników wiązanych z nadciśnieniem, np. dziedziczność, starzenie się, ćwiczenia fizyczne itd.

Słowa kluczowe: dynamiczność przepływu krwi, tlenek azotu, endotelina, nadciśnienie, sprzężenie, Waveform Age

Abstract. A new, biocybernetic theory of primary arterial hypertension was proposed in 1992. Experimental results of last years obtained by other authors have filled the theory with biochemical facts. Most important of these facts concern influence of pulsatile blood flow on endothelial secretion of several vasoactive substances as: vasodilators, e.g. prosta-cyclin, nitric oxide (NO), and vasoconstrictors, e.g. endothelin (ET-1). The paper shows how loops of negative feedback, which stabilize arterial pressure (P), are composed of peripheral resistance (PR), dynamics of blood flow alteration (DA) and NO or ET-1 secretion. DA depends on blood flow wave reflections, which are determined by difference between PR and arterial impedance. Normally, there are the loops: PR→+DA→+NO→-PR and PR→+DA→-ET-1→+PR (where '→+', '→-' means positive or negative connection, respectively, e.g. NO→-PR means that increase of NO amount causes decrease of PR). However, if the impedance starts to increase with P more quickly than PR (e.g. because of ageing), the positive connection PR→+DA changes into PR→-DA. As the consequence, the negative feedbacks change into the positive ones, and only nonlinearity of the cardiovascular system (e.g. limited heart efficiency), bounds alteration of PR and P. These positive feedbacks cause either hypertension or hypotonia, which depends on the sign of the initial deviation that is amplified by the feedbacks. The presented theory concerns rather the primary hypertension type that is connected with fatigue than other types. The theory explains negative or positive influence of almost all factors connected with hypertension, as heredity, ageing, physical exercises, etc.

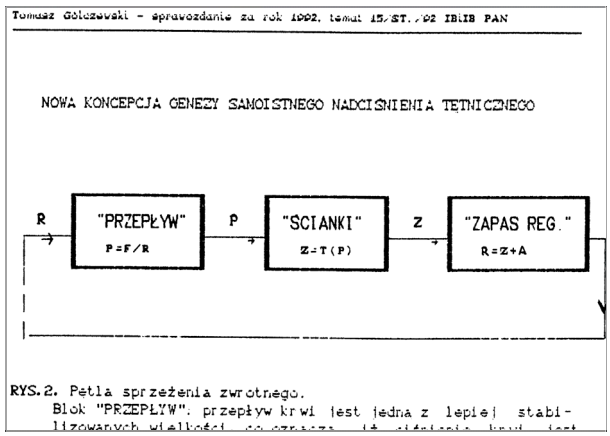
Key words: blood flow dynamicity, hypertension, feedback, nitric oxide, endothelin, Waveform Age

1. Przedmowa – medycyna teoretyczna

Czym dla fizyki jest matematyka, tym dla medycyny jest (bio)cybernetyka, a raczej: tym dla medycyny powinna być cybernetyka. W fizyce dzięki matematyce od kilku wieków można wyniki eksperymentów opisywać, systematyzować itd. To samo próbuje się robić w medycynie za pomocą cybernetyki. Ale fizycy poszli o wiele dalej: dzięki matematyce zaczęli przewidywać – często na dziesiątki lat naprzód – to, co inni później odkrywali w rzeczywistości, czy raczej: eksperymentalnie potwierdzali odkrycia dokonane „na papierze". Mówi się przecież, że np. fale elektromagnetyczne odkrył Maxwell a nie Hertz, który 27 lat później „tylko" potwierdził eksperymentalnie istnienie tych fal. Tak więc, dzięki odwadze fizyków, którzy „uwierzyli" matematyce, mogła powstać fizyka teoretyczna, która fizykę jako taką rozwinęła w stopniu – jak dotąd –

niedoścignionym dla innych nauk, dzięki niej „zaglądamy w czasy", zanim Czas w ogóle powstał, dzięki niej łatwo piszemy i wysyłamy referaty na konferencje.

Niestety, o medycynie teoretycznej wciąż raczej tylko można marzyć. Dlatego w tej pracy zostanie jedynie wspomniane, że w 1992 r. opracowano [7] – posługując się biocybernetyką tak, jak fizycy teoretyczni posługują się matematyką - nową hipotezę dotyczącą genezy pierwotnego nadciśnienia tętniczego (NT). Hipoteza ta jest częścią składową, czy raczej wnioskiem ogólniejszej teorii dotyczącej dynamiczności przepływu krwi (tj. dynamiki zmian przepływu krwi – rys. 2) i jej znaczenia dla organizmu (teoria RVR – *Range of Vascular Regulation*). Pewne podstawy tej teorii zawarte były w pracy doktorskiej autora oraz w [8].



Rys. 1. „Historyczny" rysunek z 1992 r. [7]

Rys. 1 prezentuje główną ideę hipotezy dotyczącej genezy NT. Miał on, obszerny podpis, który skrótowo wyjaśnia(ł) koncepcję, mianowicie:

Pętla sprzężenia zwrotnego.

Blok „PRZEPŁYW": przepływ krwi jest jedną z lepiej stabilizowanych wielkości, co oznacza, iż ciśnienie krwi jest funkcją oporów obwodowych (oddziaływanie na opory obwodowe jest jedną z metod regulowania ciśnienia tętniczego). Tak więc, jeśli na „wejściu" tego bloku pojawi się jakiś opór, to na „wyjściu" otrzymamy odpowiednie ciśnienie. Funkcja przejścia tego bloku zależy od bieżącego zapotrzebowania na krew.

Blok „ŚCIANKI": podatność naczyń (a więc i impedancja) zależy od ciśnienia krwi. Po zadaniu na wejście bloku pewnego ciśnienia, otrzymujemy na wyjściu impedancję.

Blok „ZAPAS REGULACJI": aby zapewnić odpowiedni zapas regulacji, blok ten daje na wyjściu opór obwodowy większy od impedancji o A= niezbędny zapas.

Opór „wyliczony" przez ten ostatni blok stanowi wejście pierwszego bloku, co zamyka pętlę sprzężenia zwrotnego. Położenie punktu stabilnego tego układu zależy od właściwości bloków, od ich funkcji przejść. W szczególnym przypadku może nie być punktu stabilnego i wówczas lawinowy wzrost ciśnienia napotyka jedynie na ograniczenia „nieliniowe" takie, jak np. wydolność serca.

W tekście [7] było jeszcze m. innymi:

[...] tłumaczy kilka zjawisk obserwowanych w rzeczywistości, w tym przede wszystkim:

- *zwiększone ryzyko nadciśnienia przy stresującym trybie życia,*
- *możliwość dziedziczenia nadciśnienia,*
- *fakt, iż nadciśnienie (również to o podłożu dziedzicznym) pojawia się dopiero w pewnym wieku [...],*
- *jest to hipoteza cybernetyczna, nie stoi ona w sprzeczności z hipotezami np. biochemicznymi, a jedynie może zebrać te koncepcje w jedną całość.*

Skoro medycyna teoretyczna jest na razie tylko marzeniem, w niniejszej pracy nie będzie już mowy ani o teorii RVR ani o tym, po co organizm miałby stabilizować pewien poziom dynamiczności przepływu. Innymi słowy: w pracy tej zostanie omówiony – na podstawie wyników badań eksperymentalnych innych autorów – pewien mechanizm regulacyjny, jednak bez podawania, po co ten mechanizm istnieje. A przede wszystkim pokazane będzie, jak mechanizm ten, którego (ubocznym) celem jest stabilizacja ciśnienia, może ulec zwyrodnieniu i prowadzić nie do stabilizacji, a do destabilizacji.

2. Wprowadzenie

Z jednej strony uzasadnione byłoby nazywanie jednym terminem *nadciśnienie* wszystkich przypadków zwiększonego ciśnienia tętniczego, tak jak uzasadnione jest nazywanie jednym terminem *gorączka* wszystkich przypadków podwyższonej temperatury ciała. Byłoby, gdyby podwyższone ciśnienie traktować jako objaw czy przejaw wadliwego działania organizmu, np. choroby. Poza tym ewentualne, negatywne skutki nie zależą od tego, z jakiego powodu wystąpiło NT czy gorączka.

Jednak w przypadku podwyższonej temperatury nikt (dzisiaj) nie będzie twierdził, że mamy do czynienia zawsze z jedną i tą samą chorobą. Świadomość, iż mogą być różne przyczyny, takie jak infekcja, nowotwór, nadczynność tarczycy, przestawienie ośrodka termo-regulacji itd., tj. świadomość, że gorączka jest tylko objawem, pozwala na szukanie przyczyn i zwalczanie tych przyczyn, a nie samej gorączki.

Brak dokładnej wiedzy nt. genezy NT powoduje, że jakoś gubi się świadomość, iż nawet pierwotne NT jest tylko objawem, choć groźnym. Owszem, w diagnostyce dąży się do wykluczenia „normalnych" chorób, którym towarzyszy wtórne NT. Gdy jednak takiej choroby nie stwierdzi się, NT traktuje się jako samoistną jednostkę chorobową (a nie jako objaw) i zwalcza w dość zunifikowany sposób. Można powiedzieć, że jest to uzasadnione praktycznie, skoro geneza jest nieznana. Gorzej, gdy takie zunifikowane podejście stosuje się w badaniach mających na celu określenie genezy NT.

Niniejsza praca dotyczy NT, ale nie jest próbą szukania ogólnej genezy „choroby nadciśnieniowej". NT nie jest nawet punktem wyjścia prezentowanych rozważań. W niniejszej pracy analizowane jest zwyrodnienie pewnego mechanizmu regulacyjnego. NT jawi się jako rezultat – jeden z dwóch możliwych – tego zwyrodnienia. Zostanie bowiem pokazane, jak pewne ujemne sprzężenia zwrotne (SZ) w określonych warunkach zamieniają się na dodatnie SZ i zamiast przyczyniać się do stabilizacji ciśnienia, destabilizują je: albo je nadmiernie podnoszą, albo nadmiernie obniżają.

Należy jeszcze raz bardzo mocno podkreślić, że zaprezentowana tu teoria nie tłumaczy wszystkich możliwych rodzajów NT nazywanych wspólnym mianem: pierwotne NT. Wprost przeciwnie, teoria ta tłumaczy genezę pewnego rodzaju pierwotnego NT, ale też wskazuje, iż istnieje co najmniej jeszcze jeden rodzaj pierwotnego NT (pomija się tu

takie formy NT, jak np. tzw. „izolowane nadciśnienie skurczowe" związane ze starczym wzrostem sztywności tętnic, bowiem z punktu widzenia inżyniera nie jest to choroba tylko oczywisty rezultat zmniejszania się z wiekiem stałej czasowej; tak samo trudno za chorobę uznać np. pojawianie się zmarszczek lub siwizny).

3. Pętle sprzężeń zwrotnych

3.1. Dynamiczność przepływu krwi

Gdy wrzuci się kamień do wody o dużej powierzchni, powstanie fala, która oddali się, a w miejscu wrzucenia kamienia tafla wody będzie znowu gładka. Tak samo byłoby, gdyby układ tętniczy był jedną nieskończoną elastyczną rurą: porcja krwi, wyrzucona przez serce w czasie skurczu, oddalałaby się coraz bardziej od serca w postaci jednej fali. Jednak układ tętniczy jest skończony, jest ograniczony oporami obwodowymi. W tym wypadku, tak jak po wrzuceniu kamienia do wanny, fala będzie – w mniejszym lub większym stopniu – odbijać się od ograniczenia i dawać pewną liczbę fal wtórnych.

To, w jakim stopniu fala przepływu krwi odbije się od oporów, zależy od wzajemnej relacji pomiędzy:

- związkiem między ciśnieniem i przepływem krwi przez opory obwodowe, który to związek – nazywany opornością – wyrażany jest ilościowo jako:

  $$PR=P/f \quad (1)$$

  gdzie: f – przepływ krwi, P – ciśnienie tętnicze (dalej PR oznaczać będzie zarówno opory obwodowe, jak i ich oporność);
- związkiem między zaburzeniem ciśnienia i zaburzeniem przepływu, stanowiącymi falę, który to związek – nazywany impedancją (tętniczą) – wyrażany jest ilościowo jako:

  $$Za=\Delta P/\Delta f \quad (2)$$

  (gdzie: Δ – zaburzenie, tj. przyrost danej wielkości).

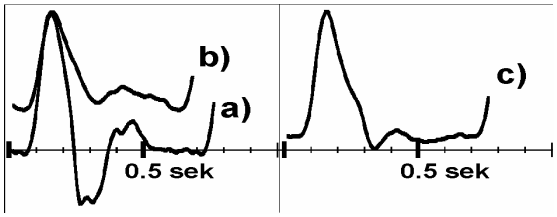
W fizjologii Za często nazywa się oporem sprężystym, gdyż zależy on od sprężystości tętnic: im sztywniejsze są tętnice, tym większego zaburzenia ciśnienia potrzeba dla danego zaburzeniu przepływu. Tak więc im sztywniejsze są tętnice, tym większa jest Za.

Stopień odbicia fal od ograniczenia, jakim jest PR, ilościowo określa współczynnik odbicia Γ wyrażony wzorem:

$$\Gamma=(PR-Za)/(PR+Za) \quad (3)$$

Oznacza on, że odbicia fal są tym mniejsze, im – przede wszystkim – mniejsza jest różnica między PR i Za. Dodatkowo odbicia maleją, jeśli sumaryczny opór sprężysty i niesprężysty jest większy. Powyższy wzór można intuicyjnie zinterpretować w następujący sposób: im fali jest łatwiej poruszać się w tętnicy niż przez PR, tym większa jej część pozostanie w tej tętnicy w postaci fali odbitej (dokładną analizę odbić fal można znaleźć w każdym opracowaniu dotyczącym zjawisk falowych generalnie, czy też zjawisk falowych w układzie krążenia).

Fale przepływu krwi odbijają się od PR ze zmianą znaku. Jeśli falę docelową, tj. falę poruszającą się w kierunku obwodu, tworzy przepływ w stronę obwodu (przepływ dodatni), to falę odbitą tworzy przepływ wsteczny (przepływ ujemny). I na odwrót. Szczegółowa analiza tego, jak z superpozycji fal docelowych i powrotnych powstaje wypadkowy obserwowany przepływ, zawarta jest w [8]. Natomiast upraszczając, można powiedzieć, że przepływ dodatni jest przez falę odbitą zmniejszany, a przepływ ujemny – zwiększany. Przy dużych odbiciach przepływ dodatni jest zmniejszany aż do wartości ujemnych, a przepływ ujemny jest podwyższany do wartości dodatnich (jak na rys. 2a).



Rys. 2. Dwa przykłady przepływu krwi w tętnicy ramieniowej o tym samym HR i zbliżonej wartości średniej przepływu (nieco większa w b). a) osoba młoda (elastyczne naczynia): duże odbicia fal + mała prędkość propagacji fal = duża dynamiczność, b) osoba w starszym wieku (sztywne naczynia): małe odbicia + duża prędkość = mała dynamiczność

Na to, w jakim stopniu fala odbita zmniejsza przepływ dodatni (zwiększa ujemny), ma również wpływ prędkość propagacji fal, czyli czas, po jakim fala odbita powróci z miejsca odbicia. Im wcześniej fala odbita nałoży się na falę docelową, tym (a) w większym stopniu zmniejszy przepływ maksymalny oraz (b) w mniejszym stopniu zmniejszy późniejszy przepływ. W efekcie zmiany przepływu będą mniej dynamiczne (jak na rys. 2b).

Prędkość propagacji zależy od sztywności tętnic, podobnie jak Za. W sumie więc zarówno z powodu spadku odbić jak i wcześniejszego powrotu fal odbitych, dynamika zmian przepływu jest tym mniejsza, im sztywniejsze są naczynia. Jeśli wielkość odbić zmienia się z powodu zmiany podatności (a nie z powodu zmiany PR – wzór (3)), to zmienia się też prędkości propagacji, co w pewnym sensie wzmacnia wpływ zmiany odbić. Dlatego poniżej, by nie zaciemniać obrazu jednoczesną analizą wpływu zmian odbić i zmian prędkości równocześnie, mowa będzie tylko o zmianie odbić. Czytelnik powinien jednak pamiętać, że wpływ zmian odbić fal ulega „wzmocnieniu" przez zmiany prędkości propagacji.

Reasumując: im większa jest podatność naczyń, tym większa jest dynamika zmian przepływu krwi, która dalej nazywana będzie w skrócie dynamicznością.

Dla tej pracy fundamentalne znaczenie ma to, że na śródbłonek oddziałuje dynamiczność, która, zależąc od wielkości odbić, zależy od wielkości różnicy PR-Za.

3.2. Starzenie a dynamiczność przepływu

Jednym z głównych przejawów starzenia się układu krążenia jest sztywnienie tętnic. Starzenie się ma duży wpływ na budowę i funkcjonowanie naczyń [2], [5]. I tak np. zmiany w błonie środkowej polegają między innymi na zmianie stosunku ilości elastyny do kolagenu na korzyść tego ostatniego. Dodatkowo, nowo utworzone włókna kolagenowe są w pewnym sensie patologiczne w porównaniu z włóknami w młodych naczyniach, ponieważ ze względu na nieuporządkowaną budowę i nieodpowiednią długość są

rekrutowane zbyt wcześnie, uniemożliwiając elastynie wypełnianie swojej funkcji. W młodych naczyniach włókna kolagenowe są rekrutowane dopiero przy nadmiernym rozciągnięciu naczynia. Z biegiem lat zwiększa się również ilość wiązanych jonów („zwapnienie" naczynia). Oba te czynniki powodują usztywnienie tętnicy. Omówione tu uogólnione, fizjologiczne zmiany błony środkowej są niezależne od ewentualnych zmian miażdżycowych, będących lokalną chorobą błony wewnętrznej, co np. wykazano w [14] przez porównanie populacji japońskiej z amerykańską (u obu podobne usztywnienie związane z wiekiem, ale zmiany miażdżycowe tylko u Amerykanów).

Również to, tj. związane ze starzeniem się, zwiększenie sztywności tętnic, powoduje wzrost wartości takich parametrów falowych, jak impedancja Za czy prędkość propagacji fal [1], [11], [18]. To z kolei prowadzi do zmniejszenia dynamiczności przepływu poprzez m.in. zmniejszenie współczynnika odbić Γ (wzór (3)). Dlatego dynamiczność przepływu krwi u osób starych jest znacznie mniejsza niż u młodych [6], [8] (rys. 2).

3.3. Dynamiczność a aktywność śródbłonka

Jeszcze całkiem niedawno uważano, że śródbłonek tętnic jest jedynie bierną ochroną ścianek tętnic. Zdawał się to potwierdzać fakt, że uszkodzenie śródbłonka może prowadzić do miażdżycy. W ostatnich latach ukazała się jednak wielka liczba prac wykazujących, że śródbłonek jest swego rodzaju rozłożonym w przestrzeni gruczołem, który produkuje i wydziela do krwi wiele ważnych substancji. Istotnymi dla tej pracy są substancje wazoaktywne [21]:

- wazodylatory (EDRF) – dalej, dla uproszczenia, będą one reprezentowane przez tlenek azotu (NO), który wydaje się mieć najsilniejsze działanie,
- grupa substancji wykazujących działanie wazokonstrykcyjne (EDCF) – dalej, dla uproszczenia, będą one reprezentowane przez endotelinę-1 (ET-1), która wykazuje najsilniejsze działanie.

Najważniejszym jednak faktem jest to, że natężenie sekrecji tych substancji zależy od sposobu, w jaki krew przepływa przez tętnice. Stosunkowo mało prac poświęcono wpływowi sposobu przepływu krwi na sekrecję ET-1 ([15,17,23]). Za to wiele prac dotyczy zależności sekrecji NO od sposobu przepływu. Są to zarówno badania wyizolowanych komórek śródbłonka, jak i badania pracy śródbłonka in vivo (np. [13], [15], [17], [22], [23], [24]). Można więc uznać za w pełni udowodnione, iż przepływ pulsacyjny powoduje wiele intensywniejszą sekrecję NO niż przepływ nie-pulsacyjny. Należy jednak uwzględnić, że:

- głównie badano sekrecję NO przy przepływie stałym versus przepływ pulsacyjny, tj. bez dokładniejszego ilościowego różnicowania sposobu pulsowania, ale
- skoro przepływ turbulentny lub oscylacyjny symulujący turbulencje za stenozą zdaje się zmniejszać, a nie zwiększać, sekrecję NO [15], [24], to wydaje się, iż dla intensywności sekrecji ważne jest to, czy dynamiczność przepływu jest naturalna (patrz: Waveform Age jako indeks fizjologiczny w [25]), czy zaburzona, jednak
- jedyna znana autorowi praca analizująca wpływ naturalnej dynamiczności na aktywność wydzielniczą śródbłonka, dotyczy sekrecji prostacykliny, a nie NO [16] (wymuszano tam dwa rodzaje przepływu o dynamicznościach prawie identycznych jak na rys. 2 – sekrecja prostacykliny była znacznie większa dla dużej dynamiczności, podobnej do tej z rys. 2a).

Dlatego też należy uznać, że dziś już w dużej części jest udowodniona, ale jednak nie do końca, następującą Teza I:

> Naturalna dynamiczność przepływu krwi wpływa na sekrecję EDRF w taki sposób, że im większa jest dynamiczność, tym intensywniejsza jest sekrecja.

Można przypuszczać, że teza ta w całości zostanie wcześniej czy później udowodniona eksperymentalnie. Główną przeszkodą wydaje się to, iż inni autorzy nie dysponują ilościową miarą dynamiczności naturalnej, a wskaźnik Waveform Age, opracowany przez autora (np. [25]), nie stał się powszechnie znany, gdyż nie był on powszechnie znany (choć pierwotną, prymitywną jego wersję udało się jednak zaprezentować w [6]).

Nieco bardziej hipotetyczną, jednak równie prawdopodobną, jest Teza II:

> Naturalna dynamiczność przepływu krwi wpływa na sekrecję EDCF w taki sposób, że im mniejsza jest dynamiczność, tym intensywniejsza jest sekrecja.

3.4. Pętle sprzężeń zwrotnych

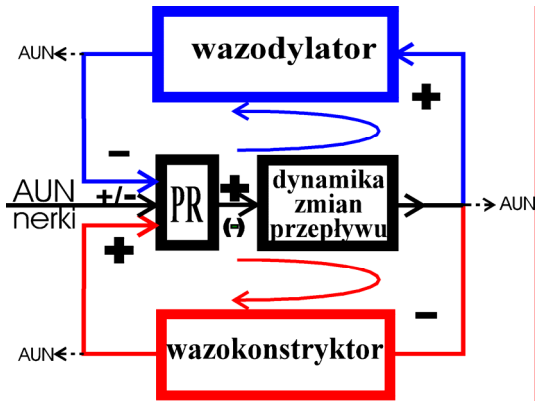
Opisane powyżej fakty pozwalają na wykazanie istnienia sprzężeń zwrotnych (SZ) między PR i aktywnością wydzielniczą śródbłonka (rys. 3).

Istnienie pierwszej pętli wydaje się być pewne. W pętli tej zmiana dynamiczności powoduje taką samą co do znaku zmianę wydzielania EDRF (Teza I). Zmiana stężenia tych wazodylatorów powoduje odwrotną co do znaku zmianę PR (z definicji). Zmiana PR powoduje taką samą, co do znaku, zmianę wielkości odbić (wzór (3)). A że zmiana odbić oznacza taką samą co do znaku zmianę dynamiczności, zamyka to pętlę SZ. Ponieważ w tej pętli zmiana znaku wystąpiła jeden raz (+EDRF→-PR), mamy do czynienia z ujemnym SZ.

W pętli drugiej mamy analogiczną sytuację, tylko że: (a) zmiana dynamiczności powoduje przeciwną co do znaku zmianę wydzielania ET-1 oraz (b) zmiana stężenia ET-1 powoduje taką samą co do znaku zmianę PR. Skoro znowu mamy jedną zmianę znaku w pętli (+dynamiczność→-ET-1), ponownie mamy do czynienia z ujemnym SZ.

Stabilizacja PR (lub stabilizacja dynamiczności - w zależności od tego, o co naprawdę chodzi) przebiega następująco: (a) jeśli PR jest nadmiernie wysokie, to poprzez zwiększoną sekrecję NO i zmniejszoną sekrecję ET-1, dochodzi do zmniejszenia PR, (b) jeśli PR jest za małe, to poprzez zwiększoną sekrecję ET-1 i zmniejszoną sekrecję NO dochodzi do wzrostu PR.

Należy podkreślić, iż rola ET-1 jest równie ważna, jak rola NO. Fakt ten wydaje się być zdecydowanie niedoceniany. Prawdopodobnie z powodu chwilowej mody, NO jest obecnie uważany za „cudowną" substancję (prawie panaceum), natomiast ET-1 jest traktowana jako „szkodnik". A przecież działanie ET-1 jest pewnym analogiem układu współczulnego, którego nikt chyba nie nazwie układem szkodliwym.

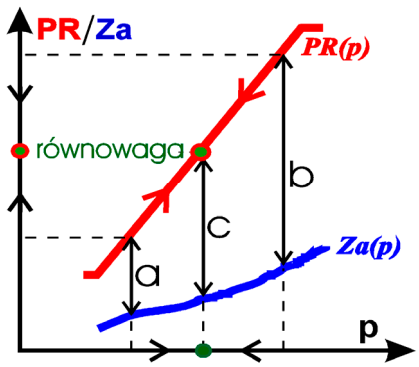


Rys. 3. System podstawowych ujemnych sprzężeń utrzymujących dynamikę zmian przepływu na poziomie wymaganym w danej chwili przez organizm. Pętla 1 (niebieska): sprzężenie zwrotne ujemne z wazodylatacją. Pętla 2 (czerwona): sprzężenie zwrotne ujemne z wazokonstrykcją. Obie pętle pracują wspólnie, choć w pewnym sensie przeciw sobie, jak np. układy: przywspółczulny i współczulny (EDRF i EDCF ich lokalnymi odpowiednikami). Jeżeli jednak sprzężenie między PR i dynamicznością zmieni znak z dodatniego na ujemny, w obu pętlach ujemne sprzężenie zwrotne zamieni się w sprzężenie dodatnie, będące przyczyną omawianego tu typu nadciśnienia (lub niedociśnienia – w zależności od znaku odchyłki, która będzie ulegać wzmocnieniu przez teraz już dodatnie sprzężenia)

Prawdopodobnie pętla z NO pracuje jako układ nadążny, np. szybko reaguje na małe uchyby i fluktuacje dynamiczności: spontaniczne lub powodowane przez czynniki zewnętrzne (np. wpływ oddechu [25]). Szybka reakcja wymaga, by nośnik sygnału nie ulegał kumulacji. W przypadku układu nerwowego mamy np. wychwyt zwrotny neurotransmiterów. W przypadku pętli z NO odpowiednikiem wychwytu byłby krótki czas istnienia NO we krwi (kilka sekund). Stąd właśnie przypuszczenie o roli pętli 1 jako układu nadążnego. Odkryty przez autora [9] komponent o bardzo wysokiej częstotliwości (VHF) (okres 2-5 sek.) w zmienności dynamiczności, który nie występuje w zmienności rytmu serca, może być właśnie przejawem działania pętli 1 jako układu nadążnego.

Pętla z ET-1 miałaby więc głównie za zadanie zgrubną stabilizację PR. Być może to ona odpowiada za komponent o bardzo niskiej częstotliwości w zmienności dynamiczności (okres ok. 1 min), który w zmienności rytmu serca jest znacząco mniej istotny [9].

Wyniki podane w [25] sugerują, że to dynamiczność wpływa na autonomiczny układ nerwowy. Skoro tak późno odkryto, iż dynamiczność wpływa na aktywność śródbłonka, to być może istnieją nie odkryte jeszcze receptory dynamiczności. Chyba jednak bardziej prawdopodobnym rozwiązaniem jest istnienie receptorów reagujących na substancje z grupy NO lub z grupy ET-1. Prawdopodobnie nie chodzi o NO, gdyż, jeśli pętla z NO działałaby jako układ nadążny i byłaby odpowiedzialna za komponent VHF w zmienności dynamiczności, to chyba powinien też istnieć taki komponent w zmienności rytmu serca, a nie istnieje [9]. Z drugiej jednak strony obecnie uważa się, że w roli neuromediatora występuje NO, a nie ET-1. Poza tym szybkość (po ok. 2 sek) i precyzja, z jaką zmiany rytmu serca powtarzają zmiany dynamiczności [25], wymagają, by nośnik informacji nie ulegał kumulacji, co wskazywałoby jednak na NO. Sprawa jest więc otwarta: musi być rozstrzygnięta w bezpośrednich eksperymentach.



Rys. 4. Zależność oporu obwodowego PR i impedancji tętniczej Za od ciśnienia tętniczego P. Gdy nachylenie Za(P) jest mniejsze niż nachylenie PR(P), wzrost ciśnienia powoduje wzrost różnicy PR-Za, co pociąga za sobą wzrost odbić fal przepływu, a więc i wzrost dynamiczności przepływu. (a) Przy zbyt małym ciśnieniu, tj. zbyt małej dynamiczności, wpływ sekrecji E-1 przewyższa wpływ sekrecji NO, co powoduje wzrost PR. (b) Przy zbyt dużym ciśnieniu, tj. zbyt dużej dynamiczności, wpływ NO przewyższa wpływ E-1, co powoduje spadek PR. (c) Dzięki ujemnemu sprzężeniu zwrotnemu (między dynamicznością i sekrecją działających przeciwnie substancji) istnieje punkt równowagi

4. Warunki stabilności

Bardziej szczegółowo działanie układu SZ z rys. 3 można prześledzić, posługując się rys. 4. Przede wszystkim rys. 4 zawiera:

- **zależność PR od ciśnienia:** W zakresie ciśnień, w którym normalnie pracuje układ tętniczy, przepływ krwi przez większość organów słabo zależy od wartości ciśnienia z powodu autoregulacji przepływu [21]. Skoro więc PR=P/f (wzór (1)) i f=const (w przybliżeniu), to PR jest proporcjonalne do P. Dlatego na rys. 4 zależność między P i PR przedstawiono w postaci linii prostej (skoro f słabo, ale jednak rośnie ze wzrostem ciśnienia, w rzeczywistości powinna to być krzywa coraz bardziej pozioma w miarę wzrostu P). Dla wysokich ciśnień możliwości autoregulacji załamują się, co symbolizuje wysoki, ale stały, niezależny od ciśnienia PR. Tak więc w normalnym zakresie ciśnień, skoro PR=(1/f)P, to nachylenie zależności PR od P (NPR) równe jest odwrotności przepływu krwi.
- **zależność Za od ciśnienia:** Podatność tętnic zależy od stopnia ich rozciągnięcia np. w związku z rekrutacją włókien kolagenowych, co powoduje, iż ze wzrostem ciśnienia rośnie zarówno prędkość propagacji fal, jak i Za [1,10-12,18,20]. W przypadku młodych, elastycznych tętnic, Za jest ok. 10 razy mniejsza niż PR (co wynika z analizy wielkości odbić, np. [8], oraz z badań bezpośrednich, np. [12,18]). Natomiast dla bardzo wysokich ciśnień, gdy tętnica jest maksymalnie rozciągnięta, Za jest bardzo duże (jak w przypadku sztywnej rury). W efekcie zależność Za od P przedstawia krzywą coraz bardziej pionową w miarę wzrostu ciśnienia. Skoro dla młodych tętnic Za jest ok. 10 razy mniejsze niż PR, to i nachylenie zależności Za od P (NZa) jest wiele mniejsze niż NPR.

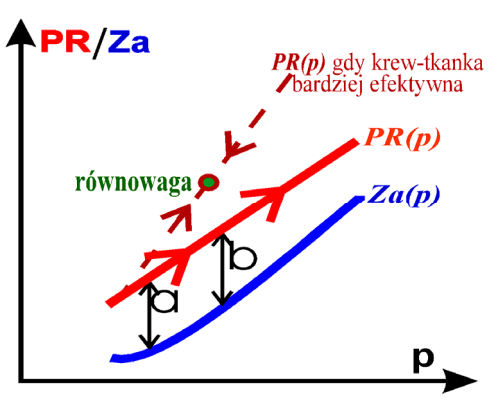
Jeśli więc, jak na rys. 4, NZa<NPR, to wzrost ciśnienia zwiększy bardziej PR niż Za. Zgodnie ze wzorem (3) spowoduje to wzrost odbić fal, a więc i wzrost dynamiczności. Wzrost dynamiczności zwiększy sekrecję NO, a zmniejszy

sekrecję ET-1, co pociągnie za sobą spadek PR. Mechanizmy dopasowujące ciśnienie do poziomu PR (objętość końcowo-skurczowa serca itd.) spowodują spadek ciśnienia. I tak zamyka się pętla ujemnego SZ stabilizującego ciśnienie.

Rys. 5. Jak na rys. 4, ale nachylenie Za(p) jest większe niż nachylenie PR(p) dla wyższych ciśnień (linia ciągła) lub dla całego zakresu ciśnień (linia przerywana). W tym wypadku mamy do czynienia nie z ujemnym sprzężeniem zwrotnym między dynamicznością i sekrecją E-1 oraz NO, a ze sprzężeniem dodatnim dla wyższych ciśnień (lub dla całego zakresu ciśnień). Brak punktu równowagi może manifestować się nadciśnieniem (jak na tym rysunku) lub – prawdopodobnie rzadziej - niedociśnieniem. a, b i c – jak na rys. 4, d – obszar braku punktu równowagi

Jeśli jednak z powodu wzrostu NZa (rys. 5) lub zmalenia NPR (rys. 6) NZa stanie się większe niż NPR przynajmniej dla wyższych zakresów ciśnień, to zamiast ujemnego SZ pojawi się dodatnie SZ, jak to wytłumaczono na rys. 5 i 6 (należy przypomnieć, że prędkość propagacji fal rośnie wraz ze wzrostem ciśnienia z powodu spadku podatności, więc dynamiczność może spadać w miarę wzrostu ciśnienia nawet wtedy, gdy NZa nie jest większe od NPR).

Sytuacja staje się szczególnie krytyczna, gdy ciśnienie wzrośnie na tyle, że mechanizmy autoregulacji zaczną zawodzić i NPR stanie się równe zeru (końcowy poziomy odcinek zależności PR od P na rys. 5).



Rys. 6. Jeśli względna efektywność krwi-tkanki jest zmniejszona, to niezbędny jest większy przepływ **f** krwi-cieczy, co powoduje, że nachylenie PR(P)=P/**f** zmaleje. Jeśli nachylenie PR(P) będzie mniejsze niż nachylenie Za(P) (ciągła, czerwona linia), to rozwijać się będzie nadciśnienie z powodu braku punktu równowagi. Zwiększenie efektywności krwi, np. za pomocą diuretyku, zwiększa nachylenie PR(P) i przywraca sprzężenie zwrotne ujemne (przerywana, brązowa)

5. Czynniki wpływające na sprzężenia

Praktycznie wszystkie czynniki, o których wiadomo, że wpływają dodatnio lub ujemnie na przebieg NT, modyfikują lub mają związek z NZa lub NPR.

Dziedziczność. Wydaje się oczywistym, że genotyp ma zasadniczy wpływ na budowę ścianek tętnic, tak jak na budowę wszystkich innych organów i części ciała. A od budowy ścianek zależy zarówno wielkość Za, jak i NZa. W przypadku niektórych genotypów żadne zmiany, np. te opisane niżej (w tym starzenie), nie są w stanie doprowadzić do tego, że NZa będzie większe od NPR. W przypadku innych genotypów nawet niewielkie zmiany mogą na tyle zmodyfikować NZa, iż ujemne SZ zamieni się w dodatnie i pojawi się NT.

Starzenie. Jak już wyżej wspominano, starzenie się układu tętniczego polega głównie na spadku podatności tętnic, tj. na generalnym wzroście Za, a więc i na wzroście NZa. Dodatkowo, wcześniejsze rekrutowanie włókien kolagenowych powoduje, iż – jak na rys. 5 – Nza może być większe niż NPR dla wyższych ciśnień, podczas gdy dla ciśnień niższych mamy wciąż jeszcze NZa<NPR. Manifestowałoby się to chwiejnym NT poprzedzającym trwałe NT.

W szczególności, związane z rytmem okołodobowym przechodzenie z zakresu ciśnień, dla którego istnieje ujemne SZ do zakresu, w którym SZ jest dodatnie, odpowiedzialne byłoby za nadmiernie duże różnice między ciśnieniem w nocy i w dzień.

Aktywność fizyczna. Ćwiczenia fizyczne zwiększają podatność układu tętniczego, jakkolwiek chyba wciąż jeszcze nie ma zgodności co do tego, jaki jest tego mechanizm [3]. Zwiększenie podatności oznacza zmniejszenie NZa. I tak, w przypadku, gdy istnieje ujemne SZ, tj. NZa jest wciąż jeszcze mniejsze niż NPR i istnieje punkt równowagi, ćwiczenia przesuwają go w stronę niższych wartości ciśnienia. W przypadku, gdy już istnieje dodatnie SZ, ćwiczenia mogą same lub łącznie z innym czynnikiem odrestaurować ujemne SZ. Dodatkowo, ćwiczenia, poprawiając efektywność krwi-tkanki, zwiększają NPR (rys. 6).

Diuretyki. W ostatnich latach są one bardzo polecane jako podstawowy lek w monoterapii, a szczególnie jako jeden z leków w terapii skojarzonej [4]. Z punktu widzenia prezentowanej tu teorii jest to jak najbardziej uzasadnione. Diuretyki powodują, że ta sama objętość krwi-tkanki dzięki zmniejszonej zawartości wody zajmuje mniejszą objętość jako krew-ciecz. A to oznacza, że ten sam przepływ krwi-tkanki odpowiada mniejszemu przepływowi krwi-cieczy, dzięki czemu NPR rośnie. Wzrost NPR, powodując wzrost dynamiczności przepływu, przyczynia się do przesunięcia punktu równowagi w stronę niższych wartości ciśnienia, jeśli istnieje jeszcze ujemne SZ lub może odrestaurować ujemne SZ, jeśli wcześniej SZ zmieniło się na dodatnie (rys. 6).

Spożywanie soli. Oczywiście, mechanizm jest odwrotny do działania diuretyków: większe uwodnienie krwi powoduje, że ten sam przepływ krwi-tkanki związany jest z koniecznością większego przepływu krwi-cieczy, co zmniejsza NPR. W młodym wieku NZa jest na tyle małe, że NPR jest wciąż od niego większe mimo spożywania nadmiaru soli. Co najwyżej punkt równowagi będzie nieco przesunięty w stronę wyższych ciśnień, ale wciąż będzie ujemne SZ. Jednak wcześniej czy

później może przyjść taki moment, że NZa wzrośnie do poziomu NPR zmniejszonego przez nadmiar soli. Ograniczenie spożycia soli opóźnia nadejście tego momentu.

Powyższe uwagi dotyczące diuretyków i soli nie stoją w sprzeczności z dotychczasowymi poglądami, które podkreślają wpływ zwiększonej objętości krwi krążącej na wzrost ciśnienia poprzez zwiększenie powrotu żylnego itd. Postulowany tu mechanizm może być drugim sposobem, w jaki spożywanie soli czy przyjmowanie diuretyków wpływa na wartość ciśnienia. Zastanawiające jest jednak, dlaczego nadużywanie soli nie powoduje NT u zdrowych, młodych osób (przecież u nich również np. działa prawo Franka-Starlinga), natomiast jest tak istotne w NT: ograniczenie spożywania soli u osób z NT może znacząco przyczynić się do obniżenia ciśnienia. Mechanizm postulowany w pracy wyjaśnia tę zagadkę: spożycie soli nabiera znaczenia dopiero wtedy, gdy NZa i NPR zaczynają mieć porównywalne wartości.

Otyłość. Im większe jest ciało, tym większy jest niezbędny przepływ krwi-tkanki, a więc i przepływ krwi-cieczy (choćby z powodu zwiększonej pracy mechanicznej niezbędnej do manipulowania dużą masą ciała). Dalej jest jak przy nadmiernym spożywaniu soli. Jeśli nawet w młodszym wieku tętnice otyłej osoby są na tyle elastyczne, że NZa jest wciąż jeszcze mniejsze niż NPR, to: (a) punkt równowagi osiągany dzięki ujemnemu SZ i tak musi się przesunąć do zakresu wyższych wartości ciśnienia, (b) wzrost NZa związany ze starzeniem się układu tętniczego (dodatkowo przyspieszonym otyłością?) powoduje, iż ujemne SZ wcześniej zamienia się w dodatnie SZ. I na odwrót: schudnięcie osoby z podwyższonym ciśnieniem: (a) przesuwa punkt równowagi w stronę niższych ciśnień, (b) może wystarczyć, by dodatnie SZ ponownie stało się ujemnym SZ bez potrzeby stosowania terapii hipotensyjnej.

Palenie. Palenie tytoniu ma wieloraki wpływ na układ krążenie, choćby poprzez wpływ na AUN. Dla prezentowanej tu teorii szczególne znaczenie ma upośledzenie utlenowania krwi (zarówno z powodu upośledzonej pracy płuc jak i blokowania hemoglobiny przez tlenek węgla). Krew-tkanka gorzej przenosząca tlen do tkanek organizmu jest mniej efektywna, wobec czego niezbędny jest zwiększony przepływ krwi-cieczy. To oznacza zmniejszenie się NPR. W konsekwencji: zmniejszenie odbić, zmniejszenie dynamiczności, spadek sekrecji NO i wzrost sekrecji E-1, wzrost oporów, wzrost ciśnienia.

Cukrzyca. Być może upośledzone przenikanie glukozy do wnętrza komórek jest traktowane jako zmniejszona efektywność krwi-tkanki. Gdyby tak było, rozwój NT byłby analogiczny jak w przypadku palenia. Oczywiście, byłby to dodatkowy mechanizm wiążący cukrzycę z NT.

Bezdech senny. Przerwy w oddychaniu prowadzą do periodycznych epizodów słabego utlenowania krwi. Mogłoby to powodować, że krew-tkanka staje się mniej efektywna, w związku z czym musi się zwiększyć przepływ krwi-cieczy, i dalej tak, jak w przypadku palenia. Jednak równie prawdopodobną reakcją na periodyczne, ale chwilowe sytuacje niedotlenienia jest wzrost oporów obwodowych w mniej ważnych regionach ciała. Tak czy inaczej, taki conocny trening mięśniówki naczyń powoduje przyrost ich masy i sprawności, prowadząc do NT, ale raczej innego typu, tj. NT ze zwiększoną dynamicznością (choć utrzymujące się uczucie zmęczenia sugerowałoby jednak zmniejszoną dynamiczność).

6. Podsumowanie

W pracy przedstawiono pewien mechanizm, który jeśli ulegnie degeneracji, prowadzi do NT. Należy podkreślić kierunek implikacji: nie „istnieje NT $\Rightarrow$ mamy do czynienia z degeneracją mechanizmu", tylko „nastąpiła degeneracja $\Rightarrow$ wystąpi NT". Oznacza to, że nie każde NT ma genezę, jak ta opisana powyżej. Wstępne wyniki badań autora (dane niepublikowane) pokazują, iż osoby z NT mają zazwyczaj dynamiczność albo mniejszą, albo większą niż norma dla ich wieku. Już to sugeruje, że istnieje inny typ NT, który jest związany z nadmiernie wysokimi oporami obwodowymi, ale nie z powodu opisanego wyżej mechanizmu. Może to być np. wadliwość działania układu renina-angiotensyna. Wydaje się, że najprostszym sposobem różnicowania byłby pomiar dynamiczności, a szczególnie sprawdzenie, czy przy obniżeniu oporów obwodowych dynamiczność spada, jak zazwyczaj bywa (wg wzoru (3)), czy też – paradoksalnie - rośnie. Być może wskazówką byłaby też duża różnica między ciśnieniem dziennym i nocnym (ale tylko w przypadkach jak na rys. 6).

Różna geneza NT u konkretnego pacjenta wymagałaby różnego rodzaju terapii, np. inhibitory ACE przy wadliwości układu renina-angiotensyna i diuretyk przy opisywanym tu typie NT. Jednak szersza analiza działania leków hipotensyjnych wykracza poza ramy tej pracy.

Prezentowany tu mechanizm powstawania NT dotyczy z pewnością tylko okresu początkowego. Należy przypuszczać, że tak jak w przypadku mięśniówki naczyń, gdzie najpierw mamy do czynienia z nadmiernym skurczem, potem z adaptacyjnym przerostem, tak samo w przypadku EDRF i EDCF najpierw mamy do czynienia z czynnościową zmianą sekrecji, potem z adaptacyjnym zanikiem zdolności sekrecji EDRF i wzrostem zdolności sekrecji EDCF.

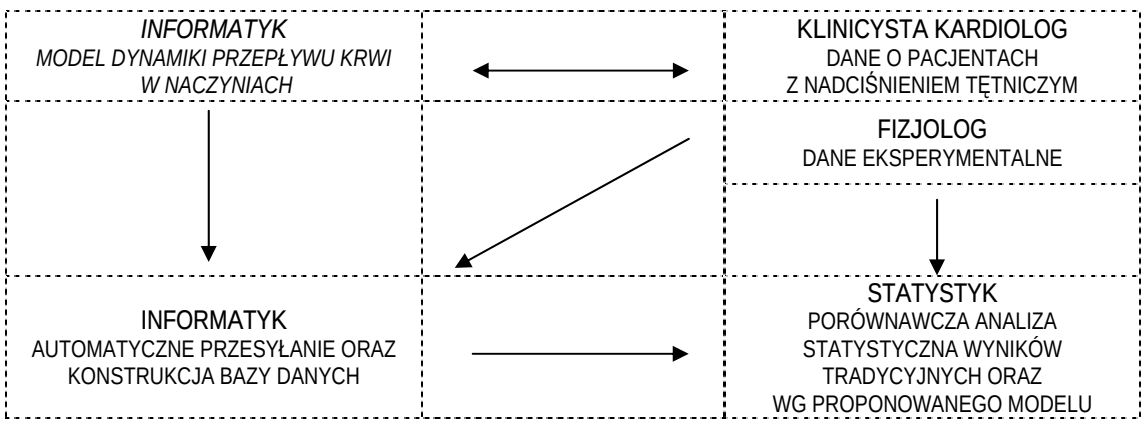
REASUMUJĄC: Przedstawiono tu pewien mechanizm prowadzący do jednego z typów NT, bazując na:

- faktach eksperymentalnych innych autorów, dotyczących zależności sekrecji wazoaktywnych substancji od sposobu przepływu krwi,
- logicznym założeniu, że różnice w sekrecji nie są skokowe, tj. tak jak ciągłe jest przejście między przepływem stałym i pulsacyjnym, tak ciągłe jest przejście między słabą sekrecją EDRF przy przepływie stałym i intensywniejszą sekrecją dla przepływu pulsacyjnego (i na odwrót w przypadku EDCF),
- hipotezie, że ta ciągłość zmian sekrecji odpowiada ciągłości zmian dynamiczności, która zależy od wielkości odbić fal przepływu.

O faktach trudno dyskutować, a logiczne założenie wydaje się jak najbardziej uzasadnione. Hipoteza również wydaje się uzasadniona, ponieważ: (a) jest ona faktem eksperymentalnym w przypadku prostacykliny, (b) byłoby rzeczą naturalną, jak w wielu innych przypadkach (np. ACTH i kora nadnerczy), że sekrecja substancji wazoaktywnych zależałaby zwrotnie od powodowanych przez nie skutków, (c) dynamiczność wydaje się bardzo ważnym parametrem przepływu krwi (np. jej związek z rytmem serca [25]).

Literatura cytowana

1. Avolio A.: Ageing and wave reflection, J. Hypertension, 10 (suppl 63): S83-S86. 1992.
2. Baskett J. J.: Biophysical ageing in systemic arteries, In: Isaacs (ed): Recent advances in geriatric medicine II, Churchill Livingstone, Edinburg 1982.
3. Cameron J., Dart A.: Exercise trainnig increases total systemic arterial compliance in humans, Am. J. Physiol., 266 (Heart Circ Physiol 35): H693-H701, 1994.
4. Chobanian A. V., Bakris G. L., Black H. R., et al.: Seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure, Hypertension, 42: 1206-1252, 2003.
5. Dobrin P. B.: Mechanical properties of arteries, Physiol. Rev., 58, 1978.
6. Gólczewski T., Krajewski A.: A vector based approach to age-related changes of Doppler velocity waveforms in the brachial artery, Ultrasound Med. Biol., 13: 15-18, 1987.
7. Gólczewski T.: Nowa koncepcja genezy samoistnego nadciśnienia tętniczego, W: Sprawozdanie za rok 1992 z realizacji tematu 15/St/92, Opracowanie wewnętrzne IBIB PAN (nie publikowane).
8. Gólczewski T.: Metoda prezentacji wybranych wielkości hemodynamicznych jako wektorów w przestrzeni $L^2(T)$ dla różnicowania stanu biologicznego – Analiza prędkości przepływu krwi w tętnicach dla oceny zaawansowania procesu starzenia się, Prace IBIB PAN, 38, 1994.
9. Gólczewski T.: Analogiczna do HR Variability, krótko-okresowa zmienność kształtu krzywej prędkości przepływu krwi, Materiały: X Konferencja Naukowa „Biocybernetyka i Inżynieria Biomedyczna": 72-76, 1997.
10. Januszewicz A., Januszewicz W., Szczepańska E., Sznajderman M. (red): Nadciśnienie tętnicze, Wyd. Medycyna Praktyczna, Kraków 2004.
11. Kawasaki T., Sasayama S., Yagi S. I., Asakawa T., Hirai T.: Non-invasive assesement of age related changes in stiffness of major branches of the human aorta, Cardiovasc. Res., 21: 678-687, 1987.
12. Merillon J. P., Fontenier G. J., Lerallut J. F., et al. : Aortic input impedance in normal man and arterial hypertension: its modification during changes in aortic pressure, Cardiovasc. Res., 16: 646-656, 1982.
13. Nakano T., Tominaga R., Nagano I., Okabe H., Yasui H.: Pulsatile flow enhances endothelium-derived nitric oxide release in the peripheral vasculature, Am. J. Physiol. Heart Circ. Physiol., 278: H1098-H1104, 2000.
14. Nakashima T., Tanikawa J.: A study of human aortic distensibility with relation to atherosclerosis and ageing, Angiology, 22: 477-490, 1971.
15. Noris M., Morigi M., Donadelli R., et al.: Nitric oxide synthesis by cultured endothelial cells is modulated by flow conditions, Circ. Res., 76(4): 536-43, 1995.
16. Ohonara T., Okadome K., Yamamura S., Mii S., Sugimachi K.: Simulated blood flow and the effects on prostacyclin production in the dog femoral artery, Circ. Res., 68: 1095-1099, 1991.
17. Orime Y., Shiono M., Hata H., et al. : Cytokine and endothelial damage in pulsatile and nonpulsatile cardiopulmonary bypass, Artif. Organs, 23(6): 508-512, 1999.
18. O'Rourke M. F.: Arterial function in health and disease, Churchill Livingstone, Edinburgh 1982.
19. O'Rourke M. F.: Basic concepts for the understanding of large arteries in hypertension, J. Cariovasc. Pharmacol., 7 (sup.2) S14-S21, 1985.
20. Simon A. C., Safar M. E., Levenson J. A., et al.: An evaluation of large arteries compliance in man, Am. J. Physiol., 237 (Heart Circ Physiol 5): H550-H554, 1979.
21. Traczyk W., Trzebski A. (red): Fizjologia człowieka z elementami fizjologii stosowanej i klinicznej, PZWL, Warszawa 1990.
22. Vedrinne C., Tronc F., Martinot S., et al.: Effects of various flow types on maternal hemodynamics during fetal bypass: is there nitric oxide release during pulsatile perfusion? J. Thorax. Cardiovasc. Surg., 116(3): 432-9, 1998.
23. Vedrinne C., Tronc F., Martinot S., et al.: Better preservation of endothelial function and decreased activation of the fetal renin-angiotensin pathway with the use of pulsatile flow during experimental fetal bypass, J. Thorax. Cardiovasc. Surg., 120(4): 770-7, 2000.
24. Ziegler T, Bouzourene K, Harrison VJ, Brunner HR, Hayoz D. Influence of oscillatory and unidirectional flow environments on the expression of endothelin and nitric oxide synthase in cultured endothelial cells. Arterioscler Throm Vasc Biol 18(5):686-92, 1998
25. Gólczewski T.: Na autonomiczny układ nerwowy wpływa dynamika zmian przepływu krwi, VI Ogólnopolska Konferencja Naukowa: Modelowanie Cybernetyczne Systemów Biologicznych, Kraków 2005.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 81-86.

NA AUTONOMICZNY UKŁAD NERWOWY WPŁYWA DYNAMIKA ZMIAN PRZEPŁYWU KRWI

DYNAMICS OF FLOW ALTERATION IS THE FEATURE OF PULSATILE BLOOD FLOW THAT INFLUENCES THE AUTONOMIC NERVOUS SYSTEM

TOMASZ GÓLCZEWSKI

Instytut Biocybernetyki i Inżynierii Biomedycznej PAN, Warszawa, tgol@ibib.waw.pl

Streszczenie. CEL: Określenie, która cecha pulsacyjnego przepływu krwi oddziałuje na autonomiczny układ nerwowy (AUN): czy sam fakt pulsacyjności, wyrażany za pomocą powszechnie znanego Pulsatility Index (PI), czy też dynamika zmian przepływu, wyrażana przez Waveform Age (WA) – indeks opracowany przez autora.
METODA: rytm serca (HR) jest traktowany jako wskaźnik napięcia AUN. Znormalizowanej funkcji korelacji wzajemnej użyto do analizy zmian HR, WA i PI wywołanych kontrolowanym oddychaniem (okres 8-24 s, próba Valsalvy).
MATERIAŁ: Prędkość przepływu krwi w tętnicy ramiennej u 10 osób.
WYNIKI: Zmiany HR i WA są silnie skorelowane (R=-0.9), WA zmienia się wcześniej o 1.7±0.7s niż HR (dla konkretnej osoby przesunięcie w czasie jest stałe i całkowicie niezależne od sposobu oddychania). PI również koreluje z HR (R=-0.9), ale tylko podczas rytmicznego oddychania z mniejszymi częstościami i wówczas zmienia się po HR.
WNIOSEK: Nie sam fakt pulsacyjności przepływu, a dynamiczność zmian przepływu zdaje się wpływać na AUN.

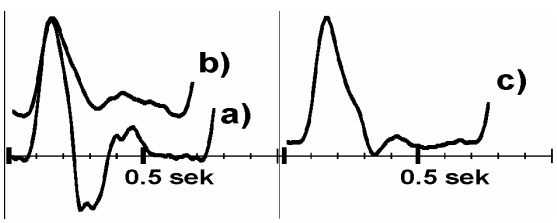
Słowa kluczowe: autonomiczny układ nerwowy, rytm serca, przepływ pulsacyjny, Pulsatility Index, Waveform Age

Abstract. GOAL: To determine which feature of the pulsatile blood flow influences autonomic nervous system (ANS): whether the pulsation degree quantified with commonly used Pulsatility Index (PI) or dynamics of flow changes quantified with Waveform Age (WA) – an index proposed previously by the author.
METHOD: Heart rate (HR) was treated as an index of ANS function. The normalized cross-correlation function was used for analysis of PI, WA, and HR changes induced by controlled breathing (with 8 – 24 sec periods, and Valsalva maneuver).
MATERIAL: 10 persons – blood flow velocity in the brachial artery.
RESULTS: Changes of WA and HR were strictly correlated (R=-0.9), and WA altered sooner than HR (1.7±0.7sec – but for a particular person, this time-shift was exactly constant for all breathing modes). PI correlated with HR (R=-0.9), but it changed later than HR (1 heart period). The dependence exists only for breathing with greater frequency.
CONCLUSION: It seems the dynamics influences ANS.

Key words: autonomic nervous system, pulsatile blood flow, heart rate, Pulsatility Index, Waveform Age

1. Wprowadzenie

To, że odbicia fal przepływu w tętnicach mają kluczowe znaczenie dla życia organizmu, jest przez autora postulowane od 20 lat [2]. Odbicia te są tym większe, im opór obwodowy jest większy od tzw. oporu sprężystego tętnic (tj. impedancji falowej) [11]. I tak, w przypadku osób młodych duża elastyczność tętnic sprawia, iż opór sprężysty jest mały, a odbicia fal duże. Daje to w efekcie dużą dynamikę zmian przepływu krwi w czasie danego cyklu pracy serca [3], [4], co ilustruje przykład na rys. 1a. Gdy opór obwodowy maleje (rys. 1b) lub sztywność tętnic rośnie (rys. 1c), odbicia maleją [4], [11], maleje więc też dynamiczność przepływu.



Rys. 1. Fale prędkości przepływu krwi w tętnicy ramiennej o różnej dynamice zmian prędkości i różnej wartości Pulsatility Index (PI): a) duże PI i duża dynamiczność (elastyczne naczynia, normalny opór obwodowy), b) małe PI i mała dynamiczność (ta sama osoba – spadek oporu po ćwiczeniu), c) duże PI ale mała dynamiczność (osoba starsza – spadek elastyczności tętnic związany ze starzeniem się)

Wyniki wielu uprzednich badań autora – z których jedne bardziej bezpośrednio, inne raczej pośrednio (poprzez analizę matematyczną i cybernetyczną) – sugerowały, że dynamiczność ma wpływ na autonomiczny układ nerwowy (AUN), a nawet na ośrodkowy (somatyczny) [6].

Wyniki innych autorów zdawały się pośrednio potwierdzać tezę wpływu na AUN. Na przykład wzrost oporów obwodowych przy niepulsacyjnym sztucznym sercu [12] oznaczałby – wg teorii autora – że: (a) AUN „rozpoznał" brak dynamiki przepływu, (b) potraktował to jako zbyt małe odbicia fal, (c) próbował zwiększyć odbicia poprzez wzrost oporów, co w normalnych warunkach spowodowałoby wzrost dynamiczności.

W ostatnich latach zaczęły ukazywać się prace (np. [1,9,13]) wykazujące już w sposób bezpośredni wpływ pulsacyjnego przepływu na AUN, a dokładniej: że niepulsacyjny przepływ zwiększa aktywność AUN. Badania te nie wskazują jednak, która cecha pulsacyjności ma znaczenie dla AUN: czy dynamika zmian przepływu – jak wynikałoby to z teorii autora – czy też jedynie sam fakt istnienia pulsującego przepływu, kwantyfikowanego za pomocą powszechnie stosowanego Pulsatility Index (PI – patrz niżej). W pracach tych nie jest nawet jednoznacznie stwierdzone, czy chodzi o pulsacyjność ciśnienia czy przepływu. Jednak inny bardzo ważny i popularny nurt badań ostatnich lat, dotyczący produkcji przez śródbłonek naczyń krwionośnych wielu ważnych substancji (tlenek azotu, endotelina, prostacyklina itd.), jednoznacznie wskazuje na rolę przepływu – a co więcej – na rolę dynamiczności przepływu (określanej często (np. w [10]) technicznym terminem *shear stress variation,* jako że w tamtych pracach chodzi bezpośrednio o tarcie płynącej krwi o śródbłonek naczyń).

Celem tej pracy jest stwierdzenie, czy dynamiczność przepływu jest tak samo ważna w przypadku wpływu na AUN, jak w przypadku wpływu na śródbłonek. Jako wskaźnik działania AUN przyjęto rytm pracy serca (HR) [7], [9].

2. Materiał i metody

2.1. Materiał i sprzęt

Prędkość przepływu krwi w tętnicach ramieniowych u 10 osób (18-40 lat) mierzono ultradźwiękowym prędkościomierzem przepływu krwi UDP-83 (Techpan) z głowicą 8MHz. Sygnał analogowy próbkowano (1 kHz) i przetwarzano za pomocą IBM PC. Badane osoby kontrolowały samodzielnie (lecz pod nadzorem) okres oddychania (8, 10, 12 sek.) wg stopera. W przypadku sprawniejszych osób zarejestrowano również przebiegi dla dłuższych okresów (14, 16, 18, 20 i 24 sek.) i w czasie próby Valsalvy.

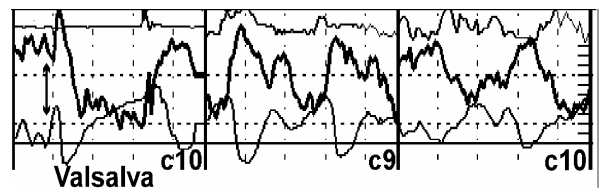
2.2. Indeksy PI i WA

PI liczono standardowo: PI=(Vmax-Vend)/Vmean (gdzie: Vmax – prędkość maksymalna, Vent – prędkość końcowo-rozkurczowa, Vmean – prędkość średnia). PI jest więc indeksem „matematycznym", tj. określa w pewien sposób to, w jakiej części przebieg jest stały, a w jakiej pulsacyjny, bez względu na to, czy przebieg ten dotyczy prędkości przepływu krwi, czy też dotyczyłby czegokolwiek innego. Natomiast przy definiowaniu indeksu Waveform Age (WA), użytego do ilościowego opisania dynamiczności, wykorzystano porównanie z rzeczywistymi, wzorcowymi przebiegami prędkości krwi o różnej dynamice zmian. WA jest więc indeksem „fizjologicznym", bowiem określa dynamikę zmian prędkości przepływu krwi, ale z uwzględnieniem rzeczywistego, fizjologicznego sposobu tych zmian.

Jako że dynamika zmian prędkości w tętnicach osób młodych jest generalnie duża, a osób starych – mała [3], za ww. wzorcowe przebiegi przyjęto [4]: duża dynamika – uśrednione przebiegi otrzymane dla młodych (Ym), mała dynamika – uśrednione przebiegi dla osób starych (Em). Dynamikę konkretnego przebiegu określono, podając w jakim stopniu jest on podobny do Ym, a w jakim do Em. Wykorzystano w tym celu aparat matematyczny przestrzeni funkcyjnej, gdzie funkcja V(t), tj. przebieg prędkości w funkcji czasu (a więc też Ym(t) oraz Em(t)), traktowana jest jako wektor tej przestrzeni. W celu ułatwienia przeprowadzania obliczeń parę wektorów wzorcowych Ym i Em orto-normalizowano, co dało prostokątny układ współrzędnych o osiach Y i E. Wyjściową miarą dynamiczności jest kąt, jaki tworzy wektor V z osią poziomą Y. I tak np. kąt równy 45° oznacza, że V leży na przekątnej układu współrzędnych, tj. że jest jednakowo oddalony od obu wzorcowych przebiegów Y i E, czyli że dynamiczność jest dokładnie średnia. Appendix zawiera więcej szczegółów na temat metody.

2.3. Sygnały HR, WA i PI

Sygnały HR, WA i PI utworzono z zarejestrowanego przebiegu prędkości przepływu w następujący sposób: (a) za początek danego cyklu pracy serca przyjęto moment, w którym pochodna prędkości przyjmuje wartość maksymalną, (b) za „wartość" HR dla tego cyklu przyjęto liczbę 60/długość_cyklu, (c) wartości HR, WA i PI dla danego cyklu traktuje się jako wartości sygnałów w chwili początku cyklu, (d) interpolacja liniowa daje wartości sygnałów dla pozostałych chwil.



Rys. 2. Przykładowe sygnały WA (linia pogrubiona), HR (cienka, dolna) i PI (cienka, górna) w trakcie próby Valsalvy (po lewej) i podczas oddychania o okresie 24 sek. (osoby C9 i C10). Pionowe linie przerywane odpowiadają 10 sek. Strzałka wskazuje początek wdechu inicjującego próbę Valsalvy

2.4. Analiza sygnałów

Związki między parami sygnałów analizowano za pomocą znormalizowanej funkcji korelacji wzajemnej, liczonej (z uwzględnieniem tego, że komputer analizuje sygnały próbkowane) wg wzoru:

$$\mathrm{FKW}(\tau) = \frac{\sum (x(t_i - \tau) - \bar{x}) \cdot (y(t_i) - \bar{y})}{\sqrt{\sum (x(t_i - \tau) - \bar{x})^2} \cdot \sqrt{\sum (y(t_i) - \bar{y})^2}}$$

gdzie: FKW(τ) – wartość funkcji dla przesunięcia w czasie między sygnałami równego τ, x(t) – sygnał, który na czas

obliczeń został wstępnie uznany za sygnał niezależny, y(t) – sygnał zależny (dalej – sygnał przyjęty czasowo za niezależny będzie prezentowany w nawiasie po sygnale zależnym, tj. y(x)).

Zauważmy, że FKW(0) jest równe zwykłemu współczynnikowi korelacji między parami danych $[x(t_i),y(t_i)]$. Jednak poprawną miarą związku między sygnałami nie jest współczynnik korelacji. Poprawną miarą jest wartość FKW(τ) dla takiego τ, dla którego FKW ma ekstremum, ponieważ jeśli nawet istniałby ścisły, przyczynowy związek między dwoma sygnałami, nie muszą one zmieniać się synchroniczne. Wprost przeciwnie: raczej bardziej prawdopodobnym jest, iż nieco czasu potrzeba, by jedna wielkość wpłynęła na wartość drugiej.

Należy podkreślić, iż znacząca wartość FKW dla pewnego τ nie musi oznaczać, że między sygnałami istnieje związek przyczynowy, czy to bezpośredni (jeden sygnał zależy od drugiego), czy pośredni. W szczególności, bez względu na to, czy między sygnałami istnieje jakikolwiek związek, czy nie, jeśli są one okresowe, to FKW ma zawsze wiele ekstremów (maksima i minima w odstępie co pół okresu). Jeżeli jednak położenie jednego z ekstremów nie zależy od okresu, to raczej nie okresowość jest jego przyczyną. Tym bardziej, jeżeli położenie takiego ekstremum jest podobne dla różnych osób.

Generalnie, by uniknąć problemów z istnieniem ekstremów FKW wynikających jedynie z okresowości, a nie z ew. rzeczywistego związku między sygnałami, bada się sygnały o charakterze szumu białego lub delty Dirac'a. W tej pracy ważne jednak też było, czy związki między HR, WA i PI zależą od częstości/głębokości oddychania. Wszystko powyższe jest powodem tego, że w niniejszej pracy: (a) analizowano sygnały okresowe, (b) ale o co najmniej trzech różnych okresach, (c) a następnie badano istnienie ekstremum o położeniu niezależnym od tych okresów.

3. Wyniki

Oddychanie powoduje znaczące zmiany HR i WA, a sygnał WA wygląda jak przesunięte w czasie lustrzane odbicie sygnału HR (rys. 2). To oznacza wysoką ujemną korelację i niesynchroniczność zmian. To, że sygnał HR jest bardziej „gładki" niż sygnał WA, wynika z faktu istnienia odkrytej przez autora [5] ultra-wysoko-częstotliwościowej składowej zmienności dynamiki przepływu, która nie występuje w zmienności rytmu serca (HRV) – pozostałe komponenty występują i w HRV i w zmienności dynamiki.

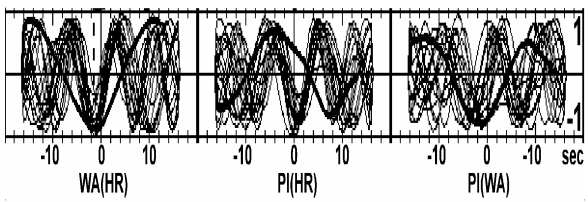
Podobieństwo zmian w sygnałach WA i HR oraz przesunięcie w czasie między nimi potwierdza analiza FKW. Funkcja ta ma dla WA(HR) szereg znaczących ekstremów (rys. 3 i 4). Jednak tylko ekstremum leżące najbliżej przesunięcia τ=0 ma znaczenie, bowiem położenie pozostałych ewidentnie zależy od okresu. Świadczy to, że są one jedynie rezultatem okresowości sygnałów (rys. 4c: gdy nie ma okresowości, nie ma pozostałych ekstremów). Dane dotyczą oddychania z okresem 8, 10, 12 sek. (dla większych okresów tylko dane dla WA(HR) nie ulegają żadnej zmianie – rys. 3, 4)

Tak więc, jak wynika z tabeli oraz rys.3 i 4, WA i HR są ze sobą silnie skorelowane dla przesunięcia, które jest zadziwiająco stałe dla różnych częstotliwości oddychania (rys. 4d) i w przybliżeniu jednakowe dla różnych osób (rys. 3). Przesunięcie to pozostaje niezmiennie takie samo dla próby Valsalvy (rys. 4c), która znacząco różni się od zwykłego oddychania, a nawet dla dodatkowej oscylacji HR i WA, pojawiającej się przy powolnym oddychaniu (rys. 2), tj. gdy częstotliwość podstawowa oddychania została wyeliminowana za pomocą górno-przepustowego filtru ruchomej średniej (rys. 4e). Co więcej, takie same wyniki otrzymano dla tętnicy szyjnej (rys. 4c) przy pojedynczych oddechach, tj. podczas nieperiodycznego oddychania z rys. 5.

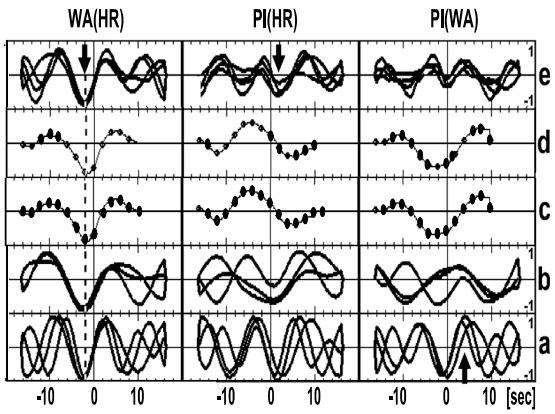
Tab. 1. Przesunięcia, dla których znormalizowana funkcja korelacji wzajemnej przybiera ekstremum

	ekstremum	przesunięcie [sek.]
WA(HR)	-0.89±0.10	-1.7±0.7
PI(HR)	-0.91±0.06	+0.9±0.7
PI(WA)	+0.83±0.14	+2.8±1.3

Jednak najważniejszym wynikiem jest fakt, iż przesunięcie dla pary WA(HR) jest ujemne, czyli WA zmienia się wcześniej niż HR.



Rys. 3. Znormalizowane funkcje korelacji wzajemnej dla poszczególnych par sygnałów HR, WA i PI (w nawiasach sygnały „niezależne") dla zbadanych osób, dla okresów oddychania 8, 10, 12 sek. (grube krzywe – przykładowe funkcje dla okresu 24 s). Oś pozioma – argument funkcji, tj. przesunięcie między sygnałami, oś pionowa – wartość funkcji



Rys. 4. Funkcje korelacji, jak na rys. 3, dla jednej osoby oddychającej z okresem: (a) 8, 10, 12 sek., (b) 14, 16, 18 sek., (c) 20, 24 sek.i próba Valsalvy, linia kropkowana – przepływ w tętnicy szyjnej w czasie aperiodycznego oddychania z rys. 5 (brak periodyczności→tylko jedno ekstremum dla WA(HR)), (d) powyższe zaprezentowane razem, (e) dla okresów 16, 18, 20 i 24 sek. po odfiltrowaniu z sygnałów częstotliwości oddychania, tj. dla „drgań własnych" układu krążenia

Tak niezmiennej relacji nie obserwuje się dla PI i HR. W przypadku szybszego oddychania (okres do 14 s) FKW ma znaczące minimum (tab. 1) dla stałego w przybliżeniu, dodatniego przesunięcia $\tau\approx$0.9 sek., tj. PI zmienia się o ok.

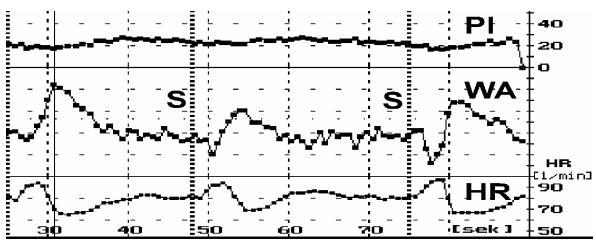
jeden cykl pracy serca później niż HR. A więc, jeśli ewentualnie istnieje związek przyczynowy między HR i PI, to HR wpływa na PI, a nie na odwrót. Taka jednorodność FKW znika dla wolniejszego oddychania (rys. 4 i pogrubiona krzywa na rys. 3). Sugeruje to, że jeżeli zmienność HR jest źródłem jakichś zmian PI, to tylko jednym z wielu. Pozostałe zaczynają dominować w miarę wzrostu głębokości oddechu.

4. Dyskusja

Zależność PI od HR można by łatwo wytłumaczyć prawem Franka-Starling'a: mniejsze HR (dłuższy okres) ⇒ pełniejsze wypełnienie lewej komory ⇒ silniejszy skurcz w następnym cyklu ⇒ większa amplituda przepływu (prędkości) w następnym cyklu ⇒ większe PI w następnym cyklu, tj. po czasie ok. 0.9 sek. Stąd ujemna korelacja oraz takie właśnie przesuniecie w czasie. Oddychanie z bardzo małą częstością wymaga bardzo głębokich oddechów, co modyfikuje powrót żylny i – niezależnie od HR – wypełnienie lewej komory. Stąd zanikanie korelacji między HR i PI dla oddychania powolnego i głębokiego.

Związek między WA i PI wydaje się niezbyt silny (rys. 3 i 4). Jest to o tyle ciekawe, że przecież oba wskaźniki charakteryzują to samo zjawisko: zmienność przepływu krwi. Widocznie – uwzględniając też różny związek z HR - ta cecha pulsacyjności, którą określa PI, jest nie tylko z matematycznego, ale też z fizjologicznego punktu widzenia całkiem czymś innym niż cecha, którą nazwano dynamicznością, a która jest kwantyfikowana przez WA.

To, że WA i PI w ogóle korelują ze sobą w przypadku mniejszych częstości oddychania, wydaje się być związane jedynie z pośrednictwem HR, a nie z tym, że oba wskaźniki opisują ten sam przebieg prędkości przepływu. Rzeczywiście, gdy HR ma dominujący wpływ na PI (płytsze oddychanie), suma przesunięć dla par PI(HR) i HR(WA) jest równa przesunięciu dla PI(WA) zarówno w przypadku wartości średnich (tab. 1), jak i wartości dla pojedynczych osób. Wygląda to tak, jak gdyby HR było czynnikiem pośredniczącym między zmianami WA i PI, tj. WA→HR→PI. I dlatego związek między WA i PI zanika całkiem dla głębszych oddechów (rys. 4), ponieważ zanika człon HR→PI.



Rys. 5. Widok ekranu komputerowego systemu TOMAg podczas dopplerowskiego badania przepływów w trybie *on-line*: tętnica szyjna – WA i HR podczas trzech głębokich oddechów (grube przerywane pionowe proste – początki wdechów) oddzielonych długim bezdechem, co może być traktowane jako trzy osobne quasi-delty Dirac'a. PI w tętnicy szyjnej jest prawie niezmienne, co może być efektem sprawnej autoregulacji przepływu mózgowego, a co było powodem badania przepływu w tętnicy ramiennej

Prezentowane wyniki sugerują więc, że jeśli już któraś z cech pulsacyjnego przepływu krwi ma wpływać na AUN, to nie jest to pulsacyjność sama w sobie, wyrażana za pomocą PI, a raczej dynamika zmian przepływu. Jest to, oczywiście, tylko sugestia, bowiem jednoznaczne rezultaty przeprowadzonej tu analizy matematycznej dowodzą tylko i jedynie, że: (a) sygnały WA i HR są niezwykle ściśle ze sobą skorelowane, (b) jeśli ta ścisła korelacja jest wynikiem związku przyczynowego między tymi sygnałami, to na pewno nie zmiany HR wpływają na zmiany dynamiki. Natomiast analiza matematyczna nie może odpowiedzieć na pytanie: czy zmiany dynamiki przepływu wpływają na HR, czy też zarówno WA, jak i HR zależą (niezależnie od siebie) od innego czynnika. Na takie pytanie odpowiedzieć może tylko fizjologia.

Jako że bezpośredni wpływ dynamiczności na HR nie był zbyt zgodny z wiedzą fizjologiczną w momencie przeprowadzenia prezentowanych tu badań, najpierw zanalizowana zostanie ewentualność istnienia czynnika, który niezależnie wpływa na dynamiczność i HR. Zakładając, że za zaobserwowane zmiany dynamiczności odpowiedzialne są tylko zmiany oporów obwodowych (jak na rys. 1a-b), trzeba wziąć pod uwagę co najmniej cztery następujące, powszechnie uznane mechanizmy fizjologiczne [8,14] (wzrost/spadek oporów obwodowych wyrażano poniżej od razu – dla większej przejrzystości – spadkiem/wzrostem WA):

Neurony sieci oddechowej: Ich bezpośredni wpływ na aktywność współczulną i przywspółczulną powoduje, iż w czasie wdechu wzrasta HR i maleje WA (w czasie wydechu na odwrót).

Baroreceptory tętnicze: Spowodowane oddychaniem fluktuacje ciśnienia tętniczego (tzw. fala oddechowa ciśnienia) winny oddziaływać na HR i WA (wzrost ciśnienia → spadek HR i wzrost WA).

Receptory sercowo-płucne: Ich znaczenie powinno szczególnie wzrastać podczas głębokiego oddychania, które znacznie moduluje wypełnienie naczyń płucnych i przedsionków.

Lokalna autoregulacja przepływu: WA powinno reagować na oddechową falę ciśnienia: wzrost ciśnienia →spadek WA i na odwrót (przeciwnie niż odruch z baroreceptorów!).

Generalnie (z wyjątkiem lokalnej autoregulacji, która bezpośrednio nie dotyczy HR), ww. mechanizmy zmieniają HR przeciwnie niż WA, co mogłoby tłumaczyć ujemną korelację między HR i WA. Jednakże, skoro zależność między zmianami HR i WA jest taka niezmienna, działanie tych różnych mechanizmów (jak i innych, tu nie wymienionych) powinno być wprost niezwykle precyzyjnie zharmonizowane i to w sposób niezależny od głębokości oddychania. Wydaje się to mało prawdopodobne. Ponadto:

- Ww. mechanizmy nie tłumaczą, dlaczego zmiany WA zachodzą wcześniej niż zmiany HR. Na przykład składowa naczyniowa odruchu z baroreceptorów pojawia się kilka sekund później, a nie wcześniej niż składowa sercowa.
- Pierwsze trzy mechanizmy związane są z aktywnością układu autonomicznego, a więc mogą wpływać na WA w tętnicy ramiennej, ale nie w szyjnej (rys. 5), gdyż krążenie mózgowe kontrolowane jest niemal wyłącznie przez czwarty mechanizm [8, 14]. A przecież zmiany WA w obu tętnicach są jednakowo związane z HR.
- Jeżeli to lokalna autoregulacja byłaby głównym mechanizmem powodującym obserwowane zmiany WA, to skąd taki ścisły związek ze zmianami HR, jeśli nie poprzez AUN?

Tak więc jest dość trudno wytłumaczyć tak ścisły związek między dynamicznością przepływu i HR za pomocą wpływu trzeciego czynnika. Być może więc to dynamiczność jest rzeczywiście tą cechą pulsacyjnego przepływu krwi, która wpływa na AUN. Należy jednak przypomnieć, iż autorzy, którzy w bezpośrednich badaniach wykazali wpływ pulsacji na AUN, nie określili jednoznacznie, czy czynnikiem jest zmienność ciśnienia podczas pulsacyjnego przepływu, czy sam przepływ. Wydawałoby się, że łatwiej byłoby wytłumaczyć wpływ ciśnienia poprzez wpływ na np. baroreceptory szyjne lub mechanoreceptory serca. W [15] podano jednak – na podstawie wspomnianej teorii autora oraz badań eksperymentalnych, dotyczących aktywności śródbłonka – pewne sugestie na temat tego, jak organizm mógłby „odczytywać" dynamiczność przepływu bezpośrednio (a nie dynamiczność ciśnienia związaną z dynamicznością przepływu).

Całkiem odrębnym pytaniem jest, dlaczego WA w ogóle zmienia się z oddechem, jeśli nie jest to wpływ HR. Istnienie w sygnale WA rytmu oddechowego spowodowane może być kilkoma czynnikami. Stosunkowa długotrwałość zmian WA podczas tzw. *próby ziewania* (stosowanej standardowo przy badaniu za pomocą systemu TOMAg) sugeruje, iż periodyczne zmiany utlenowania krwi mogą być jedną z przyczyn rytmu. Jednak, mimo że zmiany te wydają się być najbardziej oczywistą przyczyną zmian WA, WA zmienia się podobnie również w przypadku „oddychania bez oddychania" tj. podczas ruchów klatki piersiowej (i przepony), takich jak przy oddychaniu, ale z zamkniętymi ustami i nosem. Ponadto oddech, zmniejszając zawartość CO_2 we krwi, powinien zwiększyć opory w krążeniu mózgowym, tj. zmniejszyć WA, a dzieje się na odwrót: po niewielkim początkowym spadku następuje znaczny i długotrwały wzrost WA (rys. 5).

Fakty te sugerują, iż drugim, chyba istotniejszym czynnikiem mogą być zmiany objętości końcowo-rozkurczowej serca (zmiany powrotu żylnego, oporu w krążeniu płucnym, itd. spowodowane zmianami ciśnienia w klatce piersiowej), które zmieniają objętość wyrzutową. Być może wspomniana wcześniej „częstotliwość własna" układu krążenia (rys. 2, 4e), to czas powrotu fluktuacji rzutu serca, spowodowanej inicjacją wdechu, po przejściu przez cały układ krwionośny. Należy jednak przypomnieć, że to zmiany PI, które nie wydają się być skorelowane z WA, są bezpośrednio związane ze zmianami objętości wyrzutowej. Tak więc zmiany WA odzwierciedlają raczej reakcję układu krążenia na zmiany rzutu niż tylko samą zmianę rzutu.

Trzecim czynnikiem mógłby być bezpośredni wpływ nerwu błędnego, jak to się postuluje w przypadku HR. Jednak zmienność WA również w tętnicy szyjnej zdaje się wykluczać tę możliwość.

Jakie by nie były przyczyny zmian WA, to HR je naśladuje. Jest to dość zaskakujący wynik, zważywszy na to, że obecnie powszechnie uważa się, iż oddechowa zmienność HR jest bezpośrednim rezultatem powiązań na poziomie układu autonomicznego między ośrodkami: oddechowym i sercowym. Zmienność oddechowa HR ma spore znaczenie diagnostyczne: spadek zmienności $\rightarrow$ wzrost ryzyka chorób serca, choć nie jest jasne, dlaczego. Gdyby otrzymany wynik dotyczący zależności między WA i HR odzwierciedlał rzeczywisty związek przyczynowy, wyjaśnienie wzrostu ryzyka chorób serca byłoby proste: zmniejszona oddechowa zmienność HR świadczy o utracie zdolności dostosowywania pracy serca do warunków pracy układu naczyniowego.

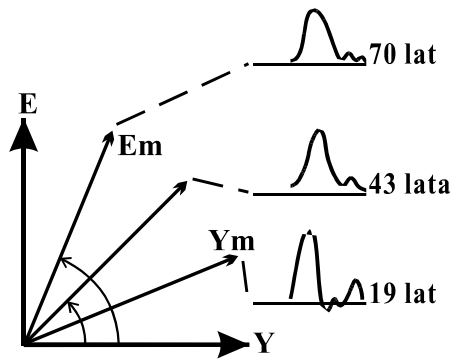
5. Wnioski

Cecha pulsacyjnego przepływu krwi charakteryzowana przez PI jest, nie tylko z matematycznego, ale też z fizjologicznego punktu widzenia całkiem czymś innym niż cecha kwantyfikowana przez WA. Jeśli inni autorzy mieliby w sposób bardziej zaawansowany niż dziś badać wpływ pulsacyjności przepływu krwi na układ nerwowy, powinni uwzględnić istnienie tej odmienności.

Zaprezentowane wyniki sugerują, że to właśnie dynamiczność wyrażana za pomocą WA jest cechą, która wpływa na AUN. Bezpośrednia, eksperymentalna weryfikacja tej tezy należy do fizjologów. Niniejsza praca sugeruje kierunek badań.

Appendix

Przestrzeń funkcyjna jest konstrukcją matematyczną, w której funkcje V(t), np. przepływ krwi w funkcji czasu, traktowane są jak (zwykłe) wektory V. Różna dynamika zmian, tj. różny kształt przebiegu prędkości krwi (KPPK) odpowiada różnemu położeniu wektora V, reprezentującego KPPK (rys. 6). Można więc określić ilościowo KPPK, określając położenie wektora, np. w postaci kąta, jaki tworzy on z pewnym wektorem odniesienia (Y na rys. 6).



Rys. 6. Przykładowe krzywe prędkości przepływu krwi i odpowiadające im wektory. Wektor Ym odpowiada przepływowi o dużej dynamiczności, Em – o małej. Z wektorów Ym i Em (w rzeczywistości Ym i Em to wektory średnie dla osób młodych i starych) utworzono układ współrzędnych o osiach Y i M. Miarą dynamiczności konkretnego przepływu jest kąt wektora z osią Y

Wiele czynników ma wpływ na KPPK w danej tętnicy. Jednak potraktowanie KPPK jako wektora V pozwoliło wykazać, że zmiany dynamiczności polegają prawie wyłącznie na obrocie wektora V w kierunku Y lub E bez względu na rodzaj czynnika, który spowodował te zmiany. Można więc miarę kątową zastosować do analizy zmian dynamiczności o dowolnej genezie, np. jak w niniejszej pracy.

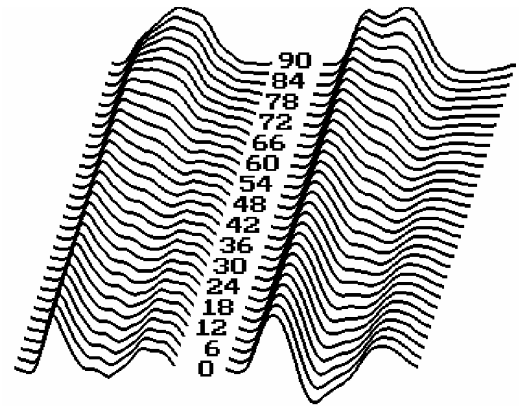
Dynamiczność i jej zmienność jest różna dla różnych tętnic. Miarę kątową dla każdej z tętnic można jednak przekształcić na wskaźnik uniwersalny, tj. jednakowy dla wszystkich tętnic, wykorzystując fakt, iż wpływ na dynamiczność u osób zdrowych w stanie spoczynku ma głównie wiek. Tym uniwersalnym wskaźnikiem jest WA liczone ze wzoru: WA=(kąt-I)/S , gdzie I, S – współczynniki regresji liniowej kąta dla danej tętnicy względem wieku. W przybliżeniu dla każdej tętnicy średnia wartość WA u osób w wieku x równa jest x (stąd nazwa WA: *Waveform Age*). Tak wyskalowany wskaźnik

pozwala między innymi na bezpośrednie stwierdzenie: (1) czy WA dla wszystkich tętnic odpowiada wiekowi metrykalnemu pacjenta (np. ogólny stan pacjenta), (2) czy WA jest jednakowe dla wszystkich tętnic, czy też dla jednej z tętnic WA ma inną wartość niż dla pozostałych (np. WA znacząco większe niż w pozostałych tętnicach sugeruje miażdżycę proksymalną).

Ostatecznie więc WA jest przeskalowanym kątem między Y i V (a dokładniej: przeskalowanym kątem między Y i rzutem V na płaszczyznę wyznaczoną przez Y i E). WA liczono ze wzoru:

$$WA = \alpha + \beta \cdot arctg\left(\left[E * (V - \overline{V})\right] / \left[Y * (V - \overline{V})\right]\right),$$

gdzie: $\overline{V}$ – wartość średnia V(t) w danym cyklu (odjęcie $\overline{V}$ powoduje, iż WA zależy jedynie od dynamiki zmian V, tj. od KPPK, natomiast nie zależy od położenia krzywej nad osią czasu); * – iloczyn skalarny w przestrzeni funkcyjnej. Współczynniki skali α i β, mające znaczenie w innych zastosowaniach (np. [4, 6]), tu są nieistotne, gdyż w tej pracy analizowane były jedynie znormalizowane odchylenia WA od wartości średniej (więc: α uległo odjęciu, β – skróceniu).

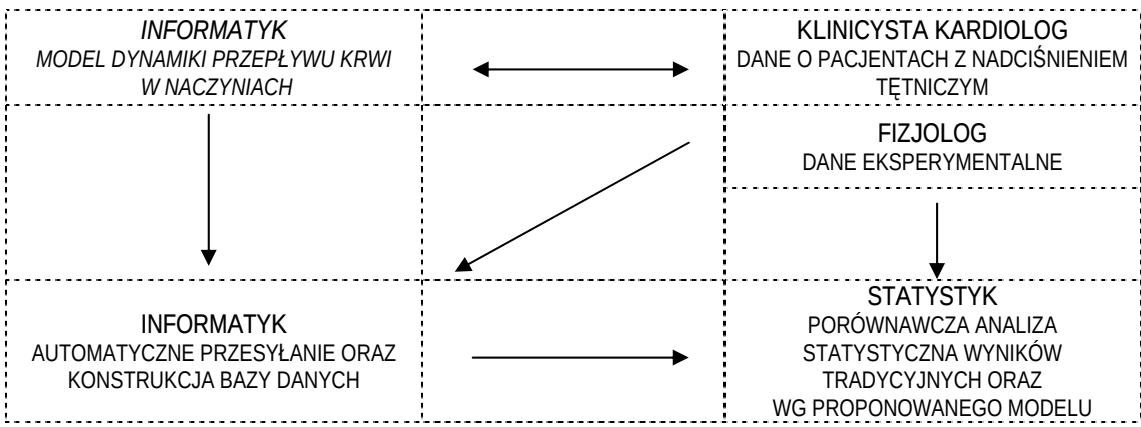


Rys. 7. Różne dynamiczności przepływu krwi w tętnicy szyjnej (po prawej) i ramieniowej (po lewej) oraz odpowiadające im wartości kąta określającego ilościowo te dynamiczności. Kształty (osobno dla tętnic szyjnych i ramieniowych) dla 0 i 90 stopni tworzą kartezjańske układy współrzędnych o osiach Y i E z rys. 6

Przedstawiona tu w wielkim skrócie metoda ilościowego opisu kształtu krzywej została szczegółowo zaprezentowana w [4].

Literatura cytowana

1. Fukae K., Tominaga R., Tokunaga S., et al.: The effects of pulsatile and nonpulsatile systemic perfusion on renal sympathetic nerve activity in anesthetized dogs, J. Thorac. Cardiovasc. Surg., 111(2): 478-84, 1996.
2. Gólczewski T.: Physiological death of old person without any disease, VI Prerovske Geriatricke Dny, Prerov 1986.
3. Gólczewski T., Krajewski A.: A vector based approach to age-related changes of Doppler velocity waveforms in the brachial artery, Ultrasound Med. Biol., 13: 15-18, 1987.
4. Gólczewski T.: Metoda prezentacji wybranych wielkości hemodynamicznych jako wektorów w przestrzeni $L^2(T)$ dla różnicowania stanu biologicznego – Analiza prędkości przepływu krwi w tętnicach dla oceny zaawansowania procesu starzenia się, Prace IBIB PAN, 38, 1994.
5. Gólczewski T.: Analogiczna do HR Variability, krótko-okresowa zmienność kształtu krzywej prędkości przepływu krwi, X Konferencja Naukowa „Biocybernetyka i Inżynieria Biomedyczna", 1997.
6. Gólczewski T.: Czy kształt fali przepływu krwi w tętnicach oddziałuje na centralny układ nerwowy?, XI Konferencja Naukowa „Biocybernetyka i Inżynieria Biomedyczna", t1:26-30, 1999.
7. Hayano J., Sakakibara Y., Yamada A., et al.: Accuracy of assessment of cardiac vagal tone by heart rate variability in normal subjects, Am. J. Cardiol., 67: 199–204, 1991.
8. Heller L. J., Mohrman D. E.: Cardiovascular physiology, 1981.
9. Nishinaka T., Tatsumi E., Nishimura T., et al.: Cardiac autonomic nervous function during long-term nonpulsatile left heart bypass, Artif. Organs., 23(6): 500-3, 1999.
10. Ohonara T, Okadome K, Yamamura S, Mii S, Sugimachi K. Simulated blood flow and the effects on prostacyclin production in the dog femoral artery. Circ Res 68:1095-1099, 1991
11. O'Rourke M. F.: Arterial function in health and disease, Churchill Livingstone, Edinburgh 1982.
12. Runge T. M., Grover F. L., Cohen D. J., Bohls F. O., Ottmers S. E., Saadatmanesh V.: Comparison of a steady flow pump to a preload responsive pulsatile pump in left atrial-to-aorta bypass in canines, Artif. Organs., 15: 35-41, 1991.
13. Toda K., Tatsumi E., Taenaka Y., et al.: How does the sympathetic nervous system behave during non pulsatile circulation? ASAIO J., 41(3): M465-8, 1995.
14. Traczyk W., Trzebski A. (red): Fizjologia człowieka z elementami fizjologii klinicznej, 1980.
15. Gólczewski T.: Nowe spojrzenie na genezę pierwotnego nadciśnienia tętniczego – hipoteza cybernetyczna, VI Ogólnopolska Konferencja Naukowa: Modelowanie Cybernetyczne Systemów Biologicznych, Kraków 2005.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 87-88.

ANALIZA AKTYWNOŚCI FALI WOLNEJ W PRZEWODZIE POKARMOWYM SZCZURA Z ZASTOSOWANIEM METODY *FFT*

ANALYSIS OF THE SLOW WAVE ACTIVITY IN THE ALIMENTARY TRACT OF RATS WITH APPLICATION OF THE *FFT* METHOD

ALEKSANDRA JUNG[*], GRZEGORZ KRÓLCZYK[**], PIOTR THOR[**]

[*]*Wydział Fizyki i Informatyki Stosowanej Akademii Górniczo-Hutniczej, Kraków, jung@novell.ftj.agh.edu.pl*
[**] *Katedra Patofizjologii Uniwersytetu Jagiellońskiego, Collegium Medicum, Kraków*

Streszczenie. Celem pracy było zastosowanie transformaty Fouriera do znalezienia częstotliwości fali wolnej generowanej w przewodzie pokarmowym szczurów w normalnych warunkach fizjologicznych, jak i po zadziałaniu odpowiedniego bodźca, co może okazać się pomocne w opracowaniu nieinwazyjnych metod diagnozowania wybranych schorzeń przewodu pokarmowego.
Sygnał elektryczny był zbierany poprzez elektrody umocowane na trzonie żołądka i końcowym odcinku dwunastnicy, następnie wygładzany i poddawany analizie. Do analizy sygnału zastosowano szybką transformatę Fouriera (FFT), która pozwoliła na znalezienie częstotliwości charakterystycznych. Częstotliwość fali wolnej była większa dla dwunastnicy niż dla żołądka, a podanie wazopresyny powodowało zwiększenie częstotliwości fali wolnej w żołądku.

Słowa kluczowe: fala wolna, transformata FFT, EMG

Abstract. The goal of this study was to assess the frequency of the basic electric rhythm generated in the alimentary tract of rats, both in the physiological conditions and when pharmacological stimulus was affecting. The obtained results should be helpful to work out a new non-invasive diagnostic method for selected gastric diseases.
The electrical activity was recorded using silver electrodes implanted on the stomach and the duodenum, and connected to amplifiers. The smoothed signal of electrical activity was analyzed by means of fast Fourier transform. The characteristic frequencies were determined showing that the basic electric rhythm was much higher for the duodenum than for the stomach. Moreover, the application of vasopressin caused significant increase of the basic electric rhythm in the stomach.

Key words: basic electric rhythm, Fourier transform, EMG

1. Wstęp

Mechanizmem leżącym u podstaw skurczów mięśni przewodu pokarmowego są wahania potencjału błonowego, zwane falami wolnymi lub podstawowym rytmem elektrycznym (*basic electrical rhythm* – BER) oraz potencjały czynnościowe (*spike potentials*). Ich źródłem są komórki mięśniowe o zmiennym potencjale błonowym, zwane komórkami rozrusznikowymi (*interstitial cells of Cajal*). Fale wolne przesuwają się z komórki do komórki warstwy podłużnej, indukując prądy elektrotoniczne w wewnętrznej warstwie okrężnej. Fala wolna nie wywołuje skurczów bezpośrednio, ale wyznacza częstość potencjałów czynnościowych. Od amplitudy i liczby tych potencjałów czynnościowych zależy siła skurczu. Istnieje cały szereg czynników wywołujących potencjały czynnościowe, takich jak mechaniczne rozciąganie, neurotransmitery, hormony żołądkowo-jelitowe oraz czynniki humoralne. Zarejestrowana zmiana częstotliwości skurczów może stanowić informację o zaburzeniach motoryki przewodu pokarmowego i stanowić podstawę diagnostyki nieinwazyjnej [1], [2], [3].

Zastosowanie transformaty Fouriera do analizy widma zapisu elektromiograficznego (EMG) z przewodu pokarmowego umożliwia wykonanie równoległego zapisu zmian aktywności elektrycznej fali wolnej dla żołądka, dwunastnicy i dalszych części jelita [4]. Jest to możliwe dzięki jednoczesnej ocenie zmian częstotliwości charakterystycznych fali wolnej w poszczególnych fragmentach przewodu pokarmowego szczura zarówno w warunkach fizjologicznych, jak i patologicznych.

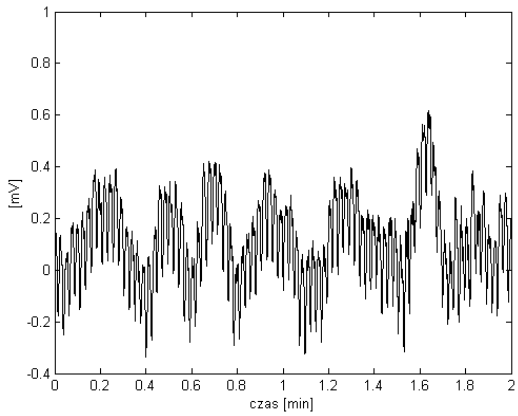
2. Materiał i metody

Metody doświadczalne

Badania były prowadzone na przytomnych szczurach, a sygnał elektryczny był zbierany przez elektrody (srebrny drut pokryty teflonem o średnicy 75 μm, A-M Systems) założone na trzon żołądka i końcowy odcinek dwunastnicy i wyprowadzone podskórnie na kark zwierzęcia. Rejestracja odbywała się poprzez system wzmacniaczy Powerlab (Adinstruments) z wykorzystaniem oprogramowania Chart v. 4.02&Extensions). Częstotliwość próbkowania wynosiła 100 Hz.

Zaburzenia motoryki (gastropareza) były wywoływane podaniem analogu wazopresyny (Terlipressin-Remestyp, Ferring-Lėčiva, 0.1 mg/kg i.p.) [5].

Przykładowy zapis aktywności fali wolnej jest przedstawiony na rys. 1.



Rys. 1. Fragment zapisu aktywności fali wolnej 5 minut od momentu podania wazopresyny u przytomnego szczura

Analiza sygnałów

Każdy sygnał był rejestrowany przez co najmniej godzinę, tak aby zapewnić odpowiednią liczebność próby niezbędną do analizy fourierowskiej. Wysoka częstotliwość próbkowania zapewniała zminimalizowanie przecieku częstotliwościowego. Następnie sygnał był wygładzany przy użyciu procedury Smoothing (Chart 4.02), a także poddawany filtracji dolnoprzepustowym filtrem odcinającym częstotliwości większe od 5 Hz i analizowany z zastosowaniem szybkiej transformaty Fouriera (FFT) [6].

Aby zmniejszyć przeciek w widmie zastosowano okno Cosine-Bell, opisane wzorem (1):

$$w(n) = 0.54 + 0.46\cos\left(\frac{2\pi \cdot n}{N}\right) \qquad (1)$$

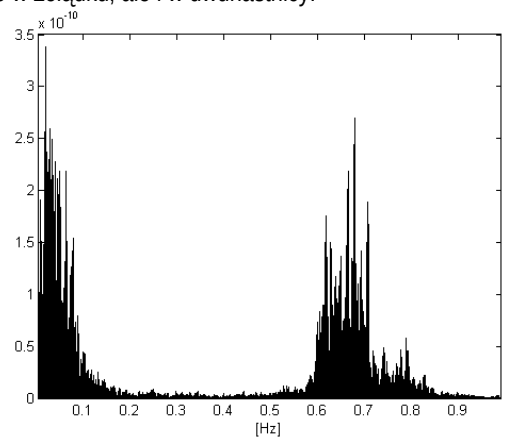
gdzie n jest wartością pomiędzy $-N/2$ a $(N/2)-1$, a N to liczba próbkowanych punktów.

3. Wyniki

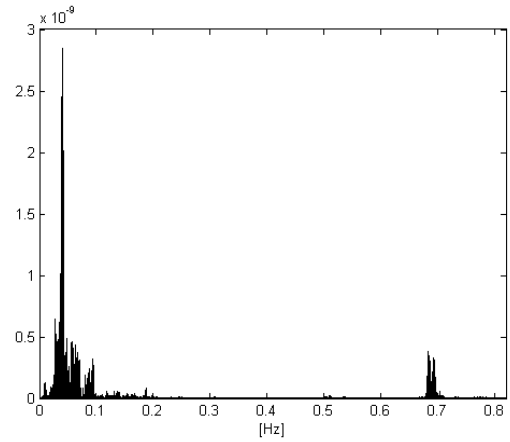
Wstępna analiza spektralna FFT przeprowadzona dla szczurów przytomnych wykazała, że częstotliwość fali wolnej dla dwunastnicy jest większa niż dla żołądka (0.073±0.010 Hz oraz 0.646±0.051 Hz, $p<0.001$) (rys. 2). Podanie wazopresyny powoduje wzrost częstotliwości fali wolnej dla żołądka (0.086±0.010 Hz) i zmniejszenie w dwunastnicy (0.571± 0.019), ale tylko w przypadku żołądka różnica między kontrolą a gastroparezą jest istotna statystycznie ($p<0.05$) (rys. 3).

4. Wnioski

Uzyskane wyniki wskazują, że zastosowanie transformaty Fouriera może być pomocne w jednoczesnej detekcji częstotliwości fali wolnej charakterystycznej dla żołądka i dwunastnicy. Można wstępnie oszacować, że badania takie mogą stanowić podstawę do opracowania nowej metody diagnostycznej. W tym celu jednak konieczne będzie przeprowadzenie większej ilości eksperymentów, co pozwoli także na statystycznie istotne potwierdzenie hipotezy, że podanie wazopresyny zwiększa częstotliwość fali wolnej nie tylko w żołądku, ale i w dwunastnicy.



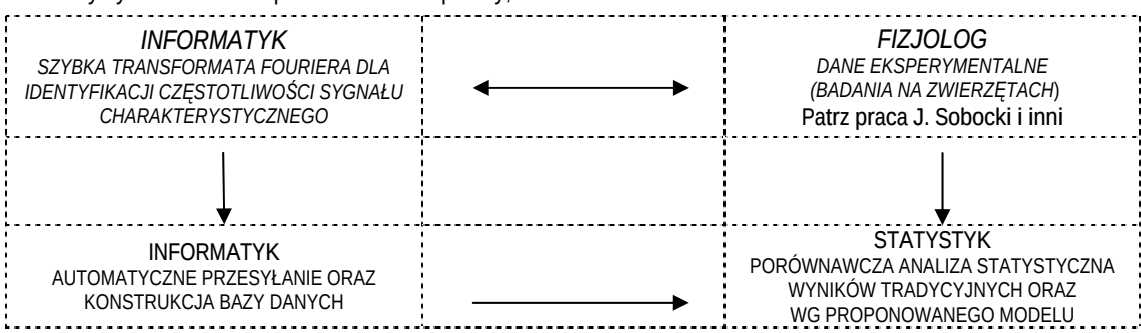
Rys. 2. Widmo częstotliwości fali wolnej dla przytomnego szczura dla zapisu aktywności elektrogastrycznej prowadzonej przez 1 godzinę



Rys. 3. Widmo częstotliwości fali wolnej dla przytomnego szczura dla godzinnego zapisu aktywności elektrogastrycznej rejestrowanej po podaniu wazopresyny

Literatura cytowana

1. Szurszewski J. H. , Elveback L. R., Code C. F.: Configuration and frequency gradient of electric slow wave over canine small bowel, Am. J. Physiol., 218(5): 1468-73, 1970.
2. Szurszewski J. H.: Electrical basis for gastrointestinal motility, In: Physiology of the Gastrointestinal Tract, Vol 2., ed. L. R. Johnson, Raven Press, N. Y., 1435-66, 1981.
3. Daniel E. E., Chapman K. M.: Electrical Activity of gastrointestinal tract as an indicator of mechanical activity, Am. J. Dig. Dis., 54: 54-102, 1963.
4. Lyons R. G.: Wprowadzenie do cyfrowego przetwarzania sygnałów, Warszawa WKŁ 2000.
5. Nowak L., Królczyk G., Sobocki J., Zurowski D., Thor P.: Gastric stimulation is effective in reversing vasopressin induced gastroparesis, Folia Med. Crac., vol. XLV, 1-2, 2004.
6. The "Spectrum" Chart extension. Application Note, ADInstruments, June 2000.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 89-98.

PATOGENNY DEFICYT MOTORYCZNY SNU I KONCEPCJA PREWENCJI ODDZIAŁYWANIAMI HARMONICZNYMI

PATHOGENIC SLEEP MOTORIC DEFICIENCY AND CONCEPT FOR PREVENTING IT BY HARMONIC EXPOSURE

MACIEJ POKORA

Instytut Biocybernetyki i Inżynierii Biomedycznej PAN, Warszawa, maciej.pokora@ibib.waw.pl

Streszczenie. Z różnicy pomiędzy warunkami regeneracji mózgu oraz pozostałych części ciała podczas nocnego wypoczynku wynika systemowy konflikt neurosomatyczny, wywołujący niekorzystny, wielogodzinny deficyt motoryczny snu (DMS). Doniesienia literaturowe potwierdzają zjawisko nasilonego występowania zgonów, a także ostrych objawów chorobowych w nocy oraz rano. Na tej podstawie oparta jest hipoteza, że powszechnie stosowane łóżka są zbyt statyczne, aby zapewniały efektywną i zarazem bezpieczną regenerację całego organizmu. Doświadczenia laboratoryjne potwierdzają pozytywny wpływ kołysania na stan żywej tkanki. Także potwierdzone klinicznie są oddziaływania łóżek kołyskowych na wentylację płuc, profilaktykę bezdechu, poprawę jakości snu, jak również polepszenie cyrkulacji płynów ustrojowych. Dla ograniczenia negatywnych skutków zjawiska DMS zaproponowano koncepcję stymulacji oscylacyjnej snu (SOS). Subtelne, harmoniczne oddziaływania inercyjne mogą przynieść w normie i w patologii korzystne efekty zdrowotne: polepszenie warunków regeneracji organizmu, unormowanie funkcji organów wewnętrznych i systemu nerwowego, a także ograniczenie bezsenności i dolegliwości bólowych.

Słowa kluczowe: sen, deficyt motoryczny, stymulacja oscylacyjna, oddziaływania harmoniczne, profilaktyka

Abstract. Different requirements for relax of brain during sleep and for the remaining body nocturnal restoration create systemic neurosomatic conflict which generates many-hour-long sleep motoric deficiency (SMD) unfavourable for health. Reports confirm the enhanced number of deaths and acute health crises during the night and in the morning. The hypothesis based on this phenomenon is that commonly used beds are far too static to secure an effective, safe restoration of the whole organism. Laboratory experience cited corroborates positive influence of rocking on condition of the biologic tissue. Also clinically confirmed are assistive features of the rocking beds for breathing function, for sleep apnea prevention, for improvement of sleep quality and for body fluids circulation. To reduce negative effects of the SMD phenomenon, the concept of sleep oscillatory stimulation (SOS) is proposed. Subtle, harmonic inertial impacts should bring significant health benefits both in normal conditions and pathology, e.g. improvement of conditions for body nocturnal restoration, conditioning of functions of inner organs and nervous system, as well as reduction of sleeplessness and pain.

Key words: sleep, motor deficit, oscillatory stimulation, harmonic impacts, prevention

1. Konflikt neurosomatyczny

Pojawienie się wieczorem symptomów fizycznego i psychicznego zmęczenia oznacza, że organizm potrzebuje snu – niezbędnego dla życia stanu z dominacją oddziaływań hamujących [100]. Sen jest niezbędny dla regeneracji mózgu, ale o ile pozostałe organy mogą w pełni wypocząć bez snu, to dla mózgu jego niedobór jest szkodliwy. Opisane w literaturze eksperymenty wykazały, że długotrwałe pozbawienie człowieka snu prowadzi kolejno do wystąpienia: rozdrażnienia, złego humoru, otępienia, agresji, trudności w utrzymaniu koncentracji uwagi, występowania iluzji, zaburzeń pamięci, wahań nastroju, kilkusekundowych epizodów snu na jawie, urojeń, napadów chorobliwej podejrzliwości, aż wreszcie – po ok. 100 godzinach deprawacji snu – prowadzi do depersonalizacji, tj. utraty tożsamości. Natomiast objawy somatyczne tak długotrwałej bezsenności są zazwyczaj nikłe: pieczenie powiek, bóle oczu, upośledzenie czucia, a także drżenie kończyn – przy jednoczesnym braku zaburzeń metabolicznych [10], [62], [57], [111].

W celu regeneracji organizmu pożądane jest zredukowanie obciążeń w stawach i obniżenie napięcia mięśni, ale najwięcej kłopotów sprawia mózg, bo jego regeneracja wymaga izolowanych warunków: przede wszystkim ciszy, a dla wydzielania melatoniny – hormonu szyszynki, regulującego m.in. fazy snu oraz czuwania – potrzebne jest także zaciemnienie. Ponieważ podczas snu następuje całkowite wyłączenie świadomości, w konsekwencji sen – konieczny stan czasowej dysfunkcji centralnego układu nerwowego – oznacza utratę możliwości kontroli równowagi, a tym samym zachowania pozycji stojącej lub siedzącej. Wymuszona w ten sposób

pozycja leżąca podczas snu umożliwia trwanie w stanie bezruchu, którego efektami są: upośledzenie cyrkulacji płynów, częściowa blokada klatki piersiowej oraz zwiększenie powierzchni podporu ciała, poddawanej ciągłemu uciskowi. O ile dla właściwej regeneracji mózgu pełne wyłączenie świadomości – oznaczające bezruch – jest warunkiem koniecznym, to przy regeneracji pozostałych elementów organizmu długotrwały bezruch jest wręcz szkodliwy. W tym powtarzającym się każdej nocy konflikcie zwycięzcą jest zawsze mózg (a właściwie jego zmęczenie), ale cenę za to szczególne zwycięstwo „umysłu" nad „ciałem" płaci często cały organizm.

Konflikt neurosomatyczny podczas snu jest konfliktem systemowym i możliwe są tu tylko kompromisy. Przystosowanie sytuacyjne wykazują zwierzęta, z których wiele naprzemiennie śpi raz prawą, a raz lewą półkulą mózgu [10], [23], [57]. Zjawisko stwierdzono – jak dotąd – u ptaków migrujących, u tygrysa, a także u zwierząt żyjących w wodzie, np. u delfina butelkonosego, żyjącego w Morzu Czarnym 200-kilogramowego ssaka, u którego epizod „półsnu" trwa do 60 minut, a po podaniu środków nasennych pojawiają się problemy z oddychaniem. Zatem u wielu zwierząt co najmniej połowa mózgu pozostaje w stanie czuwania. U człowieka umiejętność półsnu-półjawy nie wykształciła się i dlatego pojawia się pytanie: czy podczas snu osiągnięty ewolucyjnie kompromis neurosomatyczny jest optymalny dla organizmu?

2. Deficyt motoryczny

Propagowane są powszechnie wszelkie aktywne formy wypoczynku, ale gdy przychodzi pora snu, wciąż obowiązuje konserwatywna zasada, iż dla optymalnej regeneracji organizmu wskazany jest całkowity spokój, a to oznacza wielogodzinny bezruch ciała podczas snu. Powszechne jest przekonanie, że warunki, w jakich nasz organizm funkcjonuje podczas snu, to rezultat złożonych procesów ewolucyjnych, które zachodziły na przestrzeni wielu tysięcy lat – skoro tak spali nasi przodkowie i praprzodkowie, więc nic nie należy tu zmieniać. Czy takie przekonanie jest słuszne?

Potrzeba ruchu jest naturalna i podświadoma. Usiadłszy z nogami luźno zwisającymi, bez dotykania podłogi, wkrótce zaczynamy nimi kiwać niczym wahadłami, bo taki ruch wymaga niewielkiego wysiłku i przynosi naszemu organizmowi relaks. Podobne odruchy demonstrujemy, siedząc na fotelu obrotowym lub bujanym, na huśtawce, a także przy częściowym zanurzeniu ciała w wodzie. Spontaniczna aktywność fizyczna jest potwierdzeniem, że nasz organizm źle znosi dłuższe stany bezruchu i każdą nadarzającą się okazję stara się wykorzystać, aby ograniczać bezruch kosztem możliwie małego wydatku energetycznego.

Z ewolucyjnego punktu widzenia, skoro życie biologiczne powstało i rozwinęło się w morzach, to wraz z wyjściem naszych prapraszczurów na ląd swoboda ruchu została mocno zredukowana. Zamiana 3-wymiarowej przestrzeni wodnej na 2-wymiarową płaszczyznę lądu była związana m.in. z utratą kołysania wywoływanego falowaniem wód morskich. Utraciliśmy naturalne kołysanie również podczas snu. Beztroski okres życia płodowego spędzony w owodni w łonie matki, w warunkach symulujących stan nieważkości, kończy się w momencie porodu dramatycznym przejściem na nieruchome podłoże, pod działanie przygniatających sił grawitacji. Taka nagła zmiana to stan niekorzystny dla dziecka, ale matka kołysaniem przynosi mu ulgę. Czyż kołysząc niemowlę tylko do chwili, aż przestanie dopominać się o to płaczem i zaśnie, nie blokujemy na całe późniejsze życie naszej uzasadnionej biologicznie potrzeby kołysania?

Nieświadome zmiany ułożenia ciała podczas snu, występujące średnio co kilkanaście minut oznaczają, że organizm usiłuje się bronić przed wielogodzinnym bezruchem i zagrożeniami, jakie pojawiają się w rezultacie fizycznej immobilności, tj. upośledzenia cyrkulacji krwi i limfy, obniżenia metabolizmu, groźby powstania lokalnych zatorów, utrudnienia oddechu i niedotlenienia oraz ucisku połączonego z drętwieniem kończyn. Wykazywana śladowa aktywność fizyczna podczas snu obniża stopień zagrożenia zdrowia, ale jest zbyt skąpa, by to zagrożenie wyeliminować całkowicie. Sen powinien być krzepiący, ale doświadczony lekarz dobrze wie, że przełom nocy i dnia to pora szczególna, bo właśnie wtedy nasila się liczba zgonów, zawałów serca, ataków astmy i objawów innych ciężkich chorób [120]. Ta paradoksalna regularność – kryzys w trakcie odpoczynku lub wkrótce po nim – wynika ze zjawisk relaksacji i wiąże się m.in. z nocnym deficytem motorycznym. W tych warunkach wyłania się zatem potrzeba kompensacji, tzn. uzupełnienia ruchliwości ciała poprzez zewnętrzne oddziaływania mechaniczne na organizm nie zakłócające snu.

Na uwagę zasługuje stosunkowo skromna ilość informacji zawartych w opracowaniach dotyczących fizjologii człowieka, a odnoszących się bezpośrednio do specyfiki funkcjonowania poszczególnych organów i układów podczas snu. W 800-stronicowym źródłowym dziele o fizjologii człowieka [31] w sumie zaledwie kilkanaście stron dotyczy zagadnień snu.

Wiedza o mózgu jest wciąż stosunkowo uboga, także o jego funkcjach podczas snu i oddziaływaniu we śnie na organizm. Gwałtowne ruchy gałek ocznych występujące w trakcie marzeń sennych, trwające do 20 minut i występujące 5-6 razy podczas każdej nocy praktycznie u każdego (w tym też u noworodków, a także u innych ssaków i ptaków), zostały po raz pierwszy opisane przez Eugeniusza Aserinsky'ego w 1952 roku [10]. Pomimo zakrycia gałek ocznych powiekami, te ruchy są widoczne nawet gołym okiem, ale wiemy o tym zaledwie od 53 lat! Faza snu REM (ang. *rapid eye movement*), zwana też snem paradoksalnym, w znacznym stopniu determinuje nasz stan psychofizyczny i wciąż fascynuje badaczy.

3. Czynniki ryzyka

Raporty Światowej Organizacji Zdrowia [114] przedstawiają stan zdrowotności w danym kraju przez podanie ilości zgonów. Globalnie przyczynami zgonów są następujące schorzenia (w kolejności występowania): 1 – niewydolność układu krążenia; 2 – nowotwory; 3 – niewydolność układu oddechowego. Dobowej pory zgonów nie podaje się. Przedstawianie stanu zdrowotności metodą nie-wprost pozwala uniknąć błędów diagnozowania – powszechnego zjawiska, którego skalę ilustruje analiza norweskich raportów z 222 autopsji [66]: bezpośrednie przyczyny zgonów były trafnie określone przez lekarzy w 70%, natomiast pośrednie przyczyny były zdiagnozowane prawidłowo w 75% przypadków, a w 26 przypadkach (12%) faktyczna przyczyna śmierci nie była zgodna ani z podaną w akcie zgonu, ani z diagnozą lekarską – w 6 przypadkach można domniemywać, że trafna diagnoza i odpowiednia terapia mogły uratować życie.

W normie sen wywołuje nieznaczne upośledzenie oddychania pęcherzykowego – zwiększenie CO2 (2-8 mm Hg PaCO2) oraz spadek O2 (3-11 mm Hg PaO2), co w rezultacie prowadzi do 2%-go obniżenia natlenienia hemoglobiny w tętnicach; obwodowe ciśnienie krwi i szybkość pulsu serca stopniowo obniżają się podczas snu, natomiast wydajność serca obniża się lub pozostaje stała; występowanie cyklicznej zmienności w oddychaniu, ciśnieniu krwi i szybkości pulsu jest normą; hemodynamika jest bardziej zmienna w fazach REM niż NREM i dlatego podwyższone ryzyko wystąpienia stanu niedokrwienia mięśnia sercowego łączy się z fazą REM [91].

Zbadano, że wskaźnik umieralności jest najniższy u osób, które sypiają 7-8 godzin na dobę (jest to norma snu typowa dla osób zdrowych, czynnych zawodowo) i wzrasta w miarę wydłużania się lub skracania czasu snu. Dla snu poniżej 4 godzin wskaźnik ten wzrasta 2,5-krotnie [10], [111]. Ta regularność nie oznacza wprost, że niedobór lub nadmiar snu szkodzą, bo najczęściej to właśnie znaczne odchylenia od normy są symptomami choroby – która może być przyczyną przedwczesnej śmierci – a nie sam sen. Odchyłki od normy mogą – ale nie muszą – oznaczać pojawienie się choroby. Wśród noworodków urodzonych w nocy stwierdzono wyższą śmiertelność [22].

Szczególną chorobą, objawiającą się wyłącznie podczas snu jest bezdech tj. wstrzymanie oddechu. Wystąpienie do 5 krótkich epizodów bezdechu w ciągu godziny snu na ogół uważa się za normalne; w obturacyjnym bezdechu spadki natlenienia (SaO2) są różne u poszczególnych pacjentów i różnią się pomiędzy sobą; maksymalna aktywizacja mięśni i wzrost ciśnienia występują przy zakończeniu bezdechu, podczas krótkotrwałego przebudzenia; cykliczne wzrosty ciśnienia w arteriach płuc wywołują wzrost obciążenia komorowego i mogą przyczyniać się do przerostu komory serca; niedotlenienie krwi tętniczej wywołuje: bradykardię zatokową, zatrzymania zatokowe trwające przez 2-13 sekund, blok serca 2-go stopnia oraz komorową tachykardię; przyjmuje się, że przyczyną nagłego zgonu w obturacyjnym bezdechu jest krytyczna arytmia, jednak nie wiadomo, jakie są udziały zatokowej bradykardii oraz komorowej tachykardii [91]. Jeśli autogeniczny układ kontrolny działa sprawnie, rytm oddechowy zostaje przywrócony, jednak czasem już po gwałtownym przebudzeniu, połączonym z nieprzyjemnym, często bolesnym spazmem mięśni. Bezdech stanowi zagrożenie życia, a w najlepszym przypadku wywołuje fragmentację snu [45], [59]. Wymuszonym przebudzeniom towarzyszy mobilizacja całych grup mięśni, prowadząca do hyperkapnii (stały nadmiar CO2 we krwi) i w efekcie wywołująca uczucie chronicznego zmęczenia oraz liczne negatywne efekty psychofizyczne [5], [45], [51].

U 483 pacjentów z implantowanymi rozrusznikami serca stwierdzono, że większość epizodów arytmii odnotowano późnym ranem [104]. Niedokrwienia mięśnia sercowego, zawały, udary i nagłe zgony występują najczęściej w godzinach 6-12, tj. we śnie lub zaraz po przebudzeniu się; cykliczne zmiany rytmu sercowego oraz ciśnienia krwi są przede wszystkim wywołane przez autonomiczny system nerwowy; w epizodzie bezdechu zwiększona parasympatyczna aktywność eferentna nerwu błędnego wywołuje cykliczne obniżanie się rytmu serca, a cykliczne wzrosty ciśnienia są prawdopodobnie efektem występujących peryferyjnych wazoskurczów, indukowanych przez periodyczną aktywizację sympatycznego układu nerwowego [92].

Nadciśnienie występuje w stanie czuwania u połowy pacjentów z bezdechem obturacyjnym i wiąże się z podwyższonym ryzykiem wynikającym z otyłości i starszego wieku lub otyłości u młodych mężczyzn cierpiących na bezdech. U mężczyzn chrapiących badania statystyczne wykazują podniesione ryzyko wystąpienia choroby niedokrwiennej serca i udaru, a także nagłego zgonu podczas snu, zwłaszcza w przypadkach niepodjęcia środków zaradczych [92].

W USA rocznie 250.000 zawałów serca i 38.000 nagłych zgonów następuje w nocy [56]. Analizie poddano pory ich występowania między północą a godziną 6 rano; pory nasilania się oraz spadku trendu:

- 11.633 nocnych zawałów serca:
  - nasilenie w godzinach 0 - 1;
  - minimum w godzinach 3 - 4;
- 1.200 ataków arytmii (aktywacja rozrusznika serca):
  - nasilenie w godzinach 1 - 2;
  - minimum w godzinach 4 - 5;
- 1.981 nocnych nagłych zgonów kardiologicznych:
  - nasilenie w godzinach 1 - 2;
  - minimum w godzinach 4 - 5.

Analizując 4.920 zgonów w USA w przedziale czasowym między godziną 2 w nocy a szczytem o 8 rano zaobserwowano 60%-owy wzrost ich występowania w stosunku do średniego rozkładu dobowego [69]. Autorzy sugerują, że w części przypadków wcześniejszy zgon mógł zostać zauważony dopiero rano i tak był zarejestrowany. Drugie, mniejsze nasilenie występuje ok. godziny 18. Zróżnicowanie godzinowe dotyczy głównie zgonów osób w wieku powyżej 65 lat, natomiast młodsi umierają o mniej regularnych porach. W chorobach niedokrwiennych mięśnia serca (50% wszystkich przypadków) nasilenie zgonów nastąpiło o 8 rano, w nadciśnieniu – nasilenie o godzinie 1 w nocy, ale tylko u kobiet, natomiast przy wylewach – nasilenie wystąpiło o godzinie 6 rano, ale głównie wśród mężczyzn

Rozkład dobowy nagłych zgonów kardiologicznych w domu na podstawie aktów zgonu wykazuje w USA nasilenie między 7 a 11 rano i obniżenie w porze nocnej [72]. Taki rozkład dobowy dla 2.203 przypadków – jest zdaniem autorów – podobny do podawanych w literaturze typowych pór występowania zawałów i ostrych stanów niedokrwiennych, co wydaje się potwierdzać hipotezę, iż nagłe kardiologiczne zgony spowodowane są niedokrwieniem mięśnia sercowego lub wystąpieniem arytmii. W analizie 74 nagłych zgonów na nieostrych oddziałach szpitalnych również w USA najwięcej odnotowano ich w nocy, wśród starszych pacjentów, a połowy prawdopodobnie można było uniknąć [53].

Rozkłady pór występowania zawałów serca, komorowej arytmii oraz nagłych zgonów na serce są wyraźnie nierównomierne (P<,001), co potwierdza tezę, że za wywoływanie poważnych symptomów kardiologicznych może być odpowiedzialna związana ze stanem snu zmienność aktywności autonomicznego systemu nerwowego. Wczesne maksimum (godziny 1-2 w nocy) może wskazywać, że zgony nocne istotnie bywają odnotowywane jako poranne [69], 53], choć zakończenie analizy zgromadzonych danych już na godzinie 6 rano utrudnia ocenę wiarygodności wyników [4], [72], [104].

Zmienność rytmu pracy serca jest w nocy zazwyczaj większa niż w dzień, co prawdopodobnie świadczy o dominującej roli nerwu błędnego podczas snu, a to z kolei powinno teoretycznie wiązać się z obniżeniem ryzyka wystąpienia zaburzeń pracy serca w nocy; wskaźnik LF/HF

wynikający ze spektralnej analizy zmienności rytmu sercowego, dla trzech stanów: czuwanie–NREM–REM u zdrowych przybiera wartości: 4–1.22–3, natomiast u zawałowców odmiennie: 2.4–5.11–8.9 [110]. Zawał wywołuje więc utratę zdolności fizjologicznej aktywacji nerwu błędnego, utratę nocnej zmienności rytmu pracy serca, a w rezultacie - względną dominację układu sympatycznego podczas snu. Konsekwencją tych zjawisk jest zwiększone ryzyko nagłego zgonu w nocy. U osób, które przeżyły zawał serca, lecz wkrótce potem zmarły, stwierdzono, że wartość Qtc była stała w ciągu doby, natomiast fazy QT w ciągu nocy były skrócone z 388 do 357 ms [117].

W Niemczech wśród 703 nagłych zgonów na serce większość odnotowano w godzinach 6-12 rano; drugie nasilenie wypadało po południu w godz. 15-19, przy czym osoby powyżej 65 roku życia umierają częściej podczas szczytu porannego [4]. W Japonii umiera nagle około 100.000 osób rocznie (1998), z czego 60-70% na niewydolność układu krążenia, a 1/3 nagłych zgonów następuje w nocy podczas snu [46]. Diabetycy z ciężką neuropatią autonomiczną często umierają nagle podczas snu – w zbadanych przypadkach nie ustalono bezpośredniej przyczyny zgonów, ale stwierdzono, że nie nastąpiły w następstwie dysfunkcji oddechowej czy bezdechu [13].

Gdy w latach 1981-88 wśród imigrantów z Tajlandii w USA odnotowano znaczny wzrost nagłych zgonów podczas snu, za jedną z możliwych przyczyn podano występowanie ostrej arytmii [76]. Istnieje związek między opisanym syndromem wśród młodych azjatyckich imigrantów w USA a podobnymi przypadkami odnotowanymi w Tajlandii i – zdaniem autora – prawdopodobnie przyczyna tych zgonów wynika ze szczególnych uwarunkowań genetycznych i kulturowych w tym rejonie świata [73]. Patogeneza syndromu może tkwić w zaburzeniach snu, a także w upośledzeniu przewodnictwa mięśnia sercowego [65]. W Japonii występuje podobny syndrom, nazywany „pokkuri" gdy we śnie umierają młodzi, zdrowi mężczyźni [38].

Wśród 215 zgonów pacjentów z chroniczną patologią oddechową, stosujących respirator (O2), 75% zejść śmiertelnych odnotowano w szpitalu [118]. Przyczyny: ostra niewydolność oddechu (38%); niedoczynność serca (13%); zapalenie płuc (11%); zator płuc (10%); arytmia serca (8%); nowotwór płuc (7%). W badanej grupie 26% zgonów nastąpiło nagle - stwierdzono mniejszą prężność w arteriach (Pa, CO2), zmniejszone dobowe zużycie tlenu oraz częściej występującą arytmię oddechu, a 20% zgonów nastąpiło podczas snu, ale nie stwierdzono związku ilości zgonów z porą doby.

4. Patogenna istota snu

Z przytoczonych doniesień wynika ważny wniosek, który potwierdza postawioną tezę, że w pewnych okolicznościach sen jest stanem bardziej patogennym niż stan dziennej aktywności:

- we śnie nieoczekiwanie może umrzeć zarówno niemowlę, jak i młody, zdrowy mężczyzna;
- najczęściej występującymi zagrożeniami dla osób „zdrowych" to bezdech oraz arytmia serca;
- mechanizmy nagłych ataków kardiologicznych i zgonów podczas snu nie są w pełni poznane;
- zapewne nerw błędny spełnia kluczową rolę w czuwaniu nad organizmem podczas snu;
- wydaje się, że fazy REM są dla bezpieczeństwa organizmu podczas snu okresami krytycznymi;
- ryzyko zgonu we śnie zwiększają: przebyty zawał, niedokrwienność i wady przewodnictwa;
- w miarę upływu lat prawdopodobieństwo, że zgon nastąpi podczas snu lub rano zwiększa się;
- związane z relaksacją podczas snu patogenne czynniki podnoszą wskaźniki umieralności.

Sen pozwala odświeżyć mózg, usunąć objawy zmęczenia i stresu, a także zapomnieć o kłopotach. W wielu dolegliwościach częstym zaleceniem lekarza jest leżenie w łóżku, bo sen krzepi, uspokaja i pomaga odzyskać siły i zdrowie. Jest to prawda, choć trzeba mieć świadomość, że sen jest kompromisem systemowym: w młodości sen to prawie samo zdrowie, ale z biegiem lat ta proporcja zmienia się i wreszcie może przyjść noc, podczas której w miarę upływu godzin snu organizm zamiast sukcesywnie regenerować poszczególne tkanki, a więc polepszać swoją kondycję od chwili zaśnięcia, pogrąża się w stanie krytycznym, kończącym się albo przebudzeniem z ostrymi objawami choroby, albo też... nieprzebudzeniem.

Obserwowane nasilanie się poważnych symptomów kardiologicznych w nocy oraz w godzinach rannych wydaje się potwierdzać tezę, że deficyt motoryczny w czasie snu jest niekorzystny, a w stanach krytycznych jest wręcz destrukcyjny. Trudno ocenić, jak bardzo groźne dla zdrowia są także doraźnie niezauważalne, ale prawdopodobnie kumulujące się przez lata nocne mikrokryzysy, wynikające z DMS. Zarówno analiza epidemiologii nagłych niedomagań podczas snu, jak i przytoczone dane WHO w większości dotyczą zjawisk definitywnych, tj. zgonów. Wykorzystanie tej klasy danych do analizy skutków deficytu motorycznego w nocy wynika z powszechnie stosowanej metodyki i bynajmniej nie oznacza, że względny bezruch podczas snu powoduje zawsze zagrożenie życia czy poważne ryzyko dla zdrowia. Niemniej, skoro ze snem związane są zarówno efekty pozytywne, jak i negatywne pod względem zdrowotnym, należy zadbać, aby dobra było więcej niż zła.

Przy występującym deficycie motorycznym sytuacja osoby zdrowej podczas snu przypomina stan upośledzenia fizycznego osoby niepełnosprawnej. Doprawdy, „sen równoważy pasterza i króla, prostaka i mędrca" (Cervantes), bo niezależnie od płci, statusu społecznego i materialnego, a nawet wieku – z uwagi na zagrożenia wynikające z wielogodzinnego bezruchu, szanse na przebudzenie się w niepogorszonym stanie zdrowia wydają się być wyrównane. I to w dół, skoro bezdech, zawały oraz wylewy atakują podczas snu, często z nieodwracalnym skutkiem.

5. Hipoteza

Ruch – atrybut witalności – wywiera pozytywny wpływ na funkcje biologiczne u człowieka, także gdy jest to ruch bierny, aplikowany w warunkach deficytu motorycznego podczas snu. Subtelne kołysanie, wywołane odpowiednio skonfigurowanymi w czasie i w przestrzeni periodycznymi oddziaływaniami grawitacyjno-inercyjnymi o charakterze harmonicznym, wywołuje korzystne, normujące efekty zdrowotne. Oddziaływania te stanowią podstawę zaproponowanej metody stymulacji oscylacyjnej snu (SOS).

Łańcuch przyczynowo-skutkowy, na którym oparta jest koncepcja SOS przedstawia następujący schemat:
- lądowe środowisko życia -
- wysoka specjalizacja mózgu -
- konflikt neurosomatyczny -
- deficyt motoryczny snu (DMS) -
- stymulacja oscylacyjna snu (SOS).

Oparta na zasadach fizjologii koncepcja stymulacji oscylacyjnej snu (SOS), kontestująca tradycyjne przekonanie, iż podczas snu konieczny jest całkowity bezruch ciała, powstała w efekcie wieloletniej obserwacji trybu życia sportowców niepełnosprawnych [81]. Nie poddawszy się swemu losowi, cieszą się oni życiem i sięgają po wymierne sukcesy, pamiętając o zasadach aktywnej rehabilitacji:

- należy zapewnić organizmowi warunki, w których możliwe jest jak najpełniejsze wykorzystanie jego fizjologicznych możliwości kompensacyjnych;
- dla utrzymania jak najwyższej sprawności fizycznej konieczne jest ciągłe wykorzystywanie istniejących – nawet śladowych – możliwości ruchowych, ponieważ bezruch jest patogenny;
- nawet ćwiczenia fizyczne o niewielkiej intensywności – zarówno aktywne jak i bierne – mogą przynieść odczuwalne efekty, jeżeli są odpowiednio zróżnicowane, są aplikowane w seriach wielokrotnych powtórzeń, przedzielanych relaksem i jeżeli są kontynuowane systematycznie;
- głównie poprzez ruch należy dbać o prawidłowe funkcjonowanie najważniejszych układów: krążenia, oddechowego, moczowego i odpornościowego oraz zapobiegać odleżynom;
- należy zwalczać nieracjonalne stereotypy – w każdej sytuacji przymusowej warto szukać jej pozytywnych stron i przykładowo należy wykorzystywać siły grawitacji i inercji w taki sposób, aby przeciwności przemienić w swoich sprzymierzeńców;
- warunkiem dobrego samopoczucia jest zdrowy sen;
- dla zachowania pozytywnego stanu emocjonalnego należy stymulować system nerwowy;
- aby uniknąć rozczarowań, warto liczyć przede wszystkim na siebie;
- wykorzystywanie pomocy technicznych powinno być ograniczone do rozsądnego minimum, nawet jeżeli takie ograniczenie wiąże się z odczuwalnym dyskomfortem;
- urządzenia wspomagające powinny być w jak najwyższym stopniu funkcjonalne i estetyczne, aby wśród osób zdrowych wywoływały uznanie, a najlepiej – uczucie zazdrości.

6. Uwarunkowania fizjologiczne

Zainicjowane badania laboratoryjne posłużą do opracowania modelowego urządzenia aplikacyjnego. Następnie poddane ono zostanie badaniom klinicznym w celu zweryfikowania spodziewanych efektów medycznych, a także dla zdefiniowania wskazań i przeciwwskazań proponowanej metody. Subtelna stymulacja oscylacyjna, stosowana przez dłuższy okres odpowiednio [54], [113] i tolerowana przez użytkownika, powinna umożliwić:

- poprawę funkcji układu oddechowego, a w tym ograniczenie bezdechu oraz efektu chrapania;
- poprawę cyrkulacji płynów ustrojowych i dystrybucji hormonów, enzymów, limfocytów itp.;
- unormowanie procesów metabolicznych, w tym zmniejszenie ubytku azotu, wapnia i fosforu;
- wzrost odporności – utrudnianie powstawania zastojów, zatorów, złogów, ognisk zapalnych;
- unormowanie funkcji układów: limfatycznego, dokrewnego oraz trawiennego;
- ograniczenie drętwienia przez lepsze ukrwienie kończyn i stymulowanie zmian ułożenia ciała;
- poprawę regeneracji układu kostno-stawowego przez efekt dystrakcji przy lepszej cyrkulacji;
- polepszenie odpoczynku, zmniejszenie wysiłku poprzez rozluźnienie mięśni i ułatwiony ruch;
- profilaktykę odleżynową dzięki poprawie trofiki oraz stanu tkanki w uciskanych strefach;
- korzystniejsze warunki regeneracji mózgu poprzez przedłużenie faz snu głębokiego;
- obniżenie odczuwania bólu dzięki rozluźnieniu i wprowadzeniu bodźców rozpraszających;
- ograniczenie bezsenności dzięki efektowi jednostajnego, harmonicznego kołysania.

7. Doświadczenia laboratoryjne

W rozważaniach teoretycznych nad patomorfologią DMS i możliwościami ograniczenia jego negatywnych skutków pomocne mogą być doświadczenia zgromadzone podczas badań laboratoryjnych, w których zauważono wpływ kołysania na system biologiczny. W hodowli komórek platforma kołyskowa jest powszechnie stosowana dla zapewnienia warunków natleniania zbliżonych do naturalnych, np. przy hodowli śluzówki pęcherza moczowego człowieka [85], komórek nerkowych [79] czy komórek korowych mózgu zarodka myszy [43]. W porównaniu do statycznych warunków hodowli, przy kołysaniu (2,5 cykli/min) polepszającym natlenienie i wymianę pokarmu i odpadów, kultury komórek LLC-PK1 nabłonka nerki świni zanurzane cyklicznie w surowiczym medium wykazują 1,5-krotne zwiększenie enzymatycznej aktywności mitochondrii, a także zbliżone do warunków normalnych uzależnienie ilości syntetyzowanego amoniaku od wartości pH medium [15]. Niedotlenienie hodowli komórek nerki LLC-PK1, występujące przy braku kołysania, wywołuje obniżenie metabolizmu utleniającego i upośledzenie funkcji różnicowania badanej kultury – w warunkach niedostatecznej dostawy tlenu enzym kinazy białkowej C (PKC) jest mniej aktywny i koncentruje się przy błonie zamiast we wnętrzu komórki [88].

Wzrost siatki komórek zębiny pobranej od młodego szczura jest bardziej prawidłowy w warunkach kołysania niż w bezruchu [37]. Hodowla *in vitro* tkanki jelitowej prosięcia daje lepsze efekty przy kołysaniu [78]. Poprawa morfologii kultur śródbłonkowej tkanki rogówki ludzkiej hodowanej w warunkach kołysania zwiększa szanse leczenia ran rogówki – kołysanie bowiem przyspiesza pojawienie się warstw osnowy, a także polepsza strukturę wyhodowanej tkanki [86].

Kołysanie utrudnia zrastanie się jednowarstwowych hodowli komórek mięśni gładkich i śródbłonka aorty [17]. Komórki śródbłonka w efekcie kołysania wydzielają metabolity cyklo-oxygenazowe [42]. Platforma kołyskowa służy do

produkcji transgenicznych przepiórek, w której cykl rozpoczynany jest od stadium embrionu jednokomórkowego [80].

Jak wynika z raportów wpływ kołysania na warunki aktywizacji poszczególnych mechanizmów systemu immunologicznego jest niejednoznaczny [47], [49], [50], [68], [83]. Fibroblasty C3H myszy hodowane in vitro w warunkach kołysania wykazują 3-krotnie niższy wskaźnik nieśmiertelnych mutacji o poczwórnym DNA w porównaniu z hodowlą w warunkach stacjonarnych i identycznych parametrach [16]. Istnieje pogląd, że spadek odporności immunologicznej w przypadku HIV może wynikać z upośledzenia mechanizmu przenikania limfocytów T do gruczołów dokrewnych; w warunkach dynamicznych glikoproteina 120 HIV-1 wykazuje mniejsze zdolności wiązania z aktywnymi glikoproteinami CD38, odgrywającymi istotną rolę w mechanizmie przenikania limfocytów T [11], [20].

W warunkach in vitro supresyjne działanie limfocytów śledziony na płytkotwórcze kultury Mishell-Dutton jest silniejsze w warunkach kołysania w porównaniu z warunkami stacjonarnymi [24]. Poddane stymulacji kołyskowej komórki śródbłonka wykazują obniżone przyleganie granulocytów [119]. Kołysanie zmniejsza aktywność wytworzonych in vitro makrofagów, co wynika z ich – prawdopodobnie indukowanego kontaktowo – małego zasięgu działania cytostatycznego [97].

Istnieje teoria, w której elektromagnetyczna koherentna wibracja leukocytu ludzkiego (zjawisko Frohlicha) z częstotliwością od 0,1 do 10 THz, spowodowana bliskością antygenu, wywołuje przemieszczanie się elementów spolaryzowanych i w efekcie obniża obserwowaną zdolność leukocytu do przylegania [82]. Wpływ ruchu na namnażanie się wirusów nie jest wystarczająco poznany, ale przypuszcza się, że kołysanie może być czynnikiem zwiększającym podatność komórki na infekcję oraz czynnikiem sprzyjającym rozwojowi wirusów - być może za sprawą protein stresu; natomiast hodowla wirusa polio w warunkach kołysania wykazuje wyraźny spadek wzrostu płytki przy wolniejszych oscylacjach niż 1 cykl/min [87].

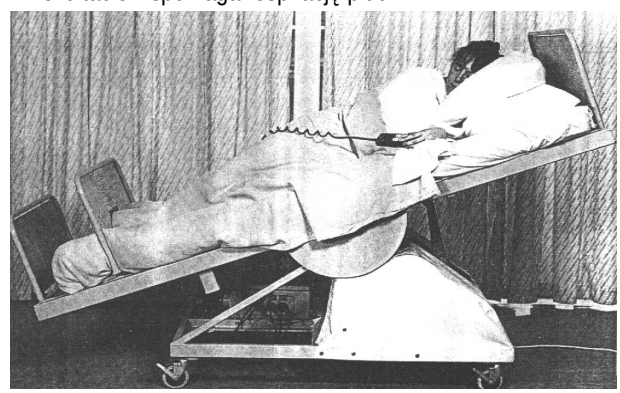
U przytomnego królika poddanego kołysaniu stwierdzono polepszenie krwiobiegu w ważniejszych częściach mózgu oraz poprawę ich natlenienia (pO2), a także obniżenie rytmu serca [8]. 3-godzinne kołysanie powoduje wzrost częstości (23%) oraz głębokości oddechu u królika, a także spadek natlenienia krwi żylnej (2%) przy niezmienionym nasyceniu O2 w tętnicach, co tłumaczone jest większą konsumpcją tlenu w efekcie zwiększenia tonusa mięśni i wzmożenia oddechu [84]. U psa 30-minutowe kołysanie wywołuje spadek (20%) wchłaniania aminokwasu (glicynina) i glukozy oraz wzrost (28%) wchłaniania pierwiastków śladowych (mangan), co może być skutkiem pojawiającego się deficytu witamin w jelicie cienkim [27].

8. Doświadczenia kliniczne

Prawdopodobnie najwcześniejsze spośród doniesień o zastosowaniu kołysania ciała w innym celu niż pielęgnacja niemowląt pochodzi z Anglii z 1932 roku [26]. Wydaje się, że na publikację tę należy zwrócić szczególną uwagę, ponieważ wywarła ona bardzo silne wrażenie w całym świecie lekarskim, które to wrażenie przetrwało do czasów współczesnych [14]. Zapewne stało się tak, ponieważ doniesienie ukazało się na łamach prestiżowego magazynu *Lancet*, a przedstawiono w nim bardzo prostą metodę postępowania, skuteczną dla ratowania życia i uzasadnioną przy pomocy przejrzystego modelu.

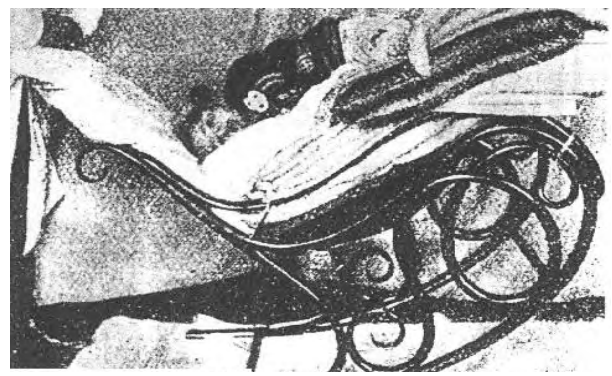
Opisany przypadek dotyczy paraliżu mięśni przepony u dwuletniej dziewczynki w następstwie przebytego dyfterytu. Dla ratowania chorego dziecka przed uduszeniem obfitą wydzieliną z oskrzeli doktor Frank C. Eve ułożył chorą na pochylonym łóżku głową do dołu. Następnie, aby uniknąć zapalenia płuc, łóżko zastąpił bujanym fotelem (Rys. 1), przy pomocy którego rodzice kołysali na zmianę swoją córkę w zakresie ± 30° w rytmie ok. 10 cykli na minutę przez 2 doby. Po powrocie funkcji przepony wystąpił z kolei zanik funkcji mięśni klatki piersiowej, który ustąpił po następnych 4 dniach i dziecko stopniowo powróciło do pełnego zdrowia.

Autor opisanej procedury terapeutycznej wytłumaczył respiracyjny efekt kołysania naprzemiennie głową do góry i w dół, podobieństwem do działania tłoka w cylindrze – cykliczna zmiana pozycji ciała pod wpływem sił grawitacji wywołuje przemieszczanie się zawartości komory brzusznej, wywierając przy tym zmienne ciśnienie na przeponę, co w rezultacie wspomaga respirację płuc.



Rys. 1. Pionierskie zastosowaniu fotela bujanego w celu wspomagania funkcji oddechowej dla ratowania życia. Źródło: [26]

Autor podał również propozycję, popartą fotografią, jak w nagłych przypadkach, przy braku fotela bujanego, zaaranżować prosty respirator, wykorzystując drabinę, poprzecznie opartą na górnej krawędzi płotu w połowie swojej długości. Taki model łóżka kołyskowego obowiązuje do dziś (Rys. 2).



Rys. 2. Współczesne łóżko kołyskowe stosowane w niewydolnościach układu oddechowego. Źródło: [14]

Pionierskie rozwiązanie prowincjonalnego lekarza z Hull zostało rozpowszechnione w latach 50-ych, gdy podczas epidemii polio zastosowano łóżka kołyskowe z napędami

mechanicznymi, oparte na oryginalnym modelu Eve'a [14], [60], [90], [95]. Z czasem dodano również ruchliwość w płaszczyźnie prostopadłej do osi długiej ciała, gdy stwierdzono, że wspomaganie funkcji oddechowych występuje również przy kołysaniu poprzecznym [35]. Początkowo łóżka kołyskowe stosowane były tylko w warunkach klinicznych do rehabilitacji funkcji oddechowej. Zasada „tłoka w cylindrze" pozwalała na praktyczne uświadomienie pacjentowi, na czym polega mechanizm wentylacji płuc i z jakich faz mechanizm ten się składa. Urządzenie umożliwiało przeprowadzanie ćwiczeń w celu wytworzenia u pacjenta nowego automatyzmu oddechowego w warunkach patologicznych, tj. w neuromięśniowych deficytach w obrębie klatki piersiowej. Podczas badań klinicznych stwierdzono, że u pacjentów z obfitą wydzieliną w oskrzelach łóżko kołyskowe umożliwia efektywny drenaż dróg oddechowych [60].

Obecnie łóżko kołyskowe, z uwagi na nieinwazyjny charakter, jest zazwyczaj pierwszą pomocą respiracyjną zalecaną w przypadkach chronicznych, do stosowania też w domu podczas snu, ponieważ w hipodynamicznej postaci niewydolności oddechowej pozycja leżąca wiąże się z pogorszeniem warunków respiracji płuc w porównaniu z pozycją siedzącą lub stojącą. W doniesieniu o rekonwalescencji pooperacyjnego porażenia nerwu przeponowego autorzy podają, że powodem zastosowania łóżka kołyskowego u 13 pacjentów była potrzeba poprawienia wentylacji płuc poprzez wywołanie pasywnych ruchów przepony i rozkurczenie dróg oddechowych [1]. Jeśli trudności nie ustępują, dodawane są inne pomoce respiracyjne [14], [19], [32], [44], [106].

Przy stabilnym stanie pacjenta długotrwałe stosowanie łóżka kołyskowego wpływa na obniżenie fragmentacji snu, m.in. w efekcie profilaktyki bezdechu, a w konsekwencji zapewnia ogólną poprawę zdrowia dzięki lepszemu dotlenieniu organizmu, wzmożonej cyrkulacji płynów ustrojowych i ustąpieniu zaburzeń przemian gazowych we krwi [5], [45], [51]. Podczas eksperymentalnego kołysania zaobserwowano wzrost aktywności elektrycznej mięśni oddechowych i przepony, a także zwiększenie się arytmii oddechowej, co jednak nie zmieniło pozytywnej oceny wpływu kołysania na wentylację płuc [34].

Analizowane publikacje potwierdzają wspomagające oddziaływanie kołysania ciała w ciężkich dysfunkcjach oddechowych [5], [34], [45], [71], [94], wynikających w szczególności z takich patologii, jak: dystrofia mięśniowa [14], miopatia [112], patologia motoneuronu [14], poliomyelitis i post-polio [14], [60], [90], [95], neuropatyczne [21] i pooperacyjne [1] porażenie nerwu przeponowego, quadriplegia [33], niedobór maltazy kwasowej [14], [106], chroniczny nieżyt oskrzeli [60], astma pochodzenia oskrzelowego [60] oraz stwardnienie rozsiane [41]. Cytowane doniesienia obejmują wieloletnie obserwacje znacznej liczby przypadków ciężkich patologii, a takim stanom chronicznym często towarzyszą infekcje i inne schorzenia, co oznacza, że te raporty w dużej mierze dotyczą osób w skrajnie złym stanie zdrowia.

W doniesieniu z 1994 roku o długotrwałym stosowaniu łóżek kołyskowych w warunkach domowych [14] zawarta jest informacja o przedwczesnym porzuceniu tej pomocy przez 14-tu pacjentów z objętej badaniami grupy 43-osobowej. Wśród powodów rezygnacji podano: zakłócenie równowagi gazowej we krwi wynikłe z niewystarczającego efektu respiracyjnego kołysania (8 przypadków), utrudniająca podróżowanie stacjonarność kołyskowego respiratora (3 przypadki), dolegliwości układu trawiennego i cofanie się kwasów żołądkowych (2 przypadki) oraz wystąpienie odleżyny (1 przypadek). W tym samym doniesieniu zwrócono uwagę na pozytywne skutki metaboliczne kołysania: zmniejszoną utratę azotu, wapnia oraz fosforu, a także obniżone ryzyko formowania się kamieni w układzie moczowym. Średni okres objętego tym raportem stosowania łóżek kołyskowych wyniósł 16,0 lat.

Potwierdzono także skuteczność łóżek kołyskowych u noworodków (także wcześniaków) w profilaktyce bezdechu [35], [108], chociaż z drugiej strony podczas badań zauważono niebezpieczeństwo obrócenia się niemowlęcia, zwłaszcza przy znacznych wychyleniach domowej kołyski podwieszanej, co może prowadzić do całkowitego mechanicznego zablokowania dróg oddechowych, a w konsekwencji nawet do nagłego zgonu przez uduszenie [2], [6], [7], [18], [70].

Stymulacja warunków zbliżonych do występujących w łonie matki – kołyskowe łóżko wodne w połączeniu z odpowiednio synchronizowaną stymulacją akustyczną – wywołuje u wcześniaków przyspieszenie rozwoju. Dlatego efekty kliniczne kombinacji obu oddziaływań oceniono pozytywnie [52]. W innych badaniach wykazano, że matki, które czują się szczęśliwe, swoje nowonarodzone niemowlęta kołyszą na ręku częściej niż matki będące w złym stanie psychicznym po porodzie [58].

Efekt kołyskowy zastosowano także do profilaktyki i terapii odleżyn [39]. Autorzy proponują kołyskową platformę do ustawienia na niej dowolnego łóżka, której napęd wywołuje powolne pochylanie powierzchni nośnej łóżka z boku na bok. Pełny cykl trwa od 20-tu do 80-ciu minut i pozwala osiągnąć naprzemiennie nachylenia w obu kierunkach w regulowanych zakresach od 0° do ±15°. Koncepcja wywodzi się od klasycznej ramy Strykera, w której przez zmienne nachylenia poprzeczne uzyskuje się lokalne, cykliczne obniżenie nacisków na strefy podporowe ciała, a w rezultacie – jeden z istotniejszych efektów przeciwodleżynowych [67]. Ponadto oscylacje mechaniczne łóżka pomiarowego, wyposażonego w elektromechaniczne czujniki drgań, wykorzystano do diagnostyki: do monitorowania epizodów bezdechów i ruchliwości podczas snu [59], [89], kaszlu, a także do wykrywania ataków epilepsji [109].

Pneumatyczne materace dwukomorowe z ciągłą oscylacją oprócz zastosowań przeciwodleżynowych w domu wykazały swoją przydatność także na oddziałach intensywnej terapii. W grupie pacjentów dorosłych (ciężka niewydolność oddechowa, koma metaboliczna, wylew, przedawkowanie leków, zakażenie) stwierdzono poprawę warunków wyprowadzenia pacjenta ze stanu krytycznego poprzez polepszenie respiracji płuc [18], [105], dopatrzono się zmniejszenia ryzyka wystąpienia zapalenia płuc [29] oraz obniżenia ryzyka – z 27 do 11% – wystąpienia stanu zapalnego dróg moczowych [63]. Natomiast u noworodków zaobserwowano skrócenie okresu, w którym niezbędne było dotchawiczne podawanie tlenu, przy jednoczesnym odnotowaniu braku negatywnego wpływu mechanicznych wymuszeń oscylacyjnych na układ krążeniowo-oddechowy i temperaturę ciała [74]. W innych badaniach z zastosowaniem oscylacyjnego łóżka wodnego u grupy wcześniaków stwierdzono obniżenie rozdrażnienia i niepokoju oraz mniejszą ilość nieskoordynowanych ruchów [51].

9. Dyskusja

Na podstawie badań laboratoryjnych można wysnuć ogólne stwierdzenie, że umiarkowane kołysanie poprawia zazwyczaj warunki funkcjonowania biosystemu. Szczególnie interesująca jest 28%-owa poprawa wchłaniania soli manganu w jelitach zwierzęcia w warunkach kołysania, a polepszenie to – zdaniem autorów – odnosi się także do innych pierwiastków śladowych [27]. Zatem spadek wchłaniania aminokwasów i glukozy powinien być kompensowany podczas programowych przerw w akcji SOS.

Z uwagi na zasygnalizowany w rozdziale 7 brak jednoznaczności co do wpływu kołysania na warunki aktywizacji poszczególnych mechanizmów układu odpornościowego, wskazana będzie duża ostrożność podczas badań klinicznych systemu SOS, zwłaszcza wobec stopnia złożoności procesów immunologicznych [28]. Bezpieczniej jest przyjąć założenie, że różne parametry kołysania mogą mieć różny wpływ na ustrój biologiczny. Również fakt, że w hodowli tkanek kołysanie jest korzystne i z powodzeniem zastępuje układ cyrkulacji płynów, bynajmniej nie oznacza, iż harmoniczne oddziaływania SOS muszą w każdych warunkach przynieść tylko same korzyści dla organizmu.

W analizie zastosowań klinicznych łóżek kołyskowych sumaryczna liczba pacjentów objęta badaniami wśród dorosłych przekracza 100 oraz 30 wśród niemowląt, a najdłuższy okres stosowania łóżka kołyskowego przez osobę z niewydolnością układu oddechowego po polio wyniósł 35 lat [14]. Brak jest informacji, aby kołysanie przeszkadzało w spaniu, ale można domniemywać, że potrzebny jest krótszy lub dłuższy okres adaptacji.

Wnioski wynikające ze stosowania łóżek oscylacyjnych nie mogą być odniesione wprost do koncepcji SOS, bowiem zasada ich działania polega na ciągłej wymianie stref podporowych w wyniku naprzemiennego napełniania powietrzem i opróżniania dwóch wzajemnie przenikających się układów komorowych. Okres oscylacji w aktywnych materacach pneumatycznych to minuty, podczas gdy przy kołysaniu według założeń teorii SOS okres to zaledwie sekundy. Z uwagi na efekt zbliżony do kołysania, łóżko oparte na koncepcji SOS powinno zachować pozytywne cechy łóżka kołyskowego. Także tu najistotniejsza różnica pomiędzy obydwoma systemami dotyczy częstotliwości roboczej, której podwyższenie w koncepcji SOS ma na celu zwiększenie oddziaływań inercyjnych przy zmniejszonym zakresie ruchu.

W przebadanej literaturze niewiele jest informacji, które dotyczą wpływu kołysania ciałem człowieka podczas snu na funkcjonowanie układu krążenia, a właściwie całkiem ich brak w odniesieniu do układu nerwowego i dokrewnego, oddziaływań w przypadkach wzmożonego napięcia mięśniowego, powysiłkowego zmęczenia, przy dolegliwościach bólowych, przy bolesnej menstruacji, w schorzeniach i stanach zapalnych narządów wewnętrznych, w chorobach psychosomatycznych, przy chrapaniu, bezsenności, w warunkach ciąży, w geriatrii, a także w onkologii. Brak jest również informacji, jaki wpływ na organizm człowieka może mieć umiarkowane, odpowiednio dawkowane kołysanie w formie oscylacji liniowych wzdłuż osi długiej ciała oraz w kierunku pionowym. Jednym z oczywistych przeciwwskazań dla SOS może być osobnicza nadwrażliwość błędnika [75]. Wyszukane w trakcie przeglądu literatury doniesienia na temat związku kołysania z chorobą lokomocyjną dotyczą oceny skuteczności badanych środków farmakologicznych, a nie zjawisk biofizycznych związanych ze stanami pogorszonego samopoczucia.

10. Podsumowanie

W powszechnym odbiorze kołysanie jest zagadnieniem trywialnym i z tego powodu wiele jego pozytywnych skutków dla zdrowia mogło umknąć uwadze badaczy. Doświadczenia kliniczne wydają się potwierdzać ogólnie pozytywny wpływ kołysania na oddychanie i na cyrkulację, a także na jakość snu u osób dorosłych i w konsekwencji na ogólny stan zdrowotny. Również u noworodków stwierdzono korzystne oddziaływanie na stan zdrowia i na ich prawidłowy rozwój. Na tej podstawie można założyć, że stymulacja oscylacyjna, wywołująca podobne ruchy harmoniczne ciała podczas snu, też powinna zapewnić odczuwalne wspomaganie układu oddechowego, poprawę cyrkulacji płynów ustrojowych i profilaktykę bezdechu, nie powodując przy tym negatywnych odczuć wywołanych cykliczną pozycją Trendelenburga (leżenie z nachyleniem ciała głową ku dołowi).

Rys. 3. Kołyska z mszału jasnogórskiego 1506-1507. Źródło: [121]

Powszechnie wiadomo, że łagodne kołysanie ciała uspokaja i dlatego tradycyjna kołyska od dawna w każdej kulturze była środkiem pielęgnacji niemowląt (Rys.3). Jednak z dobrodziejstw kołysania podczas snu powinny korzystać również osoby dorosłe z przyczyn pragmatycznych podyktowanych wiedzą biologiczną, a nie dopiero pod przymusem, gdy niewydolność oddechowa staje się krytyczna i zagraża życiu.

Literatura cytowana

1. Abd A. G., Braun N. M., Baskin M. I., O'Sullivan M. M., Alkaitis D. A.: Diaphragmatic dysfunction after open heart surgery: treatment with a rocking bed, Ann. Intern. Med., 111(11): 881-6, 1989.
2. Ackerman J., Gilbert-Barness E.: Suspended rocking cradles, positional asphyxia, and sudden infant death, Arch. Pediatr. Adolesc. Med., 151(6): 573-5, 1997.
3. Andersson S., Lundeberg T.: Acupuncture-from empiricism to science: functional background to acupuncture effects in pain and disease, Med. Hypotheses, 45 (3): 271-81, 1995.
4. Arntz H. R., Willich S. N., Oeff M., Bruggemann T., Stern R., Heinzmann A.: Circadian variation of sudden cardiac death reflects age-related variability in ventricular fibrillation, Circulation, 88(5): 2284-89, Nov 1993.
5. Barlow P. B., Bartlett D. Jr, Hauri P., Hellekson C., Nattie E. E., Remmers J. E., Schmidt-Nowara W. W.: Idiopathic hypoventilation syndrome: importance

of preventing nocturnal hypoxemia and hypercapnia, Am. Rev. Respir. Dis., 121(1): 141-5, 1980.
6. Beal S. M., Byard R. W.: Accidental death or sudden infant death syndrome? J. Paed. Child. Health., 31(4): 269-71, 1995.
7. Beal S. M., Moore L., Collett M., Montgomery B., Sprod C., Beal A.: The danger of freely rocking cradles, J. Paediatr. Child. Health., 31(1): 38-40, 1995.
8. Beketov A. I., Skoromnyi N. A. [Effect of dimephosphon on the blood circulation and oxygen tension in the brain of normal waking rabbits and in motion sickness], Farmakol. Toksikol., 51(5): 35-8, USSR 1988.
9. Bonnet M. H., Arand D. L.: We are chronically sleep deprived, Sleep, 18 (10): 908-911, Dec 1995.
10. Borbély A.: Tajemnice snu, Państwowe Wydawnictwo Naukowe: s.204, Warszawa 1990.
11. Bragardo M., Buonfiglio D., Feito M. J., Bonissoni S., Redoglia V., Rojo J. M., Ballester S.: Modulation of lymphocyte interaction with endothelium and homing by HIV-1 gp120, J. Immunol., 159(4): 1619-27, 1997.
12. Byard R. W., Beal S., Bourne A. J.: Potentially dangerous sleeping environments and accidental asphyxia in infancy and early childhood, Arch. Dis. Child., 71(6): 497-500, 1994.
13. Catterall J. R., Calverley P. M., Ewing D. J., Shapiro C. M., Clarke B. F., Douglas N. J.: Breathing, sleep, and diabetic autonomic neuropathy, Diabetes, 33(11): 1025-27, 1984.
14. Chalmers R. M., Howard R. S., Wiles C. M., Spencer G. T.: Use of the rocking bed in the treatment of neurogenic respiratory insufficiency, QJM, 87(7): 423-9, 1994.
15. Cole L. A., Scheid J. M., Tannen R. L.: Induction of mitochondrial metabolism and pH-modulated ammoniagenesis by rocking LLC-PK1 cells, Am. J. Physiol., 251(2): C293-8, 1986.
16. Curatolo L., Erba E., Morasca L.: Culture conditions induce the appearance of immortalized C3H mouse cell lines, In Vitro, 20(8): 597-601, 1984.
17. Davies P. F., Ganz P., Diehl P. S.: Reversible microcarrier-mediated junctional communication between endothelial and smooth muscle cell monolayers: an in vitro model of vascular cell interactions, Lab. Incest, 53(6): 710-8, 1985.
18. deBoisblanc B. P., Castro M., Everret B., Grender J., Walker C. D., Summer W. R.: Effect of air-supported, continuous, postural oscillation on the risk of early ICU pneumonia in nontraumatic critical illness, Chest, 103(5): 1543-7, 1993.
19. Dettenmeier P. A., Jackson N. C.: Chronic hypoventilation syndrome: treatment with non-invasive mechanical ventilation, AACN Clin. Issues. Crit. Care. Nurs., 2(3): 415-431, 1991.
20. Dianzani U., Funaro A., DiFranco D., Garbarino G., Bragardo M., Redoglia V., Buonfiglio D.: Interaction between endothelium and CD4+CD45RA+ lymphocytes. Role of the human CD38 molecule, J. Immunol., 153(3), 952-959, 1994.
21. DiBenedetto R. J., Firth M., Ham E., Causey D.: Bilateral diaphragmatic paralysis, South. Med. J., 76(10): 1312-4, 1983.
22. Domenighetti G., Paccaud F.: The night – a dangerous time to be born? Br. J. Obstet. Gynaecol., 93(12), 1262-7, 1986.
23. Dröscher V.: Reguła przetrwania, PIW, Warszawa 1982.
24. Eardley D. D., Staskawicz M. O., Gershon R. K.: Suppressor cells: dependence on assay conditions for functional activity, J. Exp. Med., 143(5): 1211-9, 1976.
25. Eipert E. F., Adorini L., Couderc J.: A miniaturized in vitro diffusion culture system, J. Immunol. Methods, 22(3-4): 283-92, Netherlands 1978.
26. Eve F. C.: Actuation of the inert diaphragm, Lancet, 2: 995-997, 1932.
27. Faitelberg R. O., Udalov U. F., Semik L. I., Gladkij T. V., Malachovska V., Balan N. N.: [Absorptive activity of the intestine following administration of the preparation aerovit during rocking], Fiziol. Zh., 24(2): 252-7, USSR 1978.
28. Faleszko W., Gołąb J., Jakóbisiak M., Lasek W., Stokłosa T., Zagożdżon R.: Fascynująca immunologia, Wydawnictwo BART, s.147, Warszawa 1998.
29. Fink M. P., Helsmoortel C. M., Stein K. L., Lee P. C., Cohn S. M.: The efficacy of an oscillating bed in the prevention of lower respiratory tract infection in critically ill victims of blunt trauma. A prospective study, Chest, 97(1): 132-7, 1990.
30. Friedman M. J., Birch S., Tiller W. A.: Towards the development of a mathematical model for acupuncture meridians, Acupunct. Electrother. Res., 14(3-4): 17-26, 1989.
31. Ganong W. F.: Fizjologia lekarska, Państwowy Zakład Wydawnictw Lekarskich, s.859, Warszawa 1973.
32. Gilmartin M. E.: Body ventilators. Equipment and techniques, Respir. Care Clin. N. Am., 2(2): 195-222, 1996.
33. Goldberg A. I., Cane R. D., Childress D., Wu Y. C., Vesely L. C., Pfrommer M.: Combined nasal intermittent positive-pressure ventilation and rocking bed in chronic respiratory insufficiency. Nocturnal ventilatory support of a disabled person at home, Chest, 99(3), 627-9, 1991.
34. Goldstein R. S., Molotiu N., Skrastins R., Long S., Contreras M.: Assisting ventilation in respiratory failure by negative pressure ventilation and by rocking bed, Chest, 92(3): 470-4, 1987.
35. Groswasser J., Sottiaux M., Rebuffat E., Kimon T., Vandeweyer M., Kelmanson I., Blum D., Kahn A.: Reduction in obstructive breathing events during body rocking: a controlled polygraphic study in preterm and full-term infants, Pediatrics, 96(1): 64-68, 1995.
36. Gustafsson H.: Vasomotion and underlying mechanisms in small arteries. An in vitro study of rat blood vessels, Acta Physiol. Scand. Suppl., 614(1-44), Sweden 1993.
37. Hasegawa N.: Effects of various culture conditions on matrix formative functions of rat incisor odontoblasts in a pulp-dentin slice culture system, Shika Kiso Igakkai Zasshi, 31(4): 392-403, Japan 1989.
38. Hayashi M., Murata M., Satoh M., Aizawa Y., Oda E., Oda Y., Watanabe T., Shibata A.: Sudden nocturnal death in young males from ventricular flutter, Jpn Heart J., 26(4): 585-591, Jul 1985.
39. Höök O., Gabrielsson L., Lagerman U.: Prophylaxis and treatment of decubitus ulcers with a rocking bed, Scand. J. Rehabil. Med., 14(1): 33-7, Sweden 1982.
40. Hoppensteadt F. C.: Intermittent chaos, self-organization, and learning from synchronous synaptic activity in model neuron networks, Proc. Natl. Acad. Sci. USA, 86(9): 2991-5, 1989.
41. Howard R. S., Wiles C. M., Hirsch N. P., Loh L., Spencer G. T., Newsom-Davis J.: Respiratory involvement in multiple sclerosis, Brain, 115(Pt 2): 479-94, Apr 1992.
42. Hughes J. H.: Physical and chemical methods for enhancing rapid detection of viruses and other agents, Clin. Microbiol. Rev., 6(2): 150-175, 1993.
43. Humphreys P., Jones S., Hendelman W.: Three-dimensio- nal cultures of fetal mouse cerebral cortex in a collagen matrix, J. Neurosci. Method., 66(1): 23-33, Netherlands 1996.
44. Hyland R. H., Jones N. L., Powles A. C., Lenkie S. C., Vanderlinden R. G., Epstein S. W.: Primary alveolar hypoventilation treated with nocturnal electrophrenic respiration, Am. Rev. Respi.r Dis., 117(1): 165-72, 1978.
45. Iber C., Davies S. F., Mahowald M. W.: Nocturnal rocking bed therapy: improvement in sleep fragmentation in patients with respiratory muscle weakness, Sleep, 12(5): 405-412, 1989.
46. Ito H., Yoshimoto N.: [Nocturnal sudden death], Nippon Rinsho, 56(2): 439-443, Japan Feb 1998.
47. Johnson K. W., Kim D. H., Munson A. E., Holsapple M. P.: Dependence on intact cells for the in vitro activation of dimethylnitrosamine to an immunosuppressive form, Mutat. Res., 182(4): 211-21, Netherlands 1987.
48. Kahn A., Van de Merckt C., Dramaix M., Magrez P., Blum D., Rebuffat E., Montauk L.: Transepidermal water loss during sleep in infants at risk for sudden death, Pediatrics, 80(2): 245-250, Aug 1987.
49. Kim B. S., Yang K. H., Haggerty H. G., Holsapple M. P.: Production of DNA single-strand breaks in unstimulated splenocytes by dimethylnitrosamine, Mutat. Res., 213 (2): 185-93, Netherlands 1989.
50. Kim D.H., Yang K.H., Johnson K.W., Holsapple M.P.: Suppression of in vitro antibody production by dimethylnitrosamine in mixed cultures of mouse primary hepatocytes and mouse splenocytes. Toxicol Appl Pharmacol 87 (1 32-42 1987
51. Korner A. F., Lane N. M., Berry K. L., Rho J. M., Bron B. W. Jr: Sleep enhanced and irritability reduced in preterm infants: differential efficacy of three types of waterbeds, J. Dev. Behav. Pediatr., 11(5): 240-6, Oct 1990.
52. Kramer L. I., Pierpont M. E.: Rocking waterbeds and auditory stimuli to enhance growth of preterm infants. Preliminary report, J. Pediatr., 88(2): 297-299, 1976.
53. Kubalak G., Rhodes M., Boorse D., DqAmelio L. F.: Unexpected death on the non-ICU trauma ward, J. Trauma, 31(9): 1258-1262, Sep 1991.
54. Kuraszkiewicz B. komunikacja osobista, IBIB PAN, Warszawa 1999 – 2000
55. Laszlo A., Muranyi L., Boda D.: [Rocking-respirator-treatment in asphyxia of the newborn], Acta Paediatr. Acad. Sci. Hung, 16(3-4): 257-65, Hungary 1975
56. Lavery C. E., Mittleman M. A., Cohen M. C., Muller J. E., Verrier R. L.: Nonuniform nighttime distribution of acute cardiac events: a possible effect of sleep states, Circulation, 96(10): 3321-3327, Nov 1997.
57. Lavie P.: Czarowny świat snu. Prószyński i S-ka, Warszawa 1998.
58. Livingood A. B., Daen P., Smith B. D.: The depressed mother as a source of stimulation for her infant, J. Clin. Psychol., 39(3): 369-75, 1983.
59. Lojander J., Salmi T., Maasilta P.: Reproducibility of oximetry with a static charge-sensitive bed in evaluation of obstructive sleep apnoea, Clin. Physiol., 18(3): 225-33, 1998.
60. Louvigne Y., deLabarthe J., deLabarthe B., Leroy D.: [Value of rocking beds with motor-produced oscillations in respiratory rehabilitation], Rev. Tuberc. Pneumol., 35(2): 137-54, France 1971.
61. Luu M., Boureau F.: [Acupuncture in pain therapy: current concepts], Ther. Umsch., 46(8): 518-25, Switzerland 1989.
62. Maas J. B.: Potęga snu. Prószyński i S-ka, Warszawa 1998.
63. MacIntyre N. R., Helms M., Wunderink R., Schmidt G., Sahn S. A.: Automated Rotational Therapy for the Prevention of Respiratory Complications during Mechanical Ventilation, Respiratory Care, 44(12), 1999.
64. Markin V. S., Liu D., Rosenberg M. D., Tsong T. Y.: Resonance transduction of low level periodic signals by an enzyme: an oscillatory activation barrier model, Biophys. J., 61(4), 1045-9, 1992.
65. Melles R. B., Katz B.: Night terrors and sudden unexplained nocturnal death, Med. Hypotheses, 26(2): 149-154, Jun 1988.
66. Midelfart J., Aase S.: [The value of autopsy in current health care. An analysis of autopsies performed during one year at the Regionsykehuset in Trondheim], Tidsskr Nor Laegeforen,117(12): 1784-5, Norway, May 1997.

67. Milewski P., Pokora M.: Nowoczesne systemy przeciwodleżynowe. Człowiek niepełnosprawny w społeczeństwie, Gdańsk. PZWL, Warszawa 1986.
68. Misiti J., Waldmann T. A.: In vitro generation of antigen-specific hemolytic plaque-forming cells from human peripheral blood mononuclear cells, J. Exp. Med., 154(4), 1069-84, 1981.
69. Mitler M. M., Hajdukovic R. M., Shafor R., Hahn P. M., Kripke D. F.: When people die. Cause of death versus time of death, Am. J. Med., 82(2): 266-274, Feb 1987.
70. Moore L., Bourne A. J., Beal S., Collett M., Byard R. W.: Unexpected infant death in association with suspended rocking cradles, Am. J. Forensic Med. Pathol., 16(2): 177-180, 1995.
71. Moore L., Bourne A. J., Shneerson J. M.: Respiratory pump failure and its treatment, Pacing. Clin. Electrophysiol., 12(5): 844-6, 1989.
72. Muller J. E., Ludmer P. L., Willich S. N., Tofler G. H., Aylmer G., Klangos I., Stone P. H.: Circadian variation in the frequency of sudden cardiac death, Circulation, 75 (1), 131-138, Jan 1987.
73. Munger R. G.: Sudden death in sleep of Laotian-Hmong refugees in Thailand: a case-control study, Am. J. Public Health, 77(9): 1187-1190, Sep 1987.
74. Murai D. T., Grant J.W.: Continuous oscillation therapy improves the pulmonary outcome of intubated newborns: results of a prospective, randomized, controlled trial. Crit Care Med 22 (7) 1147-54, 1994
75. Murphy T. P.: Mal de debarquement syndrome: a forgotten entity? Otolaryngol. Head Neck Surg., 109(1): 10-13, 1993.
76. Nademanee K., Veerakul G., Nimmannit S., Chaowakul V., Bhuripanyo K.: Arrhythmogenic marker for the sudden unexplained death syndrome in Thai men. Circulation 96 (8) 2595-2600, Oct 1997.
77. Neff R., Blasius B., Beck F., Luttge U.: Thermodynamics and energetics of the tonoplast membrane operating as a hysteresis switch. in an oscillatory model of Crassulacean acid metabolizm. J Membr Biol 165 (1) 37-43, 1998.
78. Nietfeld J. C., Tyler D. E., Harrison L. R., Cole J. R., Latimer K. S., Crowell W.A.: Culture and morphologic features of small intestinal mucosal explants from weaned pigs. Am J Vet Res 52 (7) 1142-6, 1991.
79. Nissim I., Sahai A., Sandler R. S., Tannen R. L.: The intensity of acidosis differentially alters the pathways of ammoniagenesis in LLC-PK1 cells. Kidney Int 45 (4) 1014-9, 1994.
80. Ono T., Murakami T., Mochii M., Agata K., Kino K., Otsuka K., Ohta M.: A complete culture system for avian transgenesis, supporting quail embryos from the single-cell stage to hatching. Dev Biol 161 (1) 126-30, 1994.
81. Pokora M.: Skandynawskie doświadczenia w dziedzinie lokomocji niepełnosprawnych. II Konferencja Naukowa: Rozwój Środków Lokomocji dla Osób Niepełnosprawnych. Krynica s.13-23 suplement, 1999.
82. Pokorny J., Jandova A., Kobilkova J., Heyberger K., Hraba T.: Frohlich electromagnetic radiation from human leukocytes: implications for leukocyte adherence inhibition test. J Theor Biol 102 (2) 295-305, 1983.
83. Poutsiaka D. D., Clark B. D., Vannier E., Dinarello C.A.: Production of interleukin-1 receptor antagonist and interleukin-1 beta by peripheral blood mononuclear cells is differentially regulated. Blood 78 (5) 1275-81, 1991.
84. Radkevich L. A.: [Oxygen content in the blood in prolonged rocking]. Kosm Biol Aviakosm Med 12 (1) 78-80, USSR 1978.
85. Reedy E. A., Heatfield B. M.: Histomorphometry and cell kinetics of normal human bladder mucosa in vitro. Urol Res 15 (6) 321-7, West Germany 1987.
86. Richard N. R., Anderson J. A., Weiss J. L., Binder P. S.: Air/liquid corneal organ culture: a light microscopic study. Curr Eye Res 10 (8) 739-49, 1991.
87. Richards G. P., Weinheimer D. A.: Influence of adsorption time, rocking, and soluble proteins on the plaque assay of monodispersed poliovirus. Appl Environ Microbiol 49 (4) 744-8, 1985.
88. Sahai A., Xu G., Sandler R .S., Tannen R.L.: Hypoxia-mediated impaired differentiation by LLC-PK1 cells: evidence based on the protein kinase C profile. Kidney Int 42 (5) 1145-52, 1992.
89. Salmi T., Brander P. E.: Computer assisted detection of REM and non-REM sleep for analysis of nocturnal hypoxaemia in patients with ventilatory impairment. Int J Clin Monit Comput 11 (1) 63-70, 1994.
90. Schuster E., Fisher-Williams M.: A rocking bed for poliomyelitis. Lancet 2 1074-6, 1953.
91. Shepard J. W. Jr: Gas exchange and hemodynamics during sleep. Med Clin North Am 69 (6) 1243-1264, 1985.
92. Shepard J. W. Jr: Hypertension, cardiac arrhythmias, myocardial infarction and stroke in relation to obstructive sleep apnea. Clin Chest Med 13 (3) 437-458, Sep 1992.
93. Shimizu T., Hayashi H., Kato S., Hayashi M., Tanabe H., Oda M.: Circulatory collapse and sudden death in respirator-dependent amyotrophic lateral sclerosis. J Neurol Sci 124 (1) 45-55, Jun 1994.
94. Shneerson J. M.: Respiratory pump failure and its treatment. Pacing Clin Electrophysiol 12 (5) 844-6, 1989.
95. Steljes D. G., Kryger M. H., Kirk B. W., Millar T. W.: Sleep in postpolio syndrome. Chest 98 (1) 133-40, 1990.
96. Stephan W., Frehland E.: Rate theory models for ion transport through rigid pores. II. Time-dependent analysis of the single-file model with special reference to oscillatory behavior. J Theor Biol 103 (4) 481-505, 1983.
97. Stout R. D., Suttles J.: Functional characteristics of in vitro generated macrophages: a transient refractory state precedes reinducibility of a spatially restricted, possibly contact-dependent, cytostatic mechanizm. Cell Immunol 105 (1) 33-44, 1987.
98. Sun Z.: [A study of relation between rheumatoid arthritis (RA) and blood stasis-the effect of acupuncture promoting blood circulation to remove blood stasis]. Chen Tzu Yen Chiu 20 (2) 71-5, China 1995.
99. Svenningsen N. W., Wittström C., Hellström-Vestas L.: OSCILLO-oscillating air mattress in neonatal care of very preterm babies. Technology & Health Care 3 43-46, 1995.
100. Sylwanowicz W., Michajlik A., Ramotowski W.: Anatomia i fizjologia człowieka. Państwowy Zakład Wydawnictw Lekarskich, s.526, Warszawa 1985.
101. Szczeklik A.: Katharsis. Państwowy Zakład Wydawnictw Lekarskich, s.126, Warszawa 2002.
102. Taylor W. G., Camalier R. F.: Modulation of epithelial cell proliferation in culture by dissolved oxygen. J Cell Physiol 111 (1) 21-7, 1982.
103. Thomas N., Thornhill R. A.: Stretch activation and nonlinear elasticity of muscle cross-bridges. Biophys J 70 (6) 2807-18, 1996.
104. Tofler G. H., Gebara O. C., Mittleman M.A., Taylor P., Siegel W., Venditti F.J. Jr: Morning peak in ventricular tachyarrhythmias detected by time of implantable cardioverter/defibrillator therapy. The CPI Investigators. Circulation 92 (5) 1203-1208, Sep 1995.
105. Traver G. A., Tyler M. L., Hudson L. D., Sherrill D. L., Quan S. F.: Continuous oscillation: outcome in critically ill patients. J Crit Care 10 (3) 97-103, 1995
106. Trend P. S., Wiles C. M., Spencer G. T., Morgan-Hughes J. A., Lake B. D., Patrick A.D.: Acid maltase deficiency in adults. Diagnosis and management in five cases. Brain 108 (Pt 4) 845-60, Dec 1985.
107. Tsong T. Y.: Molecular recognition and processing of periodic signals in cells: study of activation of membrane ATPases by alternating electric fields. Biochim Biophys Acta 1113 (1) 53-70, 1992.
108. Tuck S. J., Monin P., Duvivier C., May T., Vert P.: Effect of a rocking bed on apnoea of prematurity. J Arch Dis Child 57 (6) 475-7, 1982.
109. van Nimwegen C., Boter J., van Eijnsbergen B.: A method for detecting epileptic seizures. Epilepsja 16 (5) 689-92, 1975.
110. Vanoli E., Adamson P. B., Ba-Lin Pinna G. D., Lazzara R., Orr W. C.: Heart rate variability during specific sleep stages. A comparison of healthy subjects with patients after myocardial infarction. Circul 91 (7) 1918-22, 1995.
111. Volk S.: Zaburzenia snu. Springer PWN, Warszawa 1996.
112. Winter J. H., Neilly J. B., Henderson A. F., Stevenson R.D., Doyle D., Wiles C.M., Spencer G.T.: Life-threatening respiratory failure due to a previously undescribed myopathy. Q J Med 61 (236) 1171-8, 1986.
113. Wong A.: komunikacja osobista, Chang Gung Memorial Hospital, Chang Gung University, Taipei, Taiwan, 2002.
114. World Health Statistics Annual, Geneva 1995.
115. Wulbrand H., Von Zezschwitz G., Bentele K. H.: Submental and diaphragmatic muscle activity during and at resolution of mixed and obstructive apneas and cardiorespiratory arousal in preterm infants. Pediatr Res 38 (3) 298-305, Sep 1995.
116. Xie T. D., Marszalek P., Chen Y. D., Tsong T. Y.: Recognition and processing of randomly fluctuating electric signals by Na,K-ATPase. Biophys J 67 (3) 1247-51, 1994.
117. Yi G., Guo X. H., Reardon M., Gallagher M. M., Hnatkova K., Camm A. J., Malik M.: Circadian variation of the QT interval in patients with sudden cardiac death after myocardial infarction. Am J Cardiol 81 (8) 950-956, 1998.
118. Zieliński J., MacNee W., Wędzicha J., Ambrosino N., Braghiroli A., Doleñsky J., Howard P.: Causes of death in patients with COPD and chronic respiratory failure. Monaldi Arch Chest Dis 52 (1) 3-7, Feb 1997.
119. Zimmerman G. A., Wiseman G. A., Hill, H. R.: Human endothelial cells modulate granulocyte adherence and chemotaxis, J. Immunol., 134(3): 1866-74, 1985.
120. ZZZ - komunikacja osobista z lekarzami Pogotowia Ratunkowego oraz Instytutu Psychiatrii i Neurologii, Warszawa styczeń 1999.
121. Żołądź-Strzelczyk D.: Dziecko w dawnej Polsce, Wydawnictwo Poznańskie, 2002.

INFORMATYK *DEFICYT MOTORYCZNY SNU*	⟷	KLINICYSTA NEUROLOG, FIZJOLOG ANALIZA STANÓW FIZJOLOGICZNYCH W CZASIE SNU

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 99-101.

NEUROMODULACJA BŁĘDNA CZYNNOŚCI MOTORYCZNYCH ŻOŁĄDKA

VAGAL NEUROMODULACTION OF GASTRIC MOTILITY

JACEK SOBOCKI[**], PIOTR THOR[*], GRZEGORZ KRÓLCZYK*, ROMAN M. HERMAN**

[*]*Katedra Patofizjologii, Collegium Medicum Uniwersytetu Jagiellońskiego, Kraków, mfthor@cyf-kr.edu.pl*
[**]*III Katedra Chirurgii Ogólnej, Collegium Medicum Uniwersytetu Jagiellońskiego, Kraków, jsobocki@cm-uj.krakow.pl*

Streszczenie. Autorzy opisują sieć neuronalną żołądka jako zintegrowany system biocybernetyczny koordynujący funkcje wydzielnicze i motoryczne. Wyrazem, jak i podłożem zaburzeń funkcji żołądka jest dezorganizacja i desynchronizacja czynności fali wolnej. Kontrolowana neuromodulacja błędna czynności motorycznych żołądka pozwala w sposób wydolny przywrócić prawidłowe opróżnianie żołądkowe, prowadząc do ustąpienia dolegliwości subiektywnych oraz obiektywnych objawów klinicznych.

Słowa kluczowe: kontrola czynności żołądka, neuromodulacja, porażenie żołądka

Abstract. Authors describe a neuronal net of the stomach as the integrated biocybernetic system aiming to coordinate motility and secretion. The basis of gastric dismotility is disorganization and dissynchronization of the slow wave, which is reflected by ECA. Controlled vagal neuromodulation allows for efficient restoration of gastric emptying leading to the release of subjective symptoms and objective clinical signs.

Key words: gastric function control, neuromodulation, gastroparesis

1. Żołądek jako układ cybernetyczny

Żołądek jest skomplikowanym układem złożonym z jednostek motorycznych i wydzielniczych (wydzielania wewnętrznego i zewnętrznego), zintegrowanych za pomocą systemu setek tysięcy komórek nerwowych. Ta sieć neuronalna żołądka wchodzi w skład większego układu, zwanego „mózgiem trzewnym", którego zadaniem jest sterowanie pracą całego układu pokarmowego. Jakkolwiek układ ten podlega wpływom zewnętrznym, głównie neuronalnym ze strony ośrodkowego układu nerwowego, to wykazuje on wysokiego stopnia autonomię czynnościową. O tej samodzielności świadczy fakt, że nawet całkowite odnerwienie żołądka zwykle zaburza jego pracę tylko w sposób przejściowy.

System sieci neuronalnej żołądka pracuje na zasadzie funkcji zero-jedynkowej dla pojedynczych jednostek neuronalnych, zgodnie z zasadą „wszystko albo nic". Jest więc typowym przykładem biologicznego układu cybernetycznego. Jak każdy układ tego typu, może ulegać modelowaniu za pomocą odpowiednio zaprogramowanego i podłączonego procesora.

2. Porażenie motoryki żołądka

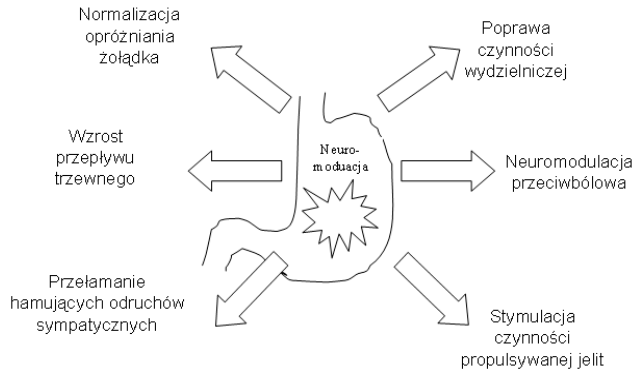
Porażenie żołądka jest trudną do leczenia chorobą, powodującą znaczne obniżenie jakości życia pacjenta. Stan ten jest szczególnie niebezpieczny, gdy wystąpi w okresie pooperacyjnym. Zagraża bowiem podstawowym procesom gojenia, co z kolei może prowadzić do poważnych, zagrażających życiu powikłań. Aktualnie dostępne leki pobudzające perystaltykę są mało efektywne, a poprzez działanie ogólne często wywołują istotne objawy uboczne. Istnieje pewna grupa pacjentów szczególnie narażonych na rozwój długotrwałych zaburzeń porażennych. Czynnikami ryzyka są tu choroby przewlekłe jak cukrzyca, przecięcie nerwów błędnych czy niektóre choroby nowotworowe jak rak trzustki [7], [8], [17].

Istota zaburzeń motorycznych żołądka leży w dezorganizacji i desynchronizacji elektrycznej czynności kontrolującej (*Electric Control Activity* – ECA). Logicznym wydaje się wiec podjęcie próby przywrócenia regularnej czynności elektrycznej za pomocą zewnętrznej stymulacji elektrycznej.

3. Neuromodulacja czynności żołądka

W wielu badaniach przyjęto, że stymulacja elektryczna żołądka wpływa głównie na czynność jego mięśniówki gładkiej. Zarówno w badaniach doświadczalnych, jak i klinicznych stosowano szeroki zakres parametrów stymulacji. Obejmowała ona zakres od niskich częstotliwości zbliżonych do częstotliwości ECA (3-4 c.p.m.) aż do wysokich częstotliwości neuralnych rzędu 30-40 Hz. Wykazano jednak, że najbardziej efektywne są wysokie częstotliwości neuralne [14]. Stymulacja niskoczęstotliwościowa powodowała w jednych badaniach częściowe ustąpienie objawów przy braku poprawy opróżniania żołądkowego [1], [4], podczas gdy w innych badaniach częściową normalizację ECA przy jednoczesnym braku poprawy ze strony objawów [6]. Największe opublikowane dotychczas badania kliniczne ze wszczepionym przewlekłym stymulatorem żołądka wykazały zupełny brak poprawy czynności motorycznej przy jedynie niewielkiej poprawie w obrazie klinicznym [2], [3]. Wyniki te doprowadziły do pewnej stagnacji i zniechęcenia wobec stymulacji elektrycznej jako metody leczenia porażenia żołądka. Sytuację tę zmieniły badania Mintchev'a i wsp. Zaproponowali oni w modelu doświadczalnym wysokoczęstotliwościową wielokanałową stymulację neuralną [9], [10].

Zmieniono więc podejście do stymulacji, jak i spojrzenie na patomechanizmy jej oddziaływania na żołądek. Sytuacja okazała się być bardziej skomplikowana niż przypuszczano. Przyjęto, że to nie mięśniówka, lecz sterujący enteralny układ nerwowy (ENS) powinien być celem stymulacji. Bezpośrednia stymulacja mięśniówki powoduje jedynie przejściową poprawę funkcji motorycznych, odpowiadającą jedynie okresowi samej stymulacji. Natomiast stymulacja neuralna poprzez modulację funkcji sieci nerwowej żołądka pozwala na długotrwałe utrzymywanie się efektu stymulacji, podtrzymywane przez pozytywne zmiany w odruchowej regulacji motoryki. Zaburzenia czynnościowe przewodu pokarmowego w okresie pooperacyjnym obejmują nie tylko jego funkcje motoryczne, ale i wiele innych czynności, jak wydzielanie endo- i egzokrynne itd. [12], [26]. Wpływ neuromodulacji ENS na czynność przewodu pokarmowego przedstawiono schematycznie na rysunku 1. Leczenia nie powinno koncentrować się więc tylko na poprawie czynności mięśniówki gładkiej, lecz całej jednostki czynnościowej, jaką jest żołądek.



Rys. 1. Schemat wpływu neuromodulacji ENS na czynność przewodu pokarmowego

Ponieważ u części pacjentów nie jest konieczna ciągła stymulacja, a stymulacja czasowa powoduje trwałe ustąpienie dolegliwości, stworzono model czasowego rozrusznika żołądka [15], [16]. Łączy o w sobie dwie zasadnicze zalety: małą inwazyjność oraz wielomiejscową dystrybucję prądu stymulującego, co odpowiada modelowi stymulacji wieloelektrodowej Mintchev'a. Wykazano wysoką skuteczność takiego postępowania [13], [18]. Neuromodulacja czynności żołądka powoduje przyspieszenie opróżniania żołądkowego [20], co w konsekwencji prowadzi do ustąpienia objawów. Stymulacja elektryczna powoduje gastrowersję, a przez to wygaszenie fałszywych rozruszników powodujących desynchronizację ECA. Wpływ na funkcję komórek Cajala prowadzi do normalizacji częstotliwości fali wolnej [22], [23]. Równocześnie wzrasta amplituda fali wolnej, co odzwierciedla znaczący wzrost siły skurczów fazowych. Efektywność neuromodulacji zarówno wobec rozszerzonych zabiegów, takich jak operacja sposobem Traverso-Longmire [24], jak również wobec żołądka odnerwionego sympatycznie i/lub parasympatycznie [21], świadczy o bezpośrednim oddziaływaniu stymulacji na ENS.

Pracę żołądka można kontrolować poprzez mikroprocesorową neuromodulację unerwiających go nerwów. Ze względów czynnościowych i anatomicznych do tego celu wybrano nerwy błędne. Efektem działania może być neurostymulacja i/lub neuromodulacja w zależności od doboru parametrów stymulacji. Neuromodulacja błędna może powodować osłabienie lub wzmożenie motoryki żołądka oraz wpływać na jego czynność wydzielniczą [5], [25].

Żołądek, podobnie jak inne trzewia, jest skomplikowanym układem sterowanym przez precyzyjny układ sieci neuronalnej, będącej w swej istocie biologicznym układem cybernetycznym. Połączenie tego biocybernetycznego układu organizmu ożywionego z mogącym je kontrolować sztucznym procesorem objęto nazwą cybernetycznego jelita – *the cybergut* [19].

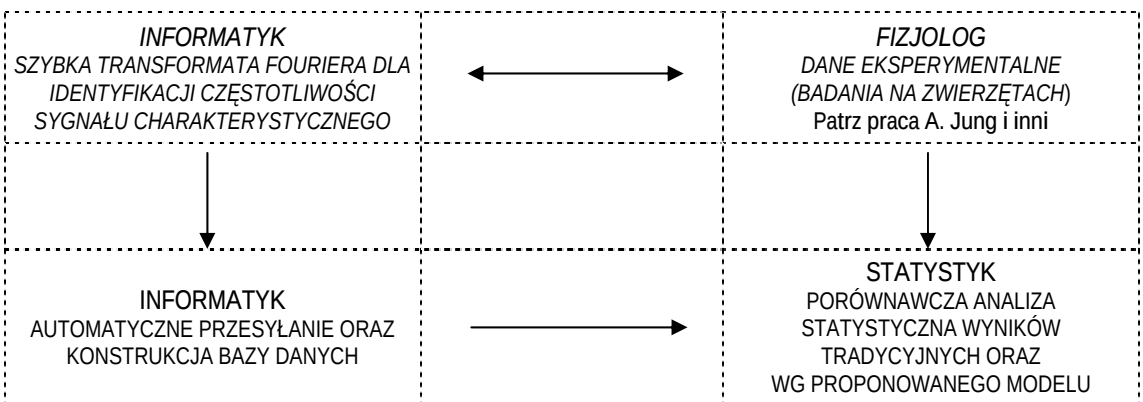
4. Podsumowanie

Nieco parafrazując sytuację, możemy przyrównać żołądek do samochodu. Podobnie jak do procesora sterującego pracą samochodu dołączamy chip, aby zwiększyć moc silnika, podobnie w żołądku do ENS dokładamy chip sterujący stymulacją, aby poprawić motorykę, czyli mechaniczną wydolność systemu napędowego – mięśniówki gładkiej. Efektem jest poprawa opróżniania żołądka.

Literatura cytowana

1. Chen J. D. Z., Lin Z. Y., McCallum R. W.: The role of electrical stimulation in the treatment of gastroparesis, Motility, 34(June): 4-7, 1996.
2. Forster J., Sarosiek I., Delcore R., Lin Z., Raju G. S., McCallum R. W.: Gastric pacing is a new surgical treatmen for gastroparesis, Am. J. Surg., 182: 676-681, 2001.
3. The GEMS Study Group.: Report of a multicenter study on electrical stimulation for the treatment of gastroparesis. Gastroenterol. 112: A735, 1997.
4. Kendall B., McCallum R.: Pacing the human gut: a technology come with age? Am. J. Gastroenterol., 89(3): 452-455, 1994.
5. Królczyk G., Żurowski D., Sobocki J., Słowiaczek M. P., Laskiewicz J., Matyja A., Zaraska Z., Zaraska W., Thor P. J.: Effects of countious microchip (MC) vagal neuromodulation on

gastrintestinal function in rats, Journal of Physiology and Pharmacology, 42(4pt1): 705-715, 2001.
6. Lin Z. Y., McCallum R. W., Schirmer B. D., et al.: Effects of pacing parameters on entraintment of gastric slow waves in patients with gastroparesis, Am. J. Physiol., 274(1pt1): G186-188, 1998.
7. Matyja A., Sobocki J., Thor P. J., Popiela T.: Long-term effects of vagotomy on gastic myoelectric activity, Neurogastroenterology & Motility, 11(4): 274, 1999.
8. Matyja A., Sobocki J., Popiela T., Thor P. J., Huszno B.: Diabetic autonomic dysfunction results in disturbances of gastric myoelectic activity an emptying, Gut, 45(supl V): A259, 1999.
9. Mintchev W. P., Sanmiguel C. P., Otto S. J., Bowes K. L.: Microprocessor controlled movement of liquid gastric content using sequential neural electrical stimulation, Gut., 43(5): 607, 1998.
10. Mintchev W. P., Sanmiguel C. P., Amaris M., Bowes K. L.: Microprocessor-controlled movement of solid gastric content using sequential neural electrical stimulation, Gastroenterol., 118(2): 258, 2000.
11. Nowak Ł., Krolczyk G., Sobocki J., Żurowski D., Thor P.: Gastric pacing is effective in reversing vasoperesin induced gastroparesis, Folia Medica Cracoviensia, 45(1-2), 2004, w druku.
12. Popiela T., Matyja A., Sobocki J., Thor P. J.: Gastrointestinal hormone profile after standard and pylorus-preserving pancreatectomy, Acta Endoscopica Polonica, 3: 135-138, 2000.
13. Popiela T., Sobocki J., Thor P. J., Matyja A.: Temporary gastric pacemaker for the treatment of postoperative gastroparesis,. Polish Journal of Surgery, 74, 10: 886-896, 2002.
14. Sobocki J., Thor P. J., Krolczyk G.: High frequency electrical stimulation of the stomach is more effective than low frequency pacing for the treatment of postoperative functional gastric stasis in human patients, Neuromodulation, 6(4): 254-6, 2003.
15. Sobocki J., Thor P., Matyja A., Wąsowicz P., Popiela T.: Intragastric electrostimulation in the treatment of postoperative gastroparesis – preliminary results. Neurogastroenterology – from the basics to the clinics,. Kulver Academic publishers, 76: 554-9, 1999.
16. Sobocki J., Thor P., Matyja A., Wąsowicz P., Herman R. M.: Stomach electrostimulation - new method of the treating gastroparesis, Folia Medica Cracoviensia, 40(3-4): 63-77, 1999.
17. Sobocki J., Matyja A., Thor P. J., Laskiwicz J.: Gastric motility disturbances in pancreatic cancer patients, Folia Medica Cracoviensia, 40(3-4): 77-83, 1999.
18. Sobocki J., Thor P., Matyja A., Popiela T.: The usefulness of gastric pacing for the treatment of postoperative paralytic ileus, Acta Endoscopica Polonica, 3: 139-141, 2000.
19. Sobocki J., Thor P., Krolczyk G., Uson J., Diaz-Guemes I., Lipinski M.: The cybergut. An experimental study on permanent microchip neuromodulation for control of gut function, Acta Chirurgica Belgica, 102: 68-70, 2002.
20. Sobocki J., Królczyk G., Thor P. J.: Temporary gastric pacemaker improves liquid gastric emptying in gastroparetic patients, Nerodastroenterology & Motility, 5(14): 586, 2002.
21. Sobocki J., Thor P. J.: Effect of gastric electrical stimulation on surgically denervated stomach, Nerodastroenterology & Motility, 5(14): 586, 2002.
22. Sobocki J., Zurowski D., Thor P. J., Matyja A.: Transmucousal electrical stimulation of physiological frequency affects electrical activity of the human stomach, Nerogastroenterology & Motility, 5(14):587, 2002.
23. Sobocki J., Krolczyk G., Laskiewicz J., Thor P. J.: Intragastric electrical stimulation reverses postoperative dysrhythmia. Journal of Physiology and Pharmacology 53 (supl 1): 82, 2002.
24. Sobocki J., Thor P. J., Popiela T.: Electrical stimulation improves gastric transit after Traverso-Longmire procedure - preliminary results. 6th Annual Meeting of European Society of Surgery, Budapest, 2002.
25. Thor P. J., Sobocki J.: Neuromodulation of vago-vagal reflex. Przegląd Lekarski, supl 5: 79-81, 2000.
26. Thor P. J., Matyja A., Popiela T., Szybiński Z., Huszno B., Sobocki J.: Early effects of standard and pylorus preserving pancreatetomy on myoelectric activity and gastric emptying, Hepato-Gastroenterol, 46(27): 1963-7, 1999.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 103-106.

TYMPANOMETRIA WIELOCZĘSTOTLIWOŚCIOWA A CYBERNETYCZNA FAZOWOŚĆ ANALIZATORA SŁUCHOWEGO

MULTIFREQUENCY TYMPANOMETRY AND CYBERNETIC PHASE OF THE HEARING ANALYSER

PAWEŁ TRĄBKA-ZAWICKI, ELŻBIETA REROŃ, JAN TRĄBKA

Katedra i Klinika Otolaryngologii, Collegium Medicum Uniwersytetu Jagiellońskiego, Kraków, ptz@o2.pl

Streszczenie. Autorzy przedstawiają zastosowanie tympanometrii w ocenie stanu funkcjonalnego ucha środkowego. Użycie tonu próbnego o różnych częstotliwościach pozwala na określenie częstotliwości rezonansowej wzmacniacza, jakim jest ucho środkowe. Poza tym przedstawiono występowanie odruchu mięśnia strzemiączkowego przy wykorzystaniu audiometrii impedancyjnej.

Słowa kluczowe: audiometria impedancyjna, częstotliwość rezonansowa ucha środkowego

Abstract. Tympanometry provides useful quantitative information about the functional state of the middle ear system. The use of the probe tone with different frequency allows to estimate the resonant frequency of middle ear. We identify the presence or absence of the stapedial reflex by means of the impedance audiometry.

Key words: impedance audiometry, resonant frequency of middle ear

Dźwięk dociera z otoczenia poprzez ucho zewnętrzne i środkowe do ucha wewnętrznego. Przechodzi on zatem ze środowiska powietrznego (ucho środkowe) do środowiska płynnego (ucho wewnętrzne). Na granicy dwóch środowisk różniących się właściwościami fizycznymi (gęstość, sprężystość, oporność) następuje znaczne rozproszenie energii akustycznej. Odnośnie ucha ludzkiego natężenie dźwięku powinno być 1000-krotnie zmniejszone. Nie dzieje się tak dzięki wzmacniającej funkcji układu transmitującego dźwięk w uchu środkowym. To wzmacniające działanie wynika z faktu, że powierzchnia błony bębenkowej jest ok. 28 razy większa od powierzchni podstawy strzemiączka. Poza tym przy przenoszeniu dźwięku w obrębie ucha środkowego wykorzystywany jest mechanizm dźwigni o różnej długości ramion, które tworzą: błona bębenkowa – rękojeść młoteczka, główka młoteczka – trzon kowadełka, wyrostek soczewkowaty kowadełka – główka strzemiączka [2]. Błona bębenkowa wraz z łańcuchem kosteczek wyrównuje różnicę między niską opornością powietrza a wysokimi wartościami oporności płynu. Dla prawidłowego działania mechanizmu ucha środkowego konieczne jest zachowanie następujących warunków:

- błona bębenkowa utrzymana w całości o niezmienionej strukturze anatomicznej,
- łańcuch kosteczek o zachowanej ciągłości i ruchomości,
- prawidłowa naprzemienna gra obu okienek,
- drożna trąbka słuchowa utrzymująca powietrzność jamy bębenkowej.

Zewnętrzny przewód słuchowy działa w przybliżeniu jak rezonator, którego częstotliwość rezonansowa jest jednocześnie częstotliwością, przy której słuch wykazuje największą czułość. W okolicy tej częstotliwości ciśnienie fali dźwiękowej jest większe w pobliżu błony bębenkowej niż przy wejściu do ucha zewnętrznego. Błona bębenkowa jest elementem ucha, który odwzorowuje zmiany ciśnienia powietrza (fali dźwiękowej) na ruch mechaniczny i energię tego ruchu przekazuje kosteczkom słuchowym. Sprawność przekazywania energii akustycznej zależy od sprężystości błony bębenkowej i od związanego z nią układu kosteczek, masy tego układy i strat energii w procesie przewodzenia. Sprężystość błony jest funkcją jej temperatury, wilgotności, różnicy ciśnień po obu jej stronach oraz stanu naprężenia mięśni i więzadeł jamy bębenkowej.

Tak więc zewnętrzny przewód słuchowy można traktować jako rezonator akustyczny zamknięty membraną o określonych właściwościach sprężystych. Układ taki można scharakteryzować tak zwaną impedancją akustyczną, która wyraża sumę przeszkód na drodze fali akustycznej przy transmisji dźwięku do ucha wewnętrznego [1]. Termin "impedancja" zaczerpnięty został z elektrotechniki, gdzie oznacza oporność układu elektrycznego dla prądu zmiennego. Całkowitą oporność akustyczną (Z) rozpatrywanego układu akustyczno-mechanicznego ucha można przedstawić jako układ drgający, składający się z pewnej masy – M, sprężystości – K, w którym występują również pewne straty energii wywołane siłami tarcia – R (1).

$$Z = \sqrt{R^2 + (X_M - X_K)^2} \quad (1)$$

Z – opór akustyczny (impedancja),
R – opór związany z tarciem (rezystancja),
X_m, X_k – opór związany z masą i ze sztywnością (reaktancja).

Ogólnie, rezystancja określa straty energii związane z przejściem jednej postaci energii w drugą, np. mechanicznej w cieplną lub elektrycznej w mechaniczną itp. W uchu ludzkim na część rezystancyjną impedancji ma wpływ przede wszystkim ślimak oraz straty energii związane z siłami tarcia występującymi w łańcuchu kosteczek [7]. Część urojona impedancji akustycznej – reaktancja określa straty energii związane z magazynowaniem jej w układzie zawierającym element sprężystości i masy. W uchu środkowym element sprężystości jest związany z błoną bębenkową, łańcuchem kosteczek słuchowych i więzadłami śródusznymi, objętością powietrza zawartą w jamie bębenkowej oraz z obu okienkami ślimaka. Z masą układu jest związana jego bezwładność. Układ taki nie będzie posiadał takiej samej ruchomości jak cząsteczki powietrza znajdującego się w ruchu falowym wywołanym przez źródło dźwięku. Ruchy takiego układu są opóźnione w stosunku do ruchu cząstek powietrza. Podstawiając do podanego wcześniej wzoru na impedancję wyrażenia określające reaktancję związaną z masą – $X_m = 2\pi fM$ oraz reaktancję związaną ze sztywnością – $X_k = \frac{K}{2\pi f}$ otrzymujemy zależność (2):

$$Z = \sqrt{R^2 + (2\pi fM - \frac{K}{2\pi f})^2} \quad (2)$$

Ze wzoru wynika, że X_m i X_k są związane z częstotliwością dźwięku. Przy niskich częstotliwościach decydujący wpływ na opór akustyczny ma sztywność, a przy wysokich – masa. Całkowita reaktancja (X_t) wynosi (3):

$$X_t = X_m + (-X_k) \quad (3)$$

X_k ma ujemny znak, gdyż występuje różnica w fazie w stosunku do X_m o 180°. Jeżeli bezwzględne wartości X_m i X_k będą sobie równe, to X_t wyniesie 0 (4). Warunek ten jest spełniony przy częstotliwości rezonansowej (fr) ucha środkowego (5).

$$2\pi fM = \frac{K}{2\pi f} \quad (4)$$

$$fr = \frac{1}{6,28} \times \sqrt{\frac{K}{M}} \quad (5)$$

W podsumowaniu akustycznych cech ucha środkowego można stwierdzić, że gdy częstotliwość dźwięku jest równa częstotliwości ucha środkowego, to wówczas panują najbardziej optymalne warunki dla przenoszenia dźwięku. Poniżej częstotliwości rezonansowej decydujący wpływ na impedancję ma reaktancja związana ze sztywnością – mówimy wówczas, że układ znajduje się pod kontrolą sztywności. Natomiast powyżej częstotliwości rezonansowej układ znajduje się pod kontrolą masy. Ucho środkowe człowieka ma częstotliwość rezonansową w zakresie od 800 Hz do 1200 Hz.

Audiometria impedancyjna zajmuje się oznaczaniem akustycznej oporności lub podatności ucha. Obecnie większość tympanometrów dokonuje pomiarów podatności, zwanej też admitancją. Dla lepszego zrozumienia istoty akustycznej podatności można, dokonując pewnych uproszczeń, przedstawić ją jako podatność komory o sztywnych ścianach, to jest o stałej i niezmiennej pojemności. Jeżeli do układu takiego wtłoczymy powietrze, to ciśnienie wewnątrz komory będzie tym większe, im większa była objętość wtłoczonego powietrza. Można stwierdzić, że podatność C będzie równa stosunkowi objętości powietrza V do ciśnienia P (6).

$$C = \frac{V}{p} \quad \text{jednostka podatności}$$

$$1mho = \frac{1000\,cm^3}{10000\,mmH_2O} \quad (6)$$

Z powyższych założeń wynika, że do komory o większej pojemności możemy wtłoczyć większą objętość powietrza, zachowując niezmienione ciśnienie. Podatność sztywnego układu zależy głównie od jego pojemności. Znając pojemność, można obliczyć podatność.

Ucho zewnętrzne i środkowe nie spełnia warunków układu o sztywnych ścianach. Ściany ucha są w pewnym stopniu elastyczne, a poza tym nie znamy jego pojemności. Stąd też podatność ucha ludzkiego określamy przez porównanie do układu o sztywnych ścianach, mającego taką samą podatność. Wyrażamy ją zatem w objętości tego układu o równoważnej podatności. Dlatego też stosowaną jednostką podatności jest jednostka objętości, np. mililitr.

Tympanometria jest pomiarem podatności układu przenoszącego dźwięk w funkcji zmian ciśnienia powietrza w przewodzie słuchowym zewnętrznym. Jest metodą nieinwazyjną, dostarczającą informacje szybko i obiektywnie. Można zatem stwierdzić, że istotą tympanometrii jest rejestracja zmian podatności błony bębenkowej i związanych z nią struktur ucha środkowego w zależności od sztucznie wywołanych zmian ciśnienia w przewodzie słuchowym zewnętrznym.

Istnieje kilka klasyfikacji tympanogramów, uwzględniających różne parametry. Najbardziej znany i rozpowszechniony jest podział zaproponowany przez Jergera [4], [6] .

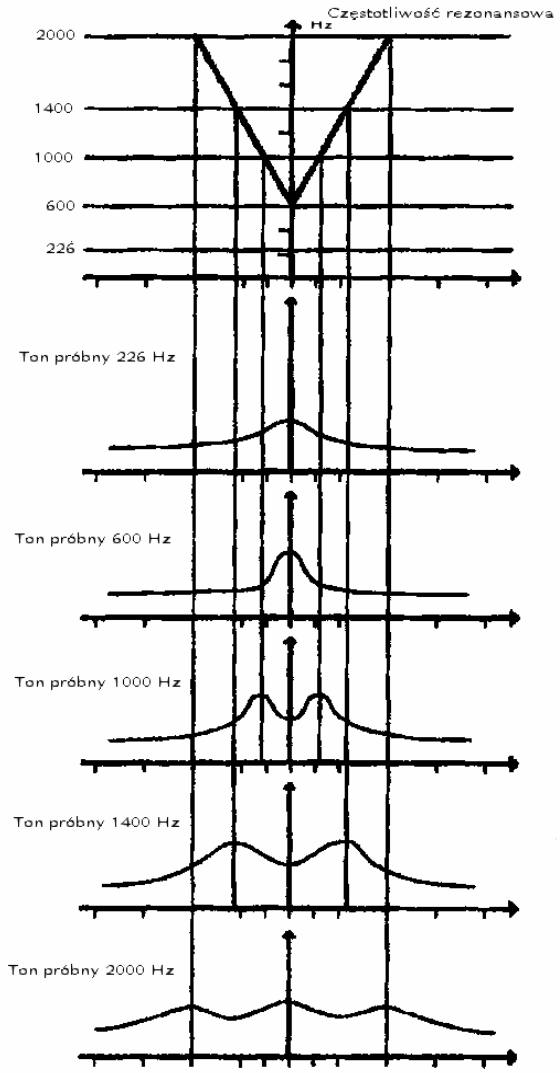
Podzielił on krzywe tympanometryczne w zależności od kształtu i wielkości ciśnienia, w którym przypada maksimum. I tak wyróżniamy następujące typy tympanogramów:

- typ A w kształcie dzwonu, a maksimum krzywej odpowiada ciśnieniu od 0 daPa do ± 100 daPa,
- typ B – krzywa płaska, brak maksimum,
- typ C – kształt dzwonu, maksimum krzywej odpowiada ujemnemu ciśnieniu powyżej –100 daPa.

W 1969 roku Linden [5], po raz pierwszy wykorzystując ton próbny o częstotliwości większej niż 226 Hz, uzyskał tympanogramy dwuszczytowe, które zostały określone jako rys. 1:

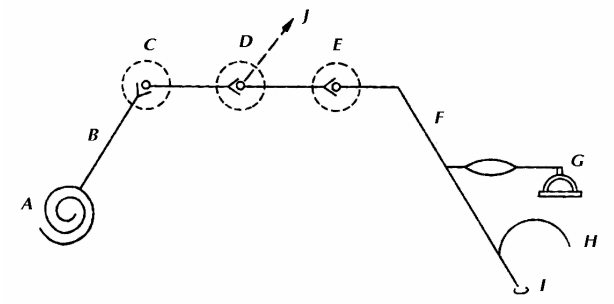
- typ D – tympanogram w kształcie litery M i dwa maksima leżą blisko siebie,
- typ E – tympanogram w kształcie litery M, lecz maksima są bardziej oddalone.

Znany jest fakt, że tympanogram posiada maksimum, jeżeli ciśnienie po obu stronach błony bębenkowej jest równe. W warunkach tych układ jest najbliższy częstotliwości rezonansowej ucha środkowego. Jeżeli częstotliwość tonu próbnego jest poniżej częstotliwości rezonansowej, otrzymujemy tympanogram w kształcie dzwonu. Jeżeli z kolei częstotliwość tonu próbnego jest większa od częstotliwości rezonansowej to tympanogram ma kształt litery M (rys. 1).



Rys. 1. Częstotliwości rezonansowe w funkcji zmian ciśnienia w przewodzie słuchowym zewnętrznym oraz odpowiadające tympanogramy wg Cotrona [3]

Przy ciśnieniu równym po obu stronach błony bębenkowej rejestrujemy odruch mięśnia strzemiączkowego, to jest zmianę podatności ucha środkowego powstałą w wyniku skurczu mięśnia strzemiączkowego jako reakcję na dźwięk o natężeniu znacznie przekraczającym (70-90 dB) próg słuchowy w audiometrii tonalnej. Odruch strzemiączkowy, niezależnie od sposobu wywołania, występuje symetrycznie w obu uszach. Współczesne metody badania pozwalają dokonywać pomiaru progu odruchu na drodze pobudzenia kontralateralnego – pobudzenie ucha przeciwstronnego do ucha testowanego lub na drodze ipsilateralnej – poprzez pobudzenie ucha testowanego. Łuk odruchowy przy pobudzeniu kontralateralnym składa się z następujących neuronów: zwój spiralny, jądro ślimakowe brzuszne, jądro oliwkowe górne przyśrodkowe, jądro nerwu twarzowego po stronie przeciwnej. Natomiast przy badaniu ipsilateralnym łuk odruchowy kończy się w jądrze nerwu twarzowego po tej samej stronie (rys. 2). Czas utajenia odruchu strzemiączkowego u człowieka przy kontralateralnym podawaniu bodźca waha się od 150 ms przy natężeniach okołoprogowych do 25-35 ms przy dużych natężeniach dźwięku.



Rys. 2. Schemat drogi nerwowej dla odruchu strzemiączkowego. A – ślimak, B – nerw ślimakowy, C – jądro nerwu słuchowego, D – jądro oliwy górnej, E – jądro ruchowe nerwu twarzowego, F – nerw twarzowy, G – mięsień strzemiączkowy i strzemiączko, H – struna bębenkowa, I – otwór rylcowo-sutkowy, J – połączenie z płatem skroniowym

Istnieje szereg hipotez wyjaśniających czynnościową rolę mięśnia strzemiączkowego, wśród których na podkreślenie zasługują:

- ochrona ucha wewnętrznego, szczególnie w zakresie 0,5-4 kHz przed ekspozycją na nadmierne natężenie dźwięku,
- poszerzenie zakresu dynamicznego ucha w odniesieniu do natężenia,
- zmniejszenie efektu rezonansowego ucha środkowego,
- zmiana ciśnienia płynów w ślimaku.

W odbiorczym aparacie analizatora słuchowego charakterystyki parametrów fali akustycznej ulegają trzem kolejnym zmianom fazowym, spowodowanym różnicami budowy anatomicznej ośrodków przewodzących: powietrznego, płynowego i ciała stałego. Różna oporność akustyczna tych ośrodków zmienia intensywność (amplitudę) i wysokość (częstotliwość) oraz czas trwania dźwięku. Aby zapewnić zjawisku receptorowemu prawidłowy przebieg, każdy etap środowiskowych zmian fazowych musi być poddany ścisłej regulacji cybernetycznej.

Zmiana fazowa polega na zachowaniu symetrycznej bądź (znacznie częściej) asymetrycznej relacji między spójnością (koherencją) oraz dyfuzją (dekoherencją) energii bodźca słuchowego. Największą dyfuzyjność posiada powietrze; stąd powstaje echo, które zanika, gdy wartość koherencji przyjmuje wartość zerową. Z kolei najmniejszą rozpraszalnością dysponuje ciało stałe.

W czasie transmisji bodźca z zewnętrznego przewodu słuchowego do narządu Cortiego występuje nieustanna ale regulowana cybernetycznie gra między koherencją a dyfuzją, wyrażająca się ich asymetrią, rzadko asymetrią.

Błona bębenkowa na granicy ucha zewnętrznego i środkowego działa jak wzmacniacz – koncentruje energię fonów unoszonych przez falę akustyczną, zwiększając jej spójność. Maksimum spójności dźwięku zapewnia łańcuch kosteczek słuchowych. W następnej zmianie fazowej dzieje się coś zgoła

przeciwnego: aby komórki słuchowe mogły przeanalizować skoncentrowaną, wysoką falę akustyczną pod względem intensywności, czasu trwania oraz widma częstotliwościowego, fala ta musi być wtórnie przez system przestrzeni płynowych ślimaka rozciągnięta, czyli ulec dyfuzji kosztem koherencji, która automatycznie maleje. Aby skompensować ową częściową upływność energii, analizator musi oprzeć swą pracę na optymalnym utrzymaniu stałej charakterystyki częstotliwościowej, co z kolei gwarantuje najmniejszą stratę zawartości informacyjnej transmitowanej drogami neutralnymi, czyli parametru najważniejszego dla rozumienia mowy na poziomie jądra analizatora słuchowego. Stąd właśnie regulacyjne impulsy drogami odśrodkowymi analizatora docierają aż na obwód, unerwiając mięśnie napinacza błony bębenkowej oraz strzemiączka, a także poprzez pęczek Rasmussena do błony podstawnej ślimaka. Tak zamyka się obwód kontroli cybernetycznej, regulującej sprawne działanie analizatora słuchowego, które na pierwszym etapie zmian fazowych badamy za pomocą tympanometrii wieloczęstotliwościowej. Użycie jednej częstotliwości tonu próbnego dałoby zbyt wycinkowy obraz fazowości zmian fali akustycznej przy przechodzeniu przez struktury obwodowej części narządu słuchu.

W przedstawionej tak pracy wykorzystano tony próbne o częstotliwości 226 Hz, 660 Hz i 1000 Hz, co wiązało się z próbą wyznaczenia częstotliwości rezonansowej ucha środkowego, przy której istnieją najbardziej optymalne warunki dla przenoszenia dźwięku. Różne stany chorobowe deformują fazy, które mogą przybierać postać ciągłą, skokową oraz pośrednią. Regulacja cybernetyczna fazy wyraża się wyraźnymi zmianami amplitudy przy zachowaniu stałej częstotliwości, począwszy od niedosłuchu czy nadwrażliwości. Ubytki słuchu mogą być ciągłe lub tylko fragmentaryczne, dające na krzywej różnej głębokości załamki.

Literatura cytowana

1. Bekesy G.: Zur Theorie des Horens bei der Schallaufnahme durch Knochenleitung, Ann Physik, 13(11): 1932.
2. Bystrzanowska T.: Audiologia kliniczna, PZWL, 1978.
3. Cotrona V.: Understanding Impedance Measurement, Oticon A/S, 1989.
4. Jereger J.: Clinical experience with impedance audiometry, Arch Otolaryngol, 92(311), 1970.
5. Linden G., Peterson J., Björkman G.: Tympanometry, Arch Otolaryngol, 92(248), 1970.
6. Onusko E.: Tympanometry, Am Fam Physician, 70(9): 1713-20, 2004.
7. Zwisłocki J.: Analysis of middle ear function. Part I. Input impedance, J. Acoust. Soc. Am., 34,1514, 1996.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 107-110.

TRAINING SET SIZE IN ENSEMBLE FEATURE SELECTION FOR CLINICAL PROTEOMICS

ANALIZA ROZMIARU ZBIORU TRENINGOWEGO W ZESPOŁOWEJ METODZIE SELEKCJI CECH NA POTRZEBY PROTEOMIKI KLINICZNEJ

TOMASZ ARODŹ

Katedra Informatyki Akademii Górniczo-Hutniczej, Kraków, arodz@agh.edu.pl

Abstract. The SELDI-TOF mass spectrometry has been recently shown to be useful in diagnosis of a range of cancer types. In the procedure, samples of body fluids are subject to proteomic analysis using mass spectrometry, resulting in highly dimensional fingerprints. The fingerprints gathered from a set of cancer and control patients allow for creation of a machine learning-based method for diagnosing cancer. In this paper, we analyse the effects of the number of examples in the training set used for constructing a classifier distinguishing blood samples from normal and ovarian cancer patients. We employ a version of our FLD boosting classifier, extended to include a feature selection algorithm within a single machine-learning framework. We show that when a particular type of protein chip is used in SELDI-TOF-MS analysis, the training set containing samples from only ca. 30-40 patients is suitable for creating a classifier exhibiting ca. 95% accuracy, sensitivity and specificity to ovarian cancer. The feature selection procedure incorporated into the classifier reduces to 2 the number of mass/charge values used for discrimination.

Key words: ensemble feature selection, boosting, SELDI-TOF mass spectrometry, clinical proteomics

Streszczenie. Spektrometria masowa typu SELDI-TOF została w ostatnich latach zastosowana do diagnostyki nowotworów. W tym celu wykorzystywane są próbki płynów ustrojowych, które poddawane są analizie proteomicznej z wykorzystaniem spektrometru masowego. W efekcie uzyskiwany jest wielowymiarowy opis pobranej próbki. Dla zbioru pacjentów z oraz bez nowotworu możliwe jest stworzenie metody diagnostycznej opartej na metodach uczenia maszynowego. W niniejszym artykule analizujemy efekt wielkości zbioru trenującego używanego do uczenia klasyfikatora rozróżniającego próbki krwi od pacjentów zdrowych i z obecnym nowotworem jajnika. Użyty klasyfikator bazuje na metodzie zespołowej typu boosting używającej reguły decyzyjnej Fishera. Klasyfikator ten został rozszerzony o metodę selekcji cech. W artykule wykazano, iż dla odpowiedniego typu mikromacierzy białkowej użytej w metodzie spektrometrycznej SELDI-TOF, zbiór treningowy zawierający ok. 30-40 próbek pozwala na stworzenie klasyfikatora wykazującego 95%-ową dokładność klasyfikacji. Zintegrowana z klasyfikatorem metoda selekcji cech pozwala na skuteczną klasyfikację przy użyciu tylko 2 cech z widma spektrometrycznego.

Słowa kluczowe: zespołowe metody selekcji cech, klasyfikatory zespołowe, spektrometria masowa typu SELDI-TOF, proteomika diagnostyczna

1. Introduction

The proteins from body fluids have been used for diagnosis of cancer for a long time. Examples include e.g. prostate-specific antigen (PSA) protein used in diagnosis of prostate cancer. However, until recently, the level of single protein was usually used as a biomarker. As a single protein is rarely a definitive biomarker of cancer, the usefulness of such diagnostic methods was limited. In particular, the determination of the particular protein to be used as a biomarker was troublesome.

Recently, fast method of detecting the level of abundance of a large range of proteins of diverse molecular weights was introduced [11] and gained much interest [10]. The method is based on *surface-enhanced laser desorption/ionization- time of flight mass spectrometry* (SELDI-TOF MS [12]) and is also known as clinical or diagnostic proteomics.

As the results of SELDI-TOF examination contain abundance for thousands of proteins, manual processing of the information is not feasible. The automatic detection of patterns associated with a given illness is essential. Therefore, computational approaches are crucial to the success of the method.

Up to date, the time-of-flight mass spectrometry has been used in diagnosis of ovarian [11], breast [9], prostate [15], bladder [18], pancreatic 6] and other cancer. Apart from diagnosis of cancer, the proteomic mass spectrometry has been found useful in other areas of medicine, e.g. in toxicology [13].

Several datasets of the proteomic mass spectrometry experimental data are available publicly at the website of the Clinical Proteomics Program Databank. (*http://home.ccr.cancer.gov/ncifdaproteomics/*). These include data for ovarian cancer from experiments described in [11], prostate cancer [14], pancreatic cancer [6] and toxicology experiments [13]. Another dataset for prostate cancer, used in an experiment described in [1], is available at the Virginia Prostate Center site (*http://www.evms.edu/vpc/seldi/*).

The proteomics data are presented in Fig. 1. The abundance of ions is measured as a function of their mass to charge ratio (*m*/*z*). The measured spectrum usually spreads over a large range of masses, e.g. from 2 kDa to 40 kDa in [15]. Combined with high resolution of the *m*/*z* ratio, the proteomic data for a single patient contain e.g. 15000 *m*/*z* points [11].

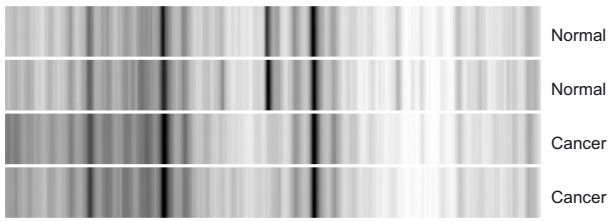


Fig. 1. Proteomic mass spectrometry data for two cancer and two non-cancer samples. Abundance of molecules as a function of their mass to charge ratio. Only parts of the spectra are shown

Several pattern recognition approaches have been used to process the proteomic data in search for biomarker patterns. The authors of an ovarian cancer study [11] follow a two-step procedure. First, a genetic algorithm was used to reduce significantly the number of features. Then, an unsupervised learning algorithm based on Kohonen self-organising maps was used.

In Ref. [15], the dimensionality of the data was reduced as the first step. This was done using peak detection and alignment, as well as estimates of discrimination ability of individual peaks. Then, decision trees and decision stumps were used with boosting to obtain final diagnosis of prostate cancer.

In Ref. [16], statistical feature selection, based on the difference between mean values of spectra for healthy and cancer samples was used. After selecting points in the spectrum with the highest difference, neural networks, support vector machines and decision trees were evaluated with ovarian cancer data.

In Ref. [8], the ovarian cancer proteomic data are used to evaluate several statistical approaches to feature selection. These statistical methods include the χ^2 test, the Correlation-based Feature Selection and the *t* statistics. Following, the selected features were classified with one of five different classification methods: k-nearest neighbour, decision trees, naive Bayes or support vector machines to evaluate the effectiveness of the feature selection.

All of the above approaches include two phases. First, the feature selection algorithm is executed. Then, the machine-learning algorithm is applied, with only the selected features used. In this paper we use a system, which combines feature selection and classifier training into a single algorithm. The method involves using a group of linear classifiers and is based on our previous study of ensembles of Fisher linear discriminants [2].

The goal of the paper is to determine the minimum size of training set required for successful construction of a robust classifier for cancer diagnosis. We use the ovarian cancer proteomics data [11] from the Clinical Proteomics Program Databank.

The rest of the paper is arranged as follows. In Section 2, the outline of the classification method used in experiments is presented. Next, in Section 3, the experimental setup is described and the results are discussed. Finally, Section 4 summarises the conclusions of the paper.

2. The Classification Method for Diagnosis of Ovarian Cancer

Ensemble machine learning methods are based on the idea of aggregating results of a diverse group of lower-level learning methods to boost the accuracy and the reliability of learning. In particular, a group of methods known as multi-classifier systems is centred on supervised learning problem. These methods train a set of individual classifiers, such as e.g. decision trees. For an unknown sample, the decisions of the individual classifiers are combined, e.g. by averaging, to form the final classification decision.

The main variants of ensemble classifier methods include boosting [5], bagging [4] and random subspace method [7]. In this paper, we use a classification method, which takes its roots from AdaBoost [5], the most widely used boosting algorithm. In this approach, the training of the final classifier is done sequentially in rounds. In each round, an instance of a simple classifier, e.g. a decision tree, is trained and incorporated into the ensemble of previously trained simple classifiers. In order to decrease the classification error of the ensemble in comparison to the errors of individual classifiers constituting it, the simple classifiers trained in each round have to yield diverse decision boundaries.

To achieve diversity among simple classifiers, each example from the training set is given a weight. During the course of training of the successive simple classifiers, the examples, which are correctly classified, have their weights reduced. On the other hand, the weights of the examples misclassified by the simple classifiers are increased. Each of the simple classifiers is trained to minimize the weighted error over the training set. Thus, the examples with higher weights gain more attention from the simple classifier. This procedure allows for shifting the attention of the classifier to the examples, which are hardest to recognize, at the cost of misclassifying the examples correctly recognized by the previously trained simple classifiers.

The nature of the simple classifier is of major importance in the boosting scheme. Approaches range from very simple decision rules, such as selecting a single feature from the feature space and using a threshold [17], to highly elaborate machine learning tools, such as decision trees [5]. In our previous studies [2], [3], we have used a well-known Fisher Linear Discriminant decision rule as a simple classifier. To increase the accuracy of the ensemble, each FLD was trained using a different, random subset of features from the feature set. This resulted in increasing the diversity of the ensemble and, in consequence, in decreasing the classification error.

To apply the method to the task of cancer recognition in highly dimensional mass-spectra, we have introduced the best feature selection instead of random selection of subset of features. In the first round, the single feature that leads to the lowest error on the training set is chosen. In subsequent rounds, the feature subset is extended to include new features from the original feature space. The feature, which by itself yields the minimal weighted training error, is chosen to be added to the subset.

However, in order to minimize the number of selected features, a new feature is not chosen in each round of

boosting. As long as the overall error on the training set decreases in subsequent rounds, no additional feature is added to the subset. However, when for a specified number of rounds the combined classifier trained so far does not reduce the training error, new feature is added. In this way, the new feature, carrying new information, is added only when the information contained in the current subset of features is not sufficient to further reduce the training error.

3. Experimental setup and analysis of results

In our study, we have used three datasets from the Clinical Proteomics Data Bank. These include the "Lancet Ovarian Data 2/16/2" dataset, prepared using the H4 protein chip, and the "4/3/2 Ovarian Study" and "Ovarian Dataset 8/7/2" dataset prepared using the WCX2 protein chip. The details on the number of cancer and normal examples in each dataset are summarized in Tab. 1.

Tab. 1. Information on the ovarian cancer datasets obtained from the Clinical Proteomics Data Bank used in this study

Dataset	Number of cancer examples	Number of normal examples	Number of features
2/16/2	100	100	15154
4/3/2	100	100	15154
8/7/2	162	91	15154

For each of the examined datasets, the estimation of classification accuracies for different sizes of the training set was carried out. The sizes of the training set ranged from 10 to 100 examples, with the step of 3 examples. For each of the training set sizes, the training and testing has been repeated 10 times, with each run using a different, randomly generated subset of the examples from the dataset used in training. The results were then averaged.

Tab. 2. Results of the tests for 100 examples in the training set

Data	Accuracy [%]	Sensitivity [%]	Specificity [%]	Number of selected features
2/16/2	81	81	80	10,2
4/3/2	93	92	94	2,0
8/7/2	98	98	99	2,1

For all three datasets, the results for the maximum number of examples, equal to 100, was summarized in Tab. 2. The accuracy on the testing set, as well as sensitivity and specificity are given.

The results show that the results of the classifier-based diagnostic method vary with dataset. In particular, the accuracy is lower and the number of selected features is significantly higher for the "2/16/2" dataset. This may be attributed to the H4 protein chip used in this particular dataset, which was different from the WCX2 chip used in the other two datasets.

The dependency between the size of the training set and the classifier accuracy for the datasets is depicted in Fig. 2-4. Again, the results for the "2/16/2" dataset are different than for "4/3/2" and "8/7/2" datasets. For this dataset, the increase in accuracy is present in the whole range of training set sizes. On the other hand, for the "4/3/2" and "8/7/2" datasets, the accuracy reaches values close to the maximum for ca. 30-40 examples.

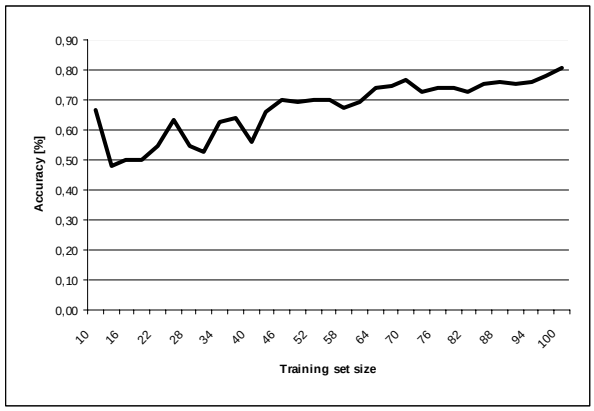


Fig. 2. Accuracy as a function of the training set size for the "2/16/2" dataset

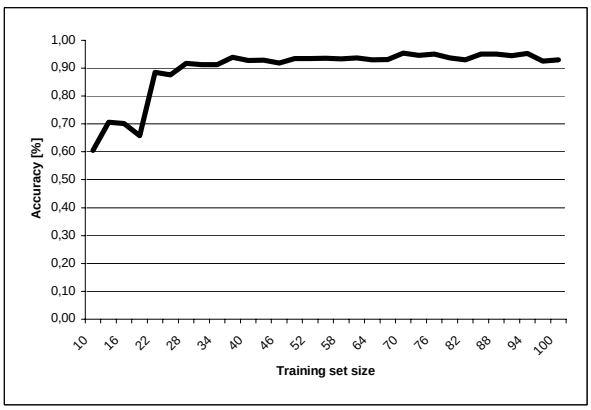


Fig. 3. Accuracy as a function of the training set size for the "4/3/2" dataset

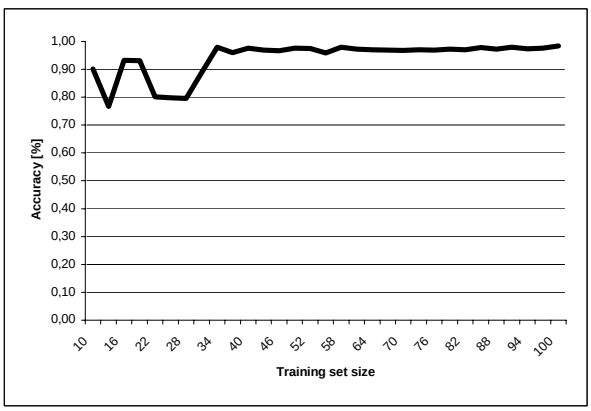


Fig. 4. Accuracy as a function of the training set size for the "8/7/2" dataset

The analysis of the number of selected features for each of the training set sizes shows that the accuracy of the classifiers depends on feature selection. For the "2/16/2" dataset (Fig. 5), the number of features increases throughout the whole range of training set sizes, correlating with the increase in accuracy. On the other hand, for the "8/7/2" dataset (Fig. 6), and similarly for the "4/3/2" dataset, the average number of selected features is 1 for the training set smaller than ca. 30-40 examples. This correlates with the

lower accuracy of the classifier in that range of training set sizes and with the observed instability in accuracy. As the number of training examples exceeds 40, two features are constantly selected, and the accuracy reaches maximum.

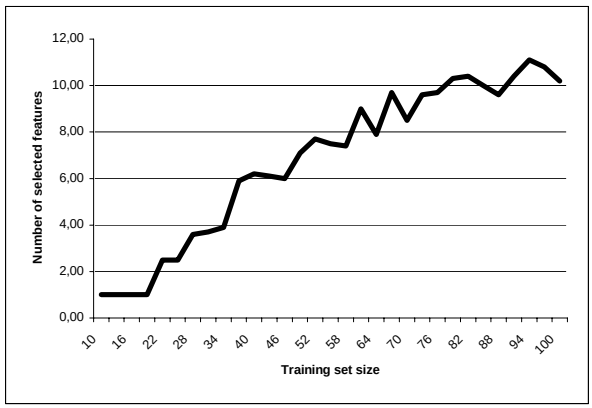


Fig. 5. Number of selected features as a function of the training set size for the "2/16/2" dataset

Fig. 6. Number of selected features as a function of the training set size for the "8/7/2" dataset

4. Conclusions

In this paper, we have used the machine learning method for the diagnosis of ovarian cancer based on mass spectrograms of blood samples. We have evaluated the ensemble feature selection method, which is an extension of our FLD boosting algorithm [2]. Using three publicly available mass spectra datasets, we have shown that, in order to achieve the accuracy, sensitivity and specificity close to 95%, approximately 30-40 examples have to be gathered for the training of the classifier. For WCX2 protein chip, such results can be obtained using only two *m/z* values from the whole 15154 values constituting the mass spectrogram. In the case of older H4 protein chip used instead of WCX2 chip in the data gathering procedure, the results show lower accuracy and higher number of features selected, despite larger training set sizes.

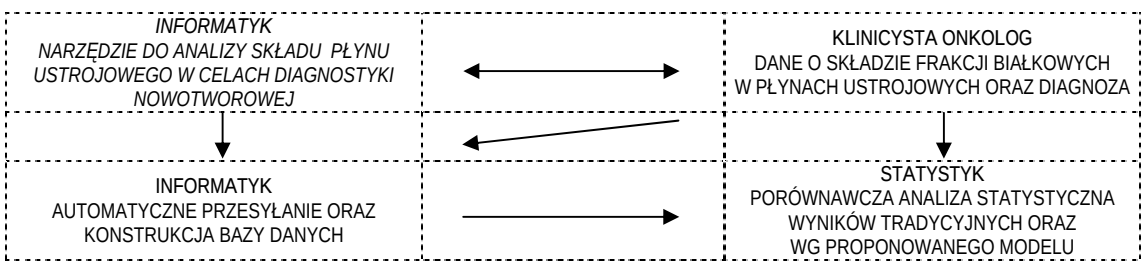
Acknowledgements

I would like to thank my Ph.D. supervisor Prof. Dr Witold Dzwinel for scientific guidance and inspiring discussions. I also gratefully acknowledge the medical expertise of Dr Tadeusz J. Popiela from Collegium Medicum, Jagiellonian University and expertise in mass spectrometry of Prof. Dr Jerzy Silberring from the Faculty of Chemistry, Jagiellonian University.

The work is supported from the Polish State Committee for Scientific Research (KBN) Grant No. 3T11C05926.

Bibliography

1. Adam B. L., Qu Y. et al.: Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. Cancer Research 62: 3609-14, 2002.
2. Arodź T.: Boosting the Fisher Linear Discriminant with Random Feature Subsets. In: Proc. CORES'05, Advances in Soft Computing. Springer-Verlag, 2005 (in press).
3. Arodź T.: Margin-based Diversity Measures for Ensemble Classifiers. In: Proc. CORES'05, Advances in Soft Computing. Springer-Verlag, 2005 (in press).
4. Breiman L.: Bagging predictors. Machine Learning 24(2): 123-140, 1996.
5. Freund Y., Schapire R.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55(1): 119-139, 1997.
6. Hingorani S., Petricoin E. et al.: Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. Cancer Cell 4(6): 437-450, 2003.
7. Ho T.K.: Random decision forests. In: Proc. 3rd Int'l Conference on Document Analysis and Recognition. Montreal, Canada, 1995.
8. Li J., Liu H., Wong L.: A comparative study on feature selection and classification methods using gene and protein expression profiles. In: Genome Informatics. Universal Academy Press, 2002.
9. Li J., Zhang Z., Rosenzweig J. et al.: Proteomics and bioinformatics approaches for identification of serum biomarkers to detect breast cancer. Clin. Chem. 48(8), 2002.
10. Liotta L. A., Ferrari, M., Petricoin, E.: Clinical proteomics: Written in blood. Nature 425: 905, 2003.
11. Petricoin E., Ardekani, A., et al.: Use of proteomic patterns in serum to identify ovarian cancer. Lancet 359(9306): 572-577, 2002.
12. Petricoin E., Liotta L.: SELDI-TOF-based serum proteomic pattern diagnostics for early detection of cancer. Current Opinion in Biotechnology 15: 24-30, 2004.
13. Petricoin E., Rajapaske et al.: Toxicoproteomics: Serum proteomic pattern diagnostics for early detection of drug induced cardiac toxicities and cardioprotection. Journal of Toxicologic Pathology 32: 1-9, 2004.
14. Petricoin E., Ornstein D., Paweletz C.: Serum proteomic patterns for detection of prostate cancer. Journal of the National Cancer Institute 95(6): 489-489, 2003.
15. Qu Y., Adam B. L. et al.: Boosted decision tree analysis of SELDI mass spectral serum profiles discriminates prostate cancer from noncancer patients. Clinical Chemistry 48(10):1835-1843, 2002.
16. Tang H., Mukomel Y., Fink E.: Diagnosis of ovarian cancer based on mass spectra of blood samples. In: Proc. IEEE Int'l Conf. Systems, Man and Cybernetics, 2004.
17. Viola P., Jones M. J.: Robust real-time face detection. Int. J. Comput. Vision 57:137-154, 2004.
18. Vlahou A., Schellhammer P. F. et al.: Development of a novel proteomic approach for the detection of transitional cell carcinoma of the bladder in urine. American Journal of Pathology 158: 1491-1502, 2000.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 111-112.

WYKORZYSTANIE BIOINFORMATYKI W BIOTECHNOLOGII

THE APPLICATION OF BIOINFORMATICS IN BIOTECHNOLOGY

ANNA DOBROWOLSKA

Katedra Technologii Fermentacji i Mikrobiologii Technicznej Akademii Rolniczej w Krakowie, annadobrow@interia.pl

Streszczenie. Dynamiczny rozwój bioinformatyki umożliwia zarówno planowanie badań biologicznych, jak i ich wielokierunkową analizę. Wykorzystanie komputerowych baz danych oraz specjalistycznych programów komputerowych stanowi obecnie podstawę rozwoju wielu współczesnych dziedzin biologii, w tym m.in. genetyki molekularnej. Bioinformatyka stanowi pomocne narzędzie w poznaniu struktury, mechanizmów działania i współdziałania genów.
W opisywanej pracy sekwencje nukleotydowe genów plazmidu binarnego pCAMBIA1303, dostępne w bazie danych (www.ncbi.nlm.nih.gov), stanowiły podstawę dla zaprojektowania pary starterów będących matrycą dla syntezy genu fuzyjnego, kodującego stafylokinazę (*sak*) pochodzenia prokariotycznego. Startery zostały wykorzystane do amplifikacji genu fuzyjnego CaMV::*sak–gusA–mgfp* w metodzie PCR. Produkty amplifikacji analizowano, wykorzystując program komputerowy Analysis (Gerany Soft Imaging System GmbH). Wektor pCAMBIA1303*sak* zastał wykorzystany do transformacji siewek tytoniu.

Słowa kluczowe: genetyka, pCAMBIA, PCR

Abstract. The dynamic development of bioinformatics enables both the planning of biological research and its multidirected analysis. The application of computer databases and specialised computer programmes is the basis for the development of many contemporary biological disciplines, including molecular genetics. Bioinformatics seems to be the tool neccesary to study the structure of genes, their mechanisms of operation, and their cooperation.
In this research, the nucleotide sequence of the binary plazmid pCAMBIA1303 genes available in the database, (www.ncbi.nlm.nih.gov) were the basis to design the pair of primers for the synthesis of desirable fusion gene construct with prokaryotic stafylokinase gene in PCR method. The primers were used for the amplification of fusion gene construct CaMV::*sak–gusA–mgfp*. The products of amplification were analyzed by Analysis software (Gerany Soft Imaging System GmbH). The pCAMBIA1303*sak* vector was used for tobacco seedling transformation.

Key words: genetics, pCAMBIA, PCR

1. Wstęp

Rośliny służą człowiekowi od zarania dziejów jako pożywienie i źródło substancji leczniczych dla przemysłu spożywczego i farmaceutycznego. Jednym z najważniejszych obszarów biotechnologii roślin jest produkcja metabolitów wtórnych *in vivo* oraz *in vitro*. Są to związki, zwykle o złożonej struturze chemicznej, które nie odgrywają większej roli w metabolizmie podstawowym, natomiast mają znaczenie w procesach przystosowania się roślin do przeżycia w środowisku. Wysoki koszt kultur roślinnych i separacji produktu powoduje, że w polu zainteresowania znajdują się przede wszystkim metabolity o właściwościach terapeutycznych [4].

Obecnie około jedna czwarta dostępnych leków jest pochodzenia roślinnego, a nowoczesna biotechnologia stwarza możliwości szerokiego wykorzystania własności leczniczych roślin [3]. Zamiast kosztownego i ryzykownego ekologicznie dostosowania warunków środowiska do potrzeb rośliny, dąży się do jej odpowiedniego ukształtowania genetycznego. Tak fundamentalna zmiana była możliwa tylko dzięki ogromnemu wzrostowi wiedzy w zakresie biologii, a przede wszystkim biologii molekularnej, w tym genetyki. Dało to możliwość kształtowania cech organizmów niezależnie od dystansu filogenetycznego dzielącego dawcę i biorcę genu [4].

Bioinformatyka, dynamicznie rozwijająca się dziedzina wiedzy, stanowi pomocne narzędzie w uzyskaniu odpowiedzi na pytania biologiczne, co wspomaga planowanie eksperymentów laboratoryjnych [1]. Wraz ze wzrostem dostępnych baz danych o sekwencjach i strukturach, biologiczna wiedza komputerowa zaczyna odgrywać coraz istotniejszą rolę w biotechnologii. Sekwencje nukleotydowe genów plazmidu binarnego pCAMBIA1303, dostępne w bazie danych (www.ncbi.nlm.nih.gov) oraz program komputerowy Analysis (Gerany Soft Imaging GmbH) okazały się pomocne w konstrukcji oraz analizie genu fuzyjnego CaMV::*sak-gusA-mgfp.*

2. Materiały

W pracy wykorzystano wektor plazmidowy pTrxFus*sak* [2], wektor plazmidowy pCAMBIA1303, szczep *E. coli* NM522 oraz szczep *A. tumefaciens* AGLI Rf^r.

3. Metody

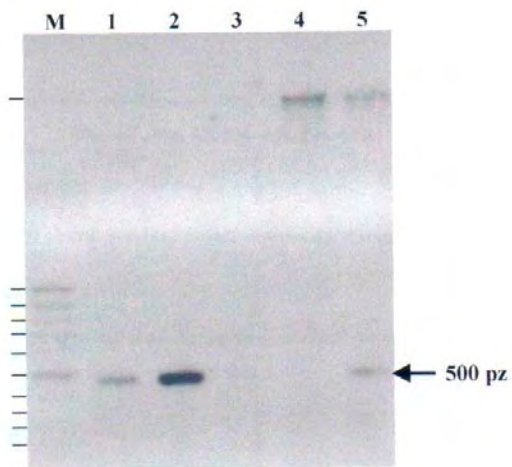
Sekwencje starterowe dla reakcji PCR zostały zaprojektowane na podstawie znanej sekwencji nukleotydowej genu stafylokinazy oraz sekwencji nukleotydowych odpowiednich miejsc restrykcyjnych na plazmidzie pCAMBIA1303 [2], amplifikację prowadzono według zaleceń producenta (MBI Fermentas) [2]. Analizy elektroforogramów prowadzono z wykorzystaniem programu komputerowego Analysis (Gerany Soft Imaging System GmbH).

4. Wyniki

Jednym z obecnie stosowanych rodzajów wektorów binarnych są wektory zawierające kasetę ekspresyjną umożliwiającą konstrukcję fuzji genowych z genami reporterowymi. Geny fuzyjne, dzięki łatwo wykrywalnemu produktowi genu reporterowego, umożliwiają czasoprzestrzenną lokalizację ekspresji badanego genu heterologicznego w komórkach roślinnych.

Obecnie z powodzeniem stosuje się geny reporterowe kodujące lucyferazę, ß-glukuronidazę czy zielono fluoryzujące białko GFP. Szczególnie często, ze względu na łatwość i wiarygodność analiz jakościowych i ilościowych, wykorzystywany jest w badaniach gen *gusA* kodujący ß-glukuronidazę. Stanowi on element plazmidu pCAMBIA1303.

W pracy podjęto próbę konstrukcji genu fuzyjnego CaMV::*sak−gusA−mgfp* w wektorze reporterowym pCAMBIA1303.



Rys. 1. Elektroforogram produktów po amplifikacji genu stafylokinazy z DNA plazmidowego wektora binarnego pCAMBIA1303*sak* wyizolowanego z komórek *E. coli* NM522 oraz po trawieniu plazmidu enzymami restrykcyjnymi *BglIIi* i *NcoI*. M. Wzorzec mas DNA (50 pz step ladder, SIGMA): 1031, 900, 800, 700, 600, 500, 400, 300, 250, 200 pz; 1. Kontrola pozytywna – produkt amplifikacji genu *sak* z wektora pTrxFus*sak*; 2. Produkt amplifikacji genu *sak* z wektora binarnego pCAMBIA1303*sak*; 3. Kontrola negatywna – PCR bez matrycy genu stafylokinazy; 4. Wektor binarny pCAMBIA1303 trawiony restryktazami *BglIIi* i *NcoI*; 5. Wektor binarny pCAMBIA1303*sak* trawiony restryktazami *BglII* i *NcoI*

Gen stafylokinazy amplifikowano na matrycy plazmidu pTrxFus*sak*, następnie produkty amplifikacji oraz DNA wektora binarnego trawiono endonukleazami. Wektor poddawano ligacji z produktami amplifikacji. Zrekombinowany wektor transformowano do szczepu *E. coli* NM522. Transformaty selekcjonowano na podłożu z kanamycyną. Rekombinanty identyfikowano, wykorzystując metodę PCR oraz analizę restrykcyjną z wykorzystaniem odpowiednich enzymów restrykcyjnych – rys. 1. Zrekombinowany wektor binarny pCAMBIA1303*sak* wprowadzono do komórek *A. tumefaciens* AGLI Rf^r. Zrekombinowany szczep *A. tumefaciens* AGLI Rf^r został wykorzystany do transformacji siewek tytoniu [2].

5. Dyskusja i podsumowanie

Transformacja roślin jest obecnie wykorzystywana w badaniach podstawowych (poznanie funkcji i regulacji ekspresji genów) i w pracach aplikacyjnych pozwalających m.in. na wykorzystanie pożądanych cech w przemyśle spożywczym i farmaceutycznym [3], [4].

W niniejszej pracy w celu konstrukcji genu fuzyjnego – nośnika genu stafylokinazy – wykorzystano wektor binarny pCAMBIA303.

Możliwość odczytu pełnej sekwencji nukleotydowej tego plazmidu w bazie danych www.ncbi.nlm.nih.gov znacznie ułatwiło zaprojektowanie odpowiednich starterów do procesu amplifikacji i konstrukcji pożądanego genu fuzyjnego pCAMBIA1303*sak.*

Wektor pCAMBIA1303 nie zawiera intronu uniemożliwiającego ekspresję transgenów w komórkach prokariotycznych. Fakt ten pozwala na badanie ekspresji konstrukcji genowych w komórkach *E. coil* i *Agrobacterium.*

Agrobacterium tumefaciens zawiera naturalny mechanizm transportu i integracji regionu T plazmidu Ti, naturalnie występuącego w *Agrobacterium,* z gnomem infekowanej komórki [5].

Konstrukcja genu fuzyjnego jest procesem wieloetapowym. Metody komputerowej analizy poszczególnych etapów pracy wspomagają planowanie dalszych eksperymentów laboratoryjnych.

Literatura cytowana

1. Baxevanis A. D., Ouellette B. F. F.: Bioinformatyka, PWN, Warszawa 2004.
2. Dobrowolska A., Kononowicz A. K.: Expression of a CaMV: sak−gusA−mgfp in tobacco plants, World Journal of Microbiology & Biotechnology, 20: 333-335, 2004.
3. Giddings G.: Transgenic plants as factories for biopharmaceuticals, Nature Biotechnology, 1, 18: 151-1155, 2000.
4. Malepszy S.: Biotechnologia roślin, PWN, Warszawa 2002.
5. Riva de la G. A., Gonzalez-Cabrera J., Vazquez-Padron R., Ayrapardo C.: Agrobacterium tumefciens : a natural tool for plant transformation, Electronic Journal of Biotechnology, 1: 1-16, 1998.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 113-116.

KOMPLEKSY BIAŁEK I UKŁADÓW CIEKŁOKRYSTALICZNYCH. MOŻLIWOŚĆ ZASTOSOWAŃ PRAKTYCZNYCH

SPECIFIC SUPRAMOLECULAR LIGANDS WITH PROTEINS. POSSIBLE PRACTICAL APPLICATION

ANNA DROZD[*], LESZEK KONIECZNY[*], JANINA RYBARSKA[*], BARBARA STOPA[*], BARBARA PIEKARSKA[*], IRENA ROTERMAN[**], MARCIN KRÓL[**], PAWEŁ SPÓLNIK[*], GRZEGORZ ZEMANEK[*]

[*]*Instytut Biochemii Lekarskiej, Collegium Medicum Uniwersytetu Jagiellońskiego, Kraków, Kopernika 7*
[**]*Zakład Bioinformatyki i Telemedycyny, Collegium Medicum Uniwersytetu Jagiellońskiego, Kraków, Kopernika 17*

Streszczenie. Wg powszechnie przyjętego mechanizmu tworzenia się specyficznych kompleksów z białkami, miejscem wiązania ligandów, w tym również pochodzenia niebiologicznego (a więc np. leków) jest grupa czynna białka. To miejsce w białku umożliwia wiązanie liganda przy ograniczonym dostępie wody, która skutecznie zapobiega asocjacjom powierzchniowym. W związku z dynamicznym charakterem struktury białka istnieje jednak możliwość penetracji ligandów do obszarów głębszych poza grupą czynną. Kompleksowanie napotyka jednak tutaj na konkurencję łańcucha polipeptydowego, który jedynie w wyniku fluktuacji przejściowo umożliwił penetrację liganda. Konkurencji nie wytrzymują z reguły związki drobnocząsteczkowe. Jednak może jej sprostać zespołowa interakcja, jaką reprezentują zasocjowane cząsteczki liganda supramolekularnego pod warunkiem, że wzajemnie oddziaływujące cząsteczki liganda – nadające mu spójność – są odpowiednio duże. Ligandy supramolekularne, które tworzą micele taśmowe, korespondują strukturalnie z łańcuchem peptydowym o konformacji Beta. Te szczególne rodzaje struktur ciekłokrystalicznych nadają się do tworzenia kompleksów poza grupą czynną białek. Plastyczność tych ligandów wybitnie ułatwia kompleksowanie.
Ponieważ dynamika natywnego białka ujawnia się głównie w wyniku ruchów związanych z funkcją, utworzone kompleksy zamrażające przejściowy stan dynamiczny mają istotny wpływ na biologiczną aktywność białka. Kompleksowanie ligandów supramolekularnych o własnościach ciekłokrystalicznych jest problemem nowym. Zjawisko otwiera duże możliwości zastosowań praktycznych.

Słowa kluczowe: struktura białek, symulacja zmian strukturalnych, kompleksy białek z ligandami ciekłokrystalicznymi

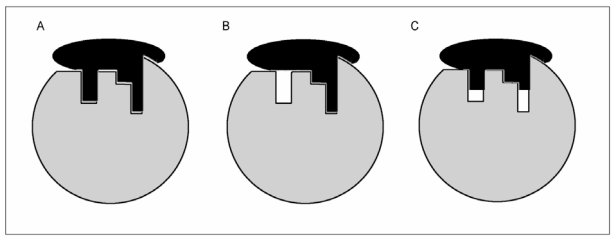
Abstract. The protein binding site is (according to commonly accepted mechanism) the only locus responsible for the specific ligation of compounds of non-biological origin, including drugs. The binding cleft favors association decreasing the access of water which prevents efficiently the non-covalent binding at the protein surface. Nevertheless, the dynamic character of protein structure may allow penetration of small compounds to the periphery or the interior of the molecule, outside the binding site. However, the complexation meets here the strong competition of polypeptide chains making the non-specific interaction of small molecules unsufficient to ensure the binding. This competition may be faced after all by collective interaction of self-assembled molecules if the local or global instability of protein structure allows penetration of a large unit of supramolecular ligand and self-assembling is strong enough to replace the polypeptide chain from its packing locus. The supramolecular ligands of Congo red family which form ribbon-shaped micellar entities and correspond in some way to polypeptide chains of Beta-conformation offer the largest interaction contact. Supramolecular structures of this kind appear capable of sustaining the competition of polypeptide chains and form stable complexes. The strong complexation depends particularly on plasticity of dye ligand improving the fitting to receptor sites in protein. Since the increased structural dynamics of native protein, allowing the penetration of dye is commonly associated with function-derived movements, the complexation may affect biological activity. The recently discovered complexation of supramolecular structures with proteins seems to be a promising phenomenon for pharmacology and medicine.

Key words: protein structure, simulation of structural changes in proteins, protein – supramolecular ligand complexes

W układach biologicznych funkcja związana jest przede wszystkim z białkami. Dlatego oczywistym jest zainteresowanie możliwościami jej modyfikacji. Najłatwiejszym i najczęściej stosowanym sposobem uzyskania efektów jest tworzenie kompleksów białek ze związkami chemicznymi. Farmakologia, toksykologia i wiele technologii stosowanych w nauce i praktyce wykorzystują tę zasadę. Najbardziej poszukiwanymi dla tych celów związkami są takie, które działają specyficznie i tworzą kompleksy odwracalne. Takim działaniem można sterować. Specyficzność kojarzy się jednoznacznie z tworzeniem

kompleksów w grupie czynnej białka, gdzie wiąże się naturalny ligand. Grupa czynna białka stwarza takie możliwości, ponieważ jest to z reguły jedyne miejsce, gdzie może powstać kompleks przy ograniczonym dostępie wody. Grupa czynna jest bowiem szczeliną w upakowanym natywnym białku, gdzie oprócz aminokwasów polarnych odpowiedzialnych za specyficzne oddziaływania obecne są również aminokwasy lub ugrupowania o cechach niepolarnych, ograniczające penetrację wody i jonów, które konkurują z wiązaniami wodorowymi i jonowymi odpowiedzialnymi za specyficzność kompleksowań opartych na oddziaływaniach niekowalencyjnych.

Klasycznymi przedstawicielami grupy ligandów białek tworzącymi kompleksy specyficzne, odwracalne są inhibitory enzymatyczne i inhibitory receptorów, które w znakomitej większości znane są jako leki lub trucizny. Wiązanie się ligandów w grupie czynnej jest, oczywiście, wiązaniem konkurencyjnym wobec naturalnego substratu i jego efekt jest określany jako hamowanie kompetycyjne. Ligand może być podobny do substratu, ale nie musi. Wystarczy odwracalne wiązanie się w grupie czynnej uniemożliwiające lub ograniczające prawidłowe dokowanie substratu. Efektywność kompetycyjnego działania liganda zależy od jego powinowactwa do grupy czynnej białka i stężenia (rys. 1).

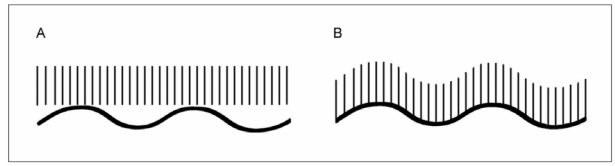


Rys. 1. Wyidealizowany modelowy obraz tworzenia kompleksów białka z podobnymi, mogącymi konkurować o miejsce wiązania ligandami przy ich różnym, wynikającym ze struktury powinowactwie (A,B,C)

To samo dotyczy substratu. W klasycznym wydaniu jest to konkurencja o miejsce wiązania bez zmiany zdolności katalitycznej, a więc bez modyfikacji elementu katalitycznego, tj. ugrupowania odpowiedzialnego bezpośrednio za katalizę. Ten typ oddziaływania jest najczęstszy. Białko ma jednak strukturę kooperatywną i odległe wiązania ligandów mogą wpłynąć na przegrupowania również w grupie czynnej, dając efekt hamowania. Takie oddziaływania nie mają cech konkurencyjności. Są one raczej rzadkie, ponieważ niekowalencyjne, odwracalne kompleksy na powierzchni upakowanego, natywnego białka tworzone przez niespecyficzne oddziaływania związków drobno-cząsteczkowych mają bardzo małą szansę na stabilność ze względu na silną konkurencję wody i jonów.

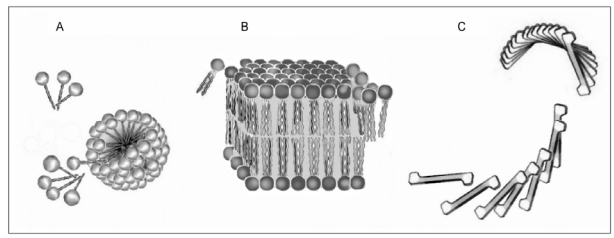
Dynamika struktury białka wyrażająca się fluktuacją łańcuchów polipeptydowych umożliwia penetrację związków o niewielkich rozmiarach do obszarów między-łańcuchowych poza grupą czynną. Te możliwości penetracji i wiązania są jednak przejściowe. Wolna przestrzeń zostaje – w przeciwieństwie do grupy czynnej – wypełniona na powrót przez łańcuch polipeptydowy, a penetrujące pojedyncze ligandy, które wiążą się niespecyficznie i z reguły słabo, zostają łatwo wyparte.

Ligandy, które mogłyby tworzyć w tych warunkach mocne kompleksy niekowalencyjne, powinny mieć charakter łańcuchowy, ponieważ tylko taka forma strukturalna może zapewnić odpowiednio duże przyleganie do powierzchni łańcuchów polipeptydowych, zapewniając tworzenia stabilnego kompleksu przez wielokrotność oddziaływań. Istotnym warunkiem skuteczności kompleksowania łańcuchowej formy liganda jest jednak plastyczność, umożliwiająca lepsze dopasowanie się do struktury receptorowej białka, reprezentowanej w tym przypadku przez łańcuch polipeptydowy (rys. 2).



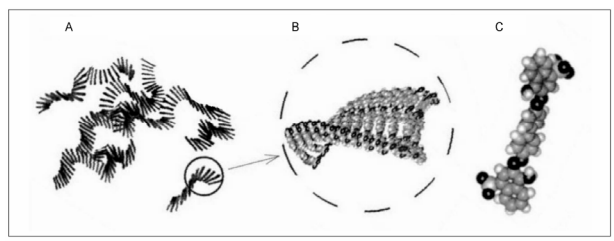
Rys. 2. Modelowe przedstawienie efektywności adhezji A – sztywnej, B – plastycznej w kompleksie liganda supramolekularnego z białkiem

Takie warunki mogą spełniać polimery łańcuchowe lub wielo-cząsteczkowe asocjaty, tworzące zintegrowane nanostruktury (rys. 3) [1].



Rys. 3. Układy supramolekularne reprezentowane przez różne formy struktur micelarnych: A – kuliste (mydła), B – płytowe (błony komórkowe), C – taśmowe (barwniki bis-azowe)

Oczekiwane własności mają, jak niedawno stwierdzono, struktury micelarne liotropowych układów ciekłokrystalicznych należące do grupy ciekłych kryształów chromoforowych [2], [3], [4], [5]. Barwniki z tej grupy, której przedstawicielem jest czerwień Kongo, tworzą, asocjując, taśmowe struktury micelarne, których kilku-cząsteczkowe fragmenty mogą na zasadzie adhezji przylegać do łańcuchów polipeptydowych Beta (rys. 4).



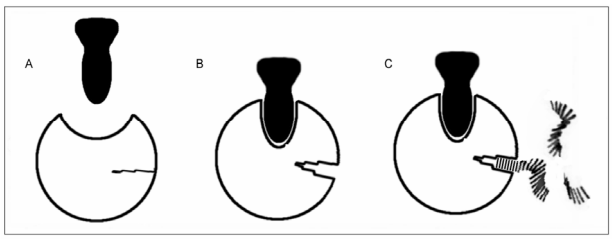
Rys. 4. Modelowe wyobrażenie mezo-fazy tworzonej przez taśmowe układy supramolekularne: A – ideowa prezentacja micelarnej struktury mezo-fazy, B – powiększony fragment miceli tworzonej przez czerwień Kongo, C – cząsteczka czerwieni Kongo – forma przestrzenna

Adhezja, umożliwiająca wielopunktowe oddziaływanie, stwarza możliwość mocnego wiązania. Istotnym warunkiem takiego kompleksowania jest jednak odpowiednio silna samoasocjacja cząsteczek liganda supramolekularnego [5], [6]. Adhezja w białku jest możliwa bowiem tylko wówczas, kiedy wielo-cząsteczkowy ligand reaguje jako spójna jednostka [7],

[8], [9], [10]. Niektóre struktury micelarne o odmiennym ukształtowaniu miceli, w tym szczególnie detergenty, jak SDS, mogą również tworzyć kompleksy z białkami. Kompleksowanie to jednak wymusza zmiany strukturalne, które nie korespondują z natywnym ufałdowaniem łańcucha polipeptydowego, co w efekcie prowadzi do denaturacji. Wytworzenie kompleksu przez duże wielocząsteczkowe ligandy napotyka jednak na bariery. Niezbędna dla wytworzenia kompleksu penetracja liganda jest możliwa tylko w niestabilnych obszarach białka, a więc tam, gdzie łańcuch polipeptydowy wykazuje silną fluktuację. Białka natywne dobrze upakowane nie tworzą trwałych kompleksów z ligandami supramolekularnymi. Taśmowy ligand wiąże się z białkiem, oddziaływując swoim fragmentem niepolarnym z łańcuchem polipeptydowym białka [10], [11], [12], [13].

Spójność liganda wynikająca z oddziaływań niekowalencyjnych ma jednak swoje optimum. Zbyt słabe i zbyt mocne oddziaływania spajające cząsteczki w ligandzie supramolekularnym nie są korzystne. Te pierwsze – ponieważ nie zapewniają odpowiedniej spójności liganda i te drugie – ponieważ utrudniają dopasowanie się liganda do struktury receptorowej (łańcucha polipeptydowego) [6].

Do białek, które z racji niestabilności lokalnej lub globalnej mogą wiązać ligandy supramolekularne, należą – poza białkami poddanymi denaturacji – białka patologiczne takie, których niestabilność jest spowodowana zmianami mutacyjnymi [14], [15]. Białka te można wykrywać, kompleksując je z barwnikami supramolekularnymi [16], [17]. Jednak najbardziej atrakcyjnym obiektem dla tworzenia kompleksów są takie białka natywne, w których przegrupowania strukturalne powstające w wyniku naprężeń towarzyszących funkcji są znaczące i w efekcie umożliwiają penetrację i wiązanie barwników supramolekularnych. (rys. 5).



Rys. 5. Schematycznie przedstawiony mechanizm tworzenia się naprężeń w cząsteczce białka reagującego z naturalnym ligandem skutkującego zwiększoną dostępnością do penetracji przez obce ligandy: A – przed reakcją z naturalnym substratem, B – zmiany strukturalne generowane wiązaniem substratu, C – zmniejszenie odwracalności wiązania substratu przez białko w wyniku przyłączenia supramolekularnego liganda

Tworząc kompleks, utrwala się stan przejściowy i modyfikuje efekty funkcyjne. Jeżeli aktywność białka jest wyrażana tworzeniem trwałych kompleksów, to zamrożenie stanu przejściowego daje wzmocnienie efektu biologicznego. Tak jest w przypadku przeciwciał reagujących z antygenem. Efektu takiego nie da się uzyskać w przypadku ligandów, które tworzą kompleksy wiążąc się w grupie czynnej białek.

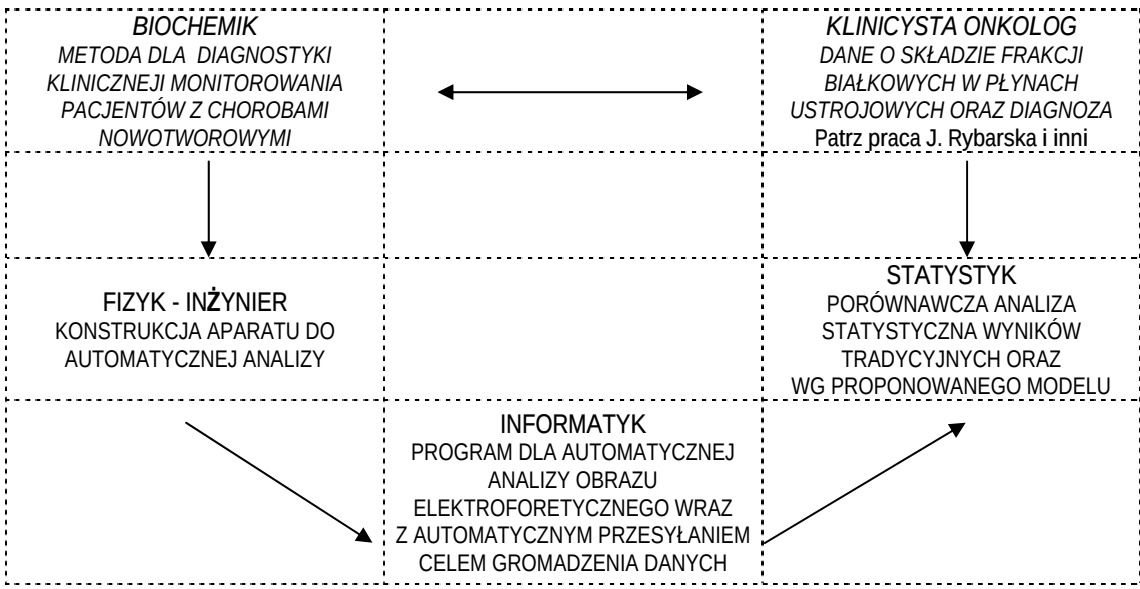
Wykorzystując ligandy supramolekularne poczyniono znaczące postępy w wyjaśnianiu mechanizmu tzw. sygnału immunologicznego [18], [19], [20], [21], [22], [23]. Sygnał generowany wiązaniem antygenu przez przeciwciała, wyzwalający aktywności efektorowe, inicjowany bezpośrednio przez fragment Fc jest blokowany przez ligandy supramolekularne, podczas gdy sama reakcja przeciwciała z antygenem jest silnie wzmacniana [6], [24]. Taśmowe ligandy supramolekularne reagują specyficznie z uaktywnioną przenoszącą sygnał cząsteczką przeciwciała. Stwarza to olbrzymie możliwości interwencji farmakologicznej w procesy mediowane przez przeciwciała. Również receptory z nadrodziny immunoglobulin wiążą ligandy supramolekularne w trakcie funkcji [25].

Wykorzystując niezwykłe cechy układów supramolekularnych, opracowuje się technikę docelowego transportu leków [26], [27]. Ligandy supramolekularne wykorzystuje się również w analizie struktury amyloidów [28]. Niestandardowy mechanizm tworzenia kompleksów z białkami przez ligandy supramolekularne otwiera nowe możliwości tworzenia kompleksów i wpływania na funkcję.

Literatura cytowana

1. Blau W. J., Fleming A. J.: Designer nanotubes by molecular self-assembly, Science, 304: 1457-1458, 2004.
2. Attwood T. K., Lydon J. E., Hall C., Tiddy G. J. T.: The distinction between chromonic and amphiphilic lyotropic mesophases, Liquid Crystals, 7: 657-668. 1990.
3. Wasilewskaya A. C., Generalova E. W., Sonin A. C.: Chromonic mesophases, Uspiechi chimi (Russ) LVIII:1574-1596, 1989.
4. Perahia D., Wachtel E. J., Luz Z.: NMR and X-ray studies of the chromonic lyomesophases formed by some xanthone derivatives, Liquid Crystals, 9: 479-492, 1991.
5. Skowronek M., Stopa B., Konieczny L., Rybarska J., Piekarska B., Szneler E., Bakalarski G., Roterman I.: Self-assembly of Congo red – a theoretical and experimental approach to identify its supramolecular organisation in water and salt solution, Biopolymers, 46: 267-281, 1998.
6. Stopa B., Piekarska B., Konieczny L., Rybarska J., Spólnik P., Zemanek G., Król M., Roterman I.: The structure and protein binding of amyloid-specific dye reagents, Acta Biochimica Polonica, 50: 1213-1227, 2003.
7. Roterman I., No K. T., Piekarska B., Kaszuba J., Pawlicki R., Rybarska J., Konieczny L.: Bis-azo dyes – studies on the mechanism of complex formation with IgG modulated by heating or antigen binding, J. Physiol. Pharmacol., 44: 213-232, 1993.
8. Roterman I., Konieczny L., Stopa B., Rybarska J., Piekarska B.: Heat-induced structural changes in the Fab fragment of IgG recognised by molecular dynamics simulation. Implications for signal transduction in antibodies, Folia Biologica (Kraków), 35: 115-128, 1994.
9. Rybarska J., Piekarska B., Stopa B., Zemanek G., Konieczny L., Nowal M., Król M., Roterman I., Szymczakiewicz-Multanowska A.: Evidence that supramolecular Congo red is the sole ligation form of this dye for L chain lambda derived amyloid proteins, Folia Histochem. Cytobiol, 39: 307-314, 2001.
10. Piekarska B., Konieczny L., Rybarska J., Stopa B., Zemanek G., Szneler E., Król M., Nowak M., Roterman I.: Heat-induced formation of a specific binding site for self-assembled Congo red in the V domain of immunoglobulin L chain lambda, Biopolymers, 59: 446-456, 2001.
11. Roterman I., Król M., Nowak M., Konieczny L., Rybarska J., Stopa B., Piekarska B., Zemanek G.: Why Congo red binding is specific for amyloid proteins, Medical Science Monitor, 7: 771-784, 2001.
12. Król M., Roterman I., Piekarska B., Konieczny L., Rybarska J., Stopa B., Local and long range structural effects caused by the removal of the N-terminal polypeptide fragment from immunoglobulin L chain lamda, Biopolymers, 69: 189-200, 2003
13. Roterman I., Rybarska J., Konieczny L., Skowronek M., Stopa B., Piekarska B., Bakalarski G.: Congo red bound to α-1 proteinase inhibitor as a model of supramolecular ligand and protein complex, Computers & Chemistry, 22: 61-70, 1998.

14. Augener W., Grey H. M.: Studies on the mechanism of heat aggregation of human IgG, J. Immunol., 105: 1024-1030, 1970.
15. Cogne M., Silvain C., Khamlichi A. A., Preud'homme J.-L.: Structurally abnormal immunoglobulins in human immunoproliferative disorders, Blood, 79: 2181-2195, 1992.
16. Spólnik P., Konieczny L., Piekarska B., Rybarska J., Stopa B., Zemanek G., Król M., Roterman I.: Instability of monoclonal myeloma protein may be identified as susceptibility to penetration and binding by newly synthetised Congo red derivatives, Biochimie, 86: 397-401, 2004.
17. Spólnik P., Piekarska B., Stopa B., Konieczny L., Zemanek G., Rybarska J., Król M., Nowak M., Roterman I.: The structural abnormality of myeloma immunoglobulins tested by Congo red binding, Med. Sci. Monit., 9: 145-153, 2003.
18. Piekarska B., Konieczny L., Rbarska J., Stopa B., Spólnik P., Roterman I., Król M.: Intramolecular signaling in immunoglobulins – new evidence emerging from the use of supramolecular protein ligands, J. Physiol. Pharmacol., 55: 487-501, 2004
19. Calvanico N. J., Tomasi T. B.: Effector sites on antibodies. In: Atassi M. Z. (ed.): Immunochemistry of proteins, Vol. 3, New York and London, Plenum Press, 1979, pp. 1-85.
20. Weininger R. B., Richards F. F.: Combining regions of antibodies. In: Atassi M. Z. (ed.): Immunochemistry of proteins. Vol. 3, New York and London, Plenum Press, 1979, pp. 123-166.
21. Metzger H.: General aspects of antibody structure and function. In Fc receptors and the action of antibodies, In: Metzger H. (ed.), Washington DC, Am. Soc. Microbiol.: pp. 7-11, 1990.
22. Schneider W. P., Wemsel T. G., Stryer L., Oi V. T.: Genetically engineered immunoglobulins reveal structural features controlling segmental flexibility, Proc. Natl. Acad. Sci. USA, 85: 2509-2513, 1988.
23. Mizutani R., Igarashi T., Tanaka T., Shimada I., Arata Y.: Effector functions of a mouse IgG that lacks the entire CH1 domain. C1q binding and complement fixation in the absence of antigen, J. Immunol., 150: 131-138, 1993.
24. Rybarska J., Konieczny L., Roterman I., Piekarska B.: The effect of azo dyes on the formation of immune complexes, Arch. Immunol. Ther. Exp., 39: 317-327, 1991.
25. Zemanek G., Rybarska J., Stopa B., Piekarska B., Spólnik P., Konieczny L., Roterman I., Bugajski A.: Protein distortion-derived mechanism of signal discrimination in monocytes revealed using Congo red to stain activated cells, Folia Histochem Cytobiol., 41: 113-124, 2003.
26. Konieczny L., Piekarska B., Rybarska J., Skowronek M., Stopa B., Tabor B., Dąbroś W., Pawlicki R., Roterman I.: The use of Congo red as a lyotropic liquid crystal to carry stains in a model immunonotargeting system – microscopic studies, Folia Histochem Cytobiol., 35: 203-210, 1997.
27. Rybarska J., Piekarska B., Stopa B., Spólnik P., Zemanek G., Konieczny L., Roterman I.: In vivo accumulation of self-assembling dye Congo red in an area marked by specific immune complexes – Possible relevance to chemotherapy, Folia Histochem Cytobiol, 42: 101-110, 2004.
28. Zemanek G., Konieczny L., Piekarska B., Rybarska J., Stopa B., Spólnik P., Urbanowicz B., Nowak M., Król M., Roterman I.: Egg yolk platelet from Xenopus laevis are amyloidogenic, Folia Histochem. Cytobiol., 40: 311-318, 2002.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 117-120.

KONSTRUKCJA SEKWENCJI KONSENSUSOWYCH RODZINY BIAŁEK β-SPEKTRYN O RÓŻNYCH PARAMETRACH PROGOWYCH ORAZ WERYFIKACJA ICH UŻYTECZNOŚCI

β-SPECTRIN CONSENSUS SEQUENCE CONSTRUCTION WITH VARIABLE THRESHOLD PARAMETERS; VERIFICATION OF USEFULNESS

ANNA FOGTMAN[*], JACEK LELUK[*], BOGDAN LESYNG[*,#]

[*]*Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego (ICM), Uniwersytet Warszawski, ul. Żwirki i Wigury 93, 02-089 Warszawa, lulu@icm.edu.pl*
[#]*Zakład Biofizyki, Wydział Fizyki, Uniwersytet Warszawski, ul. Żwirki i Wigury 93, 02-089 Warszawa, B.Lesyng@icm.edu.pl*

Streszczenie. Rozwój zarówno sekwencyjnych, jak i strukturalnych białkowych baz danych umożliwił prowadzenie wielkoskalowych badań nad podobieństwem i homologią białek. Etapem kluczowym takich badań jest poprawna konstrukcja dopasowanego zestawienia sekwencji (z ang.: *Multiple Alignment*) oraz sekwencji konsensusowych. Sekwencja taka jest swoistym markerem danej rodziny białkowej; określa ona zestaw białek homologicznych specyficznie uwydatniających cechy charakterystyczne danej rodziny. Sekwencja konsensusowa jest uśrednieniem cech całej rodziny białkowej, umożliwiającym wnikliwą analizę mechanizmów ewolucji białek oraz stworzenie ich ogólnego schematu. Taka sekwencja może zostać użyta jako narzędzie do przeszukiwania baz danych pod względem podobieństwa i homologii białek wykazujących istotny stopnień identyczności. Prawidłowa konstrukcja sekwencji konsensusowych jest istotna dla uzyskania najbardziej optymalnych wyników analizy porównawczej białek. Konstrukcja takich sekwencji oparta jest na parametrach odpowiadających liczbie białek należących do danej rodziny, długości sekwencji oraz stopniu identyczności. W naszej pracy przeanalizowaliśmy ponad 50 sekwencji aminokwasowych β-spektryn. Dopasowanie tych sekwencji oraz zestaw optymalnych sekwencji konsensusowych dla β-spektryn skonstruowano z wykorzystaniem algorytmu semihomologii genetycznej oraz programu Consensus Constructor (ICM). Przeanalizowano wpływ różnych parametrów progowych na użyteczność skonstruowanych sekwencji konsensusowych. Sekwencje te wykorzystano do przeszukania sekwencyjnych baz danych, a wyniki przeszukiwania porównano pod względem ich użyteczności w badaniach porównawczych białek.
Przedstawiona strategia konstrukcji sekwencji konsensusowych wydaje się być obiecującym, użytecznym narzędziem w precyzyjnym określaniu różnorodnych wspólnych cech homologicznych rodzin białkowych, nawet gdy obserwowany stopień podobieństwa/identyczności jest relatywnie niski.
Praca wykonana została w ramach grantu MAMBA (Centre of Excellence for Multi-scale Biomolecular Modelling, Bioinformatics and Applications) Projekt nr QLRI-CT-2002-90383.

Słowa kluczowe: homologia, podobieństwo białek, dopasowanie sekwencji, sekwencja konsensusowa, identyczność

Abstract. The development of the protein sequence and structure databases made wide scale similarity searches and homology analysis possible. The fundamental step in this research is a proper multiple alignment and consensus sequence construction. The consensus sequence is a specific marker describing the features of a given protein family. It enables the profound analysis of the evolutionary mechanisms and creation of their general structural scheme. It defines the set of homologous proteins, specifically emphasizing the distinctive features of a protein family. Such a sequence can be used as a tool to search databases for homologous proteins revealing the best actual similarity/homology score. Correct consensus sequence construction is key for obtaining the most optimum results of protein sequence homology research. The consensus sequence construction is based on parameters that apply to the number of members within a family, length of the protein sequences and the degree of identity. We analyzed over 50 sequences of β-spectrin family. The multiple alignment and the set of optimum consensus sequences for β-spectrin protein family were constructed with the aid of genetic semihomology algorithm and Consensus Constructor program (ICM). We examined the influence of variable threshold parameters on usefulness of the constructed consensus sequence. The consensus sequences of various threshold parameters were used to search the databases, and the scores were compared with respect to their usefulness in many criteria of comparative protein analysis.
The strategy of consensus sequence construction appears to be powerful and accurate optimization tool for precise description of the significant common features of homologous protein families, even when the observed similarity/identity degree is relatively low.
This work was supported by MAMBA (Centre of Excellence for Multi-scale Biomolecular Modelling, Bioinformatics and Applications) Project No. QLRI-CT-2002-90383.

Key words: protein homology, multiple alignment, consensus sequence, identity

1. Wstęp

Spektryna została po raz pierwszy zidentyfikowana jako główny składnik cytoszkieletu błony erytrocytu, kontrolujący organizację błony, jej stabilność i kształt. Obecnie wiadomo, że spektryny występują powszechnie w komórkach wszystkich tkanek kręgowców i bezkręgowców; biorą udział w wielu procesach istotnych dla prawidłowego funkcjonowania komórki. Są to białka cytoszkieletu niezbędne dla utrzymania prawidłowego kształtu komórki, sprężystości błony w odpowiedzi na stres mechaniczny, determinujące rozmieszczenia białek transmembranowych oraz organizację organelli w cytoplazmie. Spektryny poprzez oddziaływania z hydrofobowym zrębem błony komórkowej stanowią cytosolowy czynnik uczestniczący w budowie systemu filamentów aktynowych, a stosunkowo niedawno odkryto rolę tych białek jako uczestników szlaków wydzielniczych komórek [1]. Różnorodne izoformy spektryn oraz wiążących je białek – ankiryn są obecne na powierzchni aparatów Golgi'ego, transportowanych intermediatów oraz na błonach ścieżki endocytarnej. Spektryny odgrywają kluczową rolę w stabilizacji dwuwarstwy białkowo-lipidowej błon plazmatycznych, organizacji domen białek integralnych, kontroli ruchliwości receptorów błonowych, adhezji komórek, przekazywaniu impulsów nerwowych, tworzeniu pęcherzyków wydzielniczych oraz ich transporcie pomiędzy organellami, a także w rozwoju i morfogenezie komórek zarodkowych. Zmiany w strukturze pierwszorzędowej, których przyczyną są mutacje genetyczne, powodują uszkodzenia lub nieobecność białek szkieletu błonowego, co upośledza jego oddziaływanie z błoną komórkową. Zaburzenia te prowadzą do deformacji, utraty elastyczności i zmniejszenia powierzchni komórek.

Cząsteczka spektryny to heterodimer zbudowany z niekowalencyjnie połączonych ze sobą podjednostek α i β (α2β2). Łańcuchy tych podjednostek, owinięte wokół tej samej osi, ułożone są względem siebie anty-równolegle. Tetramer taki ma długość około 200 nm oraz średnicę około 3 nm i stanowi on podstawową jednostkę funkcjonalną spektryny.

β-Spektryny zawierają 19 segmentów, z których pierwszym jest wysoko konserwatywna N-końcowa domena złożona z dwóch sąsiednich domen będących homologami kalponiny. Środek cząsteczki zajmuje 17 potrójnych helikalnych powtórzeń, a koniec – domena zawierająca motyw PH – region homologiczny do plekstryny o wielkości 21.4 kDa. Miejsca wiążące ankirynę zlokalizowane są w środkowej części tetrameru, w segmencie 15.

Celem pracy była analiza przedstawicieli rodziny białkowej β-spektryn pod względem homologii, czyli podobieństwa i różnic w obrębie ich struktur pierwszorzędowych oraz konstrukcja sekwencji konsensusowych (wygenerowanych na podstawie analizy zestawienia dopasowanych wzajemnie sekwencji) o różnych, zmiennych parametrach progowych i weryfikacja ich zastosowania w poszukiwaniu homologii wśród białek. Celem pośrednim była weryfikacja narzędzia użytego w tej pracy – algorytmu semihomologii genetycznej; jego praktycznego zastosowania w teoretycznych badaniach zmienności ewolucyjnej i pokrewieństwa białek.

2. Materiały

Materiałem wyjściowym były sekwencje aminokwasowe białek β-spektryn pobrane z baz danych: SwissProt, UniProt oraz Trembl[1]. Pierwszym krokiem był wybór wzorcowej sekwencji białkowej, za pomocą której możliwe byłoby przeszukanie baz danych w celu otrzymania białek homologicznych. Jako białka wzorcowego użyto ludzkiej erytrocytarnej β-spektryny[2]. Następnym krokiem było wyszukanie sekwencji najbardziej podobnych do wzorcowej – otrzymano łącznie 85 sekwencji, z czego odrzucono 12 – były to sekwencje powtarzające się, niekompletne bądź zupełnie odbiegające swym składem aminokwasowym od pozostałych, co mogłoby zaburzyć poprawność konstrukcji sekwencji konsensusowych.

3. Metody

Ustalenie sekwencji wzorcowej. Jako sekwencję wzorcową wybrano ludzką erytrocytarną β-spektrynę (P11277) w wyniku przeszukania białkowych baz danych: SwissProt, UniProt i Trembl. Następnie wyselekcjonowano 84 sekwencje wykazujące najwyższy stopień podobieństwa do ludzkiej erytrocytarnej β-spektryny przy użyciu algorytmu BLAST[3] [2], [3]. Bazy danych do znalezienia homologów zostały wybrane domyślnie: CDS, SwissProt i PDB. Przeszukiwanie bazy sekwencji białkowych dokonano z wykorzystaniem macierzy BLOSUM62.

Dopasowanie wielu sekwencji (*Multiple Alignment*). Dopasowanie sekwencji przeprowadzano w kilku etapach. W pierwszym etapie za pomocą programu GEISHA[4] [4] dokonywano dopasowania 2 wybranych sekwencji aminokwasowych w oparciu o algorytm semihomologii genetycznej [5], [6], [7], [8], [9]. Przy tak długich sekwencjach białkowych, jakimi są β-spektryny, ilość wyrzucanych przez program wyników jest ogromna, w związku z czym prawidłowe ich odczytanie, a co za tym idzie – właściwa interpretacja, nie są możliwe. W związku z powyższym program używany był doraźnie, w przypadku pojawienia się wątpliwości co do prawidłowego dopasowania pojedynczych reszt aminokwasowych. Znacząca większość pracy dotycząca wzajemnego dopasowania sekwencji spektrynowych została wykonana manualnie.

Określanie pozycji konserwatywnych oraz wykazujących genetyczne pokrewieństwo (obsadzonych przez aminokwasy, których kodony różnią się tylko jedną pozycja). Na tym etapie wykorzystano wstępnie program Protein Calculator[5]. W celu uzyskania poprawnych wyników ostatecznych niezbędna była manualna weryfikacja. W zestawieniach dopasowanych sekwencji wyróżniono pozycje wybitnie konserwatywne (jako białe litery na czarnym tle) oraz pozycje typowe dla danej rodziny homologicznej (oznaczone kolorem szarym). Delecje oraz pozycje nie wykazujące jednoznacznej homologii pozostały niewyróżnione.

Konstrukcja wzorcowej sekwencji konsensusowej dla rodziny β-spektryn oraz sekwencji konsensusowych o zmiennych parametrach progowych. Sekwencje skonstruowano wstępnie przy pomocy programu Protein Calculator

[1] http://srs.ebi.ac.uk
[2] Nr akcesyjny: P11277.
[3] Pakiet BLAST dostępny jest z serwera NCBI (*National center for Biotechnology Information*) na stronie: http://www.ncbi.nlm.nih.gov/blast/
[4] Program jest własnością Interdyscyplinarnego Centrum Modelowania Matematycznego i Informatycznego Uniwersytetu Warszawskiego.
[5] Wersja v0.901 Beta jest własnością Interdyscyplinarnego Centrum Modelowania Matematycznego i Informatycznego Uniwersytetu Warszawskiego.

(zaznaczenie reszt identycznych), a następnie poddano manualnej weryfikacji w celu uzyskania precyzyjnego wyniku końcowego. Każdy zestaw odpowiadających sobie pojedynczych pozycji opisano w oparciu o trzy wstępnie zadeklarowane parametry definiujące identyczność, homologię i delecje (brak pozycji w podsumowaniu konsensusowym). Za reszty typowe dla danej pozycji (pozycje homologiczne) uznano takie, których kodony mogą wzajemnie przechodzić w siebie na drodze pojedynczej tranzycji/transwersji (reszty semihomologiczne), lecz żaden z tych aminokwasów nie występuje tam w wyraźnym nadmiarze. Takie pozycje zostały oznaczone jako „X" w sekwencji konsensusowej. Za *identyczne* uznano pozycje obsadzone przez tylko jeden rodzaj aminokwasu. Poprzez zmianę wartości progowych dla tych trzech parametrów uzyskano szereg różnych sekwencji konsensusowych dla tego samego zestawienia dopasowanych sekwencji. Dalszej szczegółowej analizie poddano 7 sekwencji konsensusowych skonstruowanych w ten sposób: po 2 dla każdego parametru ze skrajnymi jego wartościami przy domyślnych (optymalnych) wartościach pozostałych dwóch parametrów oraz wzorcową sekwencję konsensusową, z optymalnie dobranymi wartościami wszystkich trzech parametrów.

Przeszukiwanie sekwencyjnych baz danych. Zasoby sekwencyjnych baz danych przeszukano przy użyciu skonstruowanych sekwencji konsensusowych z wykorzystaniem programu BLAST (blastp). Delecje wynikające z dopasowania sekwencji zostały usunięte w sekwencjach konsensusowych, bowiem program przez konstruowanie na bazie sekwencji matrycy pozycyjno specyficznej uwzględnia występowanie przerw w sekwencjach.

4. Wyniki

Do analizowanej rodziny β-spektryn zostały zakwalifikowane 73 spośród 85 zgromadzonych sekwencji białkowych. Sensem dokonanej selekcji było otrzymanie jej przedstawicieli wykazujących najwyższy stopień wzajemnego pokrewieństwa, jak i wyeliminowanie błędnych lub niekompletnych rekordów zdeponowanych w bazach danych. W ten sposób otrzymano spójną grupę sekwencji spektrynowych, które poddano dalszej szczegółowej analizie teoretycznej.

Wstępnym etapem było wzajemne dopasowanie sekwencji β-spektryn, które posłużyło do skonstruowania sekwencji konsensusowej dla tej rodziny białek. Celem wzajemnego dopasowania homologicznych sekwencji białkowych jest wykazanie identyczności, podobieństw oraz różnic porównywanych struktur pierwszorzędowych. Dokonuje się tego poprzez takie wzajemne ustawienie sekwencji aminokwasowych, by odpowiadające sobie pozycje (pod względem filogenetycznym) znajdowały się w tej samej kolumnie. Kreuje to ogólny obraz homologii w obrębie analizowanej rodziny, ułatwiając jednocześnie dalsze badania nad nią. Graficznie ma to postać zestawienia, w którym pozycje konserwatywne oznaczone są jako białe symbole na czarnym tle, a reszty semihomologiczne występują na szarym polu.

Przy ustalaniu zestawienia dopasowanych sekwencji niezbędne jest branie pod uwagę delecji oraz insercji, które były jednym z mechanizmów różnicujących białka homologiczne. Delecje mogą się pojawić np. na skutek pojawienia się w wyniku pojedynczej mutacji punktowej kodonu STOP, który uniemożliwiłby funkcjonalność danego białka. Mechanizmy naprawcze mogą przypuszczalnie usunąć cały taki kodon. W związku z tym w naszej analizie delecja traktowana jest jako 21 aminokwas, w pewnym sensie odpowiadający trzem kodonom nonsensownym. W przyjętej procedurze strategia wstawiania delecji zgodna jest z kierunkiem biosyntezy białka (analiza prowadzona jest w kierunku od N-końca do C-końca), czyli wstawiana jest w miejscu najbardziej dogodnym od strony N-końcowej łańcucha.

W oparciu o dopasowane 73 sekwencje wykazujące jednoznaczną homologię do ludzkiej β-spektryny erytrocytarnej sporządzono siedem sekwencji konsensusowych. Sekwencje konsensusowe różniły się sposobem ich konstrukcji, tzn. wartości progowe parametrów identyczności, homologii
i delecji były w każdym przypadku inne. Dobór parametrów progowych dla siedmiu przeanalizowanych sekwencji konsensusowych przedstawia tab. 1.

Tab. 1. Zestawienie procentowych wartości parametrów, za pomocą których skonstruowano sekwencje konsensusowe

Nr sekw.	Wartości wybranych parametrów [%]		
	Delecja	Identyczność	Homologia
1 (opt.)	49,32 (~50)	69,86 (~70)	28,77 (~30)
2	28,77	69,86 (~70)	28,77 (~30)
3	89,04	69,86 (~70)	28,77 (~30)
4	49,32 (~50)	28,77	28,77 (~30)
5	49,32 (~50)	89,04	28,77 (~30)
6	49,32 (~50)	69,86 (~70)	19,18
7	49,32 (~50)	69,86 (~70)	64,38

Tłustym drukiem wyróżniono wartości, które są inne niż domyślne wartości optymalne. Skrót „opt." odnosi się do ustawień domyślnych, z założenia zapewniających konstruowanie optymalnej sekwencji konsensusowej.

Tak sporządzone sekwencje konsensusowe wykorzystano do przeszukania białkowych baz danych (GenBank CDS translations + PDB + SwissProt + PIR + PRF) w celu wyselekcjonowania sekwencji homologicznych do ludzkiej β-spektryny erytrocytarnej. Jednocześnie była to weryfikacja użyteczności i prawidłowej konstrukcji każdej z sekwencji konsensusowych. Przeszukiwanie dokonane było przy użyciu wersji białkowej programu BLAST (blastp).

Zwracano przy tym uwagę nie tylko na ilość znalezionych sekwencji, lecz także na spójność wyselekcjonowanej grupy białek, wartościowaną za pomocą parametrów: *score* i *E-value*.

W celu poprawności wykonanej analizy niezbędne było usunięcie wprowadzonych wcześniej delecji, wynikających z dopasowania sekwencji.

Wyniki przeszukiwania przedstawia tabela 2.

5. Dyskusja

Uzyskane rezultaty przeszukiwania białkowych baz danych w celu odnalezienia jak największej liczby sekwencji homologicznych do β-spektryn jednoznacznie wykazały istotny wpływ wyboru kwerendy (w tym przypadku sekwencja konsensusowa) na jakość i ilość wyników. Sekwencje konsensusowe skonstruowane w oparciu o różne parametry progowe istotnie zmieniają wydajność procesu przeszukiwania baz danych, co uwidocznione jest poprzez parametry: *score* i *E-value*. Wartości te niosą ze sobą istotne

informacje dotyczące uzyskanych wyników. Dla danej wartości oceny S (*score*), wartość E (*E-value*) odpowiada liczbie różnych dopasowań o wartości S, które mogą być wynikiem czysto przypadkowego podobieństwa. Innymi słowy, im większa jest wartość S, tym niższe E, czyli tym mniejsze prawdopodobieństwo, że takie podobieństwo może być czysto przypadkowe. Bardzo małe wartości E oznaczają wysoką istotność statystyczną podobieństwa, natomiast większe E oznacza podobieństwo, które może być przypadkowe i nieistotne. Wartość E równa lub bliska 0 oznacza sekwencje identyczne lub prawie identyczne. Wartości E większe od 0,1 zwykle oznaczają mało znaczące podobieństwo. Wartość ta zależy między innymi od długości sekwencji (im krótsze sekwencje, tym większe prawdopodobieństwo znalezienia przypadkowego podobieństwa, tym większe musi więc być S, by podobieństwo było znaczące) oraz od wielkości i składu przeszukiwanej bazy danych (od nich zależy dystrybucja losowych podobieństw).

Tab. 2. Wartości parametrów *score* i *E-value* wyników przeszukiwania białkowych baz danych za pomocą siedmiu sekwencji konsensusowych skonstruowanych na bazie różnych parametrach progowych

Nr sekw.	Ilość nowych sekwencji*	Ilość sekwencji o wartości E=0 na 100 otrzymanych	*Score* [bits]		*Score* dla sekwencji o E=0	
			Min.	Max.	Min.	Max.
1 (Opt.)	19	51	356	1293	1186	1293
2	24	49	355	1094	713	1293
3	24	49	354	1127	707	1127
4	24	50	354	1092	637	1092
5	23	50	354	1094	639	1094
6	25	52	377	1193	716	1193
7	19	45	322	1450	736	1450

Skrót „Opt." oznacza sekwencja konsensusową zbudowaną przy zadeklarowanych domyślnych wartościach parametrów.

* Po przeszukaniu białkowych baz danych za pomocą uzyskanych sekwencji konsensusowych, w porównaniu z wynikami przeszukiwania tych samych baz za pomocą sekwencji wzorcowej (P11277).

Jak widać na przykładzie otrzymanych wyników (tab. 2), najlepsze rezultaty w przeszukiwaniu baz danych dało zastosowanie sekwencji konsensusowej skonstruowanej na podstawie domyślnych, najbardziej optymalnych parametrów programu Protein Calculator. Odnalezione sekwencje charakteryzują się wysokim podobieństwem (najwyższe w grupie wartości *score*), a wśród odnalezionych sekwencji ponad połowa wykazuje najwyższy stopień nieprzypadkowej identyczności (E=0). Te wyniki wskazują na prawidłowe wzajemne dopasowanie sekwencji oraz prawidłowo skonstruowaną na podstawie takiego dopasowania sekwencję konsensusową.

Badania nad sekwencjami konsensusowymi dają nowe możliwości analizy zmienności ewolucyjnej całych rodzin białkowych. Stanowią one swoiste uśrednienie cech sekwencyjno-strukturalnych białek, dając ogólny obraz danej rodziny białek homologicznych. Pozwalają one na stworzenie profili rodzin białkowych, co może być pomocne zarówno przy identyfikacji nowych członków danej rodziny, jak i przy modelowaniu i przewidywaniu całych struktur. Sekwencje konsensusowe zatem otwierają nowe pole badań proteomicznych, umożliwiając wnikliwą i bardziej precyzyjną analizę porównawczą białek.

Literatura cytowana

1. Beck K. A., Nelson W. J.: A spectrin membrane skeleton of the Golgi complex, Biochem. Biophys. Acta., 1404: 153-160, 1998.
2. Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J.: Basic local alignment search tool, J. Mol. Biol., 215: 403-410, 1990.
3. Altschul S. F., Madden T. L., Schäffer A. A., Zhang J., Zhang Z., Miller W., Lipman D. J.: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res., 25: 3389-3402, 1997.
4. Górecki A., Leluk J., Lesyng B.: A Java-implementation of a genetic semihomology algorithm (GEISHA), and its applications for analyses of selected protein families, 29th FEBS Congress, 26 June - 1 July 2004, Warsaw, Poland, Eur. J. Biochem., 271, Supplement, 30, 2004.
5. Leluk J.: A New Algorithm for Analysis of the Homology in Protein Primary Structure, Computers and Chemistry, Vol. 22, No 1: 123-131, 1998.
6. Leluk J.: Regularities in mutational variabillity in selected protein families and the Markovian model of amino acid replacement, Computers and Chemistry, 24: 659-672, 2000.
7. Leluk J.: A non-statistical approach to protein mutational variability, BioSystems, 56: 83-93, 2000.
8. Leluk J., Hanus-Lorenz B., Sikorski A. F.: Application of genetic semihomology algorithm to theoretical studies on various protein families, Acta Biochimica Polonica, Vol. 48, No. 1/2001.
9. Leluk J., Konieczny L., Roterman I.: Search for structural similarity in proteins, Bioinformatics, Vol. 19, No. 1, 2003.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 121-124.

IMPLEMENTACJA OSZACOWANIA STOPNIA PODOBIEŃSTWA BIAŁEK NA POZIOMIE STRUKTURY PIERWSZORZĘDOWEJ ORAZ KONSTRUKCJA DRZEW FILOGENETYCZNYCH DLA BIAŁEK HOMOLOGICZNYCH

AN APPROACH TO SEQUENCE SIMILARITY SIGNIFICANCE ESTIMATION

ELŻBIETA GAJEWSKA*, JACEK LELUK**

**Wydział Matematyki i Nauk Informacyjnych, Politechnika Warszawska, Warszawa, egajewska@grid.icm.edu.pl*
***Interdyscyplinarne Centrum Modelowania Matematycznego i Komputerowego, Uniwersytet Warszawski, ul. Żwirki i Wigury 93, 02-089 Warszawa, lulu@icm.edu.pl*

Streszczenie. Oszacowanie stopnia podobieństwa sekwencji odgrywa istotną rolę w porównywaniu sekwencji. Aby prawidłowo określić stopień podobieństwa, należy uwzględnić następujące kryteria:
- względna zawartość identycznych pozycji,
- długość porównywanych sekwencji,
- rozmieszczenie pozycji identycznych wzdłuż porównywanego łańcucha,
- rodzaj aminokwasów zajmujących pozycje konserwatywne (dla białek),
- genetyczne i strukturalne współzależności aminokwasów występujących w odpowiadających sobie pozycjach porównywanych sekwencji (dla białek).

Podejście statystyczne opracowane przez Karlina i Altschula pozwala wyodrębnić na tyle duże podobieństwo, by można je było nazwać pokrewieństwem. Nasze rozwiązanie, bazujące na technologii Java, wykorzystuje te same postulaty, wzbogacając dostępną informację o dalsze szczegóły wynikające z dopasowania sekwencji.

Nasze podejście wyszukuje sekwencje podobne do sekwencji wzorcowej, w których następnie oszacowany zostaje stopień rozproszenia pozycji identycznych. Identyczności położone blisko siebie w sekwencji są punktowane wyżej niż rozproszone bloki pozycji identycznych, przeplatane odcinkami nieidentycznymi. Co więcej, rozważana grupa sekwencji może być zawężana, by wyodrębnić znaczące pokrewieństwo. W tym celu należy:
- zadeklarować wartość progową wartości E-value dla dalszych obliczeń,
- wyodrębnić sekwencje o zbliżonej i niskiej wartości E-value,
- wyodrębnić sekwencje podobnej długości,
- wyodrębnić sekwencje, dla których rozkład pozycji identycznych jest podobny.

Wszystkie sekwencje sklasyfikowane jako wzajemnie spokrewnione stanowią dane wejściowe do konstrukcji molekularnego drzewa filogenetycznego. Wzajemny dystans ewolucyjny obliczany jest poprzez analizę dopasowanych par sekwencji z grupy we wszystkich możliwych kombinacjach. Konstrukcja drzewa opiera się na obliczanych wartościach dla konkretnej pary sekwencji w porównaniu do maksymalnego możliwego wyniku dla tej samej względnej zawartości pozycji identycznych oraz długości sekwencji.

Praca wykonana została w ramach grantu MAMBA (Centre of Excellence for Multi-scale Biomolecular Modelling, Bioinformatics and Applications) Projekt nr QLRI-CT-2002-90383.

Słowa kluczowe: porównywanie sekwencji, kryteria podobieństwa, parametry progowe identyczności

Abstract. Problem of similarity significance estimation is crucial in sequence comparison work. In order to assess sequence similarity several parameters should be taken into account. They are:
- the percentage of identity,
- the length of compared sequences,
- distribution of the identical positions along the aligned sequences,
- the type of units occupying conservative positions (in proteins),
- genetic or structural relationships of the units at corresponding positions (in proteins).

Statistical approach presented by Karlin and Altschul [1] proposes a model that distinguishes biologically relevant similarities. Our Java application makes use of these formulas but additionally provides more detailed information concerning the alignments.

Our approach searches for similar sequences to the query sequence then estimates the distribution of identical positions. Identities which are located in direct mutual neighborhood are scored higher than regular distributed identity blocks separated by non identical blocks. Furthermore an examined group of sequences may be filtered for selecting the actual and univocal homology. This can be accomplished by :
- applying the threshold expect (E) value for computation,
- selecting the sequences for which E-values are close,
- selecting the sequences of similar length,
- selecting the sequences that reveal compact blocks of identical positions.

All sequences classified as homologous were used to construct the molecular phylogenetic tree by means of pairwise sequence comparison. Phylogenetic tree construction is based on the scores calculated for the most extreme theoretical distribution cases versus scores calculated for particular sequence pair.

This work was supported by MAMBA (Centre of Excellence for Multi-scale Biomolecular Modelling, Bioinformatics and Applications) Project No. QLRI-CT-2002-90383.

Key words: sequence comparison, similarity criteria, identity treshold parameters

Metoda porównywania sekwencji

Prawidłowe określenie stopnia podobieństwa sekwencji aminokwasowych oraz nukleotydowych jest fundamentalnym zadaniem każdej teoretycznej analizy porównawczej. Klasyczne podejście odwołuje się do ustalenia progowej wartości prawdopodobieństwa losowego trafienia na sekwencje wykazujące stopień podobieństwa uznany za istotny. Im prawdopodobieństwo uzyskania podobnego wyniku jest wyższe, tym mniejsza jest szansa, że obserwowane podobieństwo jest znaczące. Jednym ze sposobów przeprowadzania obliczeń jest wygenerowanie losowej grupy sekwencji na podstawie permutacji jednej z sekwencji. Druga, niezmieniana sekwencja, jest porównywana z wygenerowaną grupą i analizowane są wyniki o wartościach jakościowo wyższych niż wyniki uzyskane dla początkowych sekwencji wyjściowych.

Karlin i Altschul [1] zaprezentowali model statystyczny, który pozwala na oszacowanie stopnia podobieństwa sekwencji. Model ten zależny jest od wybranych sekwencji, a także od użytej bazy danych. Model wymaga następujących początkowych założeń:

1. każdy znak sekwencji jest niezależny od innych,
2. sekwencje są nieskończenie długie,
3. dopasowane sekwencje nie zawierają przerw.

Za pomocą modelu matematycznego można przedstawić liczbę oczekiwanych sekwencji, które można znaleźć przeszukując konkretną bazę danych, a których wynik dopasowania jest nie gorszy niż założony wynik *X*. Obecnie większość aplikacji (FASTA [2], [3], BLAST [4], [5], ClustalW [6]) do szacowania stopnia podobieństwa sekwencji wykorzystuje podejście statystyczne Karlina i Altschula. Podają one parametry *P-value* oraz *E-value*. *P-value* oznacza prawdopodobieństwo, że uzyskany wynik nie jest lepszy niż losowy. *E-value* jest oczekiwaną liczbą sekwencji, dających wynik nie gorszy od założonego, uzyskanych w wyniku przeszukiwania konkretnego zbioru danych. Ustalono, że dobrym oszacowaniem progu homologii są wartości *P-value* $\leq 10^{-10}$ oraz *E-value* ≤ 0.1.

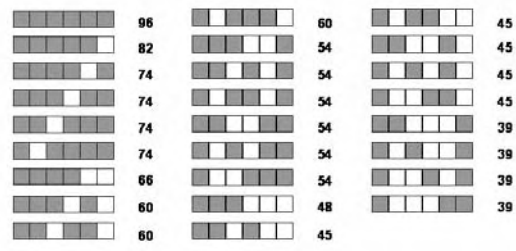
Alfabet aminokwasów oraz nukleotydów nie jest równocenny. W procedurze dopasowania sekwencji należy uwzględnić szereg różnych kryteriów, a sam proces zmienności ewolucyjnej nie może być interpretowany, jako proces markowowski [7]. Uwzględnienie wszystkich niezbędnych kryteriów oszacowania istotności wyniku jest niezbędnym warunkiem skuteczności przeprowadzonej analizy porównawczej. Podstawowymi czynnikami decydującymi o stopniu podobieństwa porównywanych sekwencji są:

1. procentowy udział pozycji identycznych,
2. długość sekwencji,
3. rozmieszczenie pozycji identycznych wzdłuż porównanych sekwencji,
4. rodzaj aminokwasów zajmujących pozycje konserwatywne (dla białek),
5. genetyczne i strukturalne współzależności aminokwasów występujących w odpowiadających sobie nieidentycznych pozycjach porównywanych sekwencji (dla białek).

Metody statystyczne są obecnie najpowszechniej stosowanymi metodami w szacowaniu stopnia podobieństwa porównywanych sekwencji, jednak nie uwzględniają wszystkich wymienionych wyżej parametrów. Model Karlina i Altschula [1] zakłada, że każda jednostka łańcucha jest *niezależna* od innych. Metody statystyczne bazują na rozkładzie wyniku w grupie sekwencji uzyskanych poprzez *permutację* sekwencji pobranej z bazy. Metody budujące drzewa filogenetyczne zakładają wzajemną *niezależność* jednostek tworzących sekwencje. Wszystkie powyższe fakty wskazują, że rozmieszczenie identyczności wzdłuż sekwencji jest pomijane.

Stopień rozproszenia pozycji identycznych i podobnych wzdłuż dopasowanych sekwencji jest istotny ze względu na oddziaływania, które wystąpić mogą pomiędzy pozycjami bliskimi sobie. Stabilizują one strukturę drugorzędową białka. Oddziaływania występują również między resztami odległymi od siebie względem numerów pozycji w łańcuchu, co jest efektem zwinięcia łańcucha białkowego w przestrzeni trójwymiarowej. Należy jednak podkreślić rangę bliskich, zgrupowanych identyczności z uwagi na ich pierwotny wpływ, zanim łańcuch białkowy osiągnie ostateczną, zwiniętą formę.

Prezentowana aplikacja wykorzystuje metodę szacowania stopnia podobieństwa zaproponowaną we wcześniejszych pracach [7], [8], [9], [10]. Operuje ona na dwóch sekwencjach białkowych lub nukleotydowych. Po dokonaniu ich wzajemnego dopasowania, każdej parze pozycji identycznych można przypisać wartość 1, a każdej nieidentycznej parze, lub przerwie – wartość 0. Utworzony w ten sposób łańcuch skanowany jest ramką, która zlicza zgrupowane identyczności. Ramka przesuwa się wzdłuż sekwencji co 1 pozycję, a rozmiar okna obejmuje sześć pozycji.



Rys. 1. Zasady punktowania zgrupowanych identyczności w oknie obejmującym sześć pozycji (identyczności oznaczone czarną barwą)

Ramka obejmująca sześć sąsiadujących pozycji ma za zadanie wykryć jedynie identyczności rozlokowane blisko siebie, które są zgrupowane w liczbie przynajmniej trzech na sześcioelementowym odcinku (rys. 1). Każdej identyczności przypisana jest waga 12 punktów. Jeśli ramka zawiera sześć identyczności, wynik powiększony jest dodatkowo o 24 punkty (maksymalne zgrupowanie identyczności). Trzem identycznościom rozlokowanym w najmniej korzystnym układzie przyznawane są dodatkowe 3 punkty.

Konstrukcja drzew filogenetycznych

Istnieje kilka metod konstruowania drzew filogenetycznych. W pracy rozważano te, które bazują na wyliczonej odległości (różnicy) między sekwencjami.

Clustering oznacza grupowanie obiektów posiadających podobne cechy. Do skonstruowania drzewa buduje się macierz odległości. Na przekątnych mamy, oczywiście, odległości równe 0 i nie są one brane pod uwagę w dalszych rozważaniach. Następnie wyszukiwana jest odległość minimalna. Ma ona identyfikować parę takich obiektów, które są najbliższe sobie pod względem przyjętej miary odległości.

Para ta łączona jest z domniemanym wspólnym przodkiem. Jeśli minimalna odległość występuje jednokrotnie w macierzy, para najbliższych osobników stanowi jedynych potomków węzła nadrzędnego. Natomiast jeśli występuje zbiór *k* par minimalnych odległości w macierzy oraz jeśli tworzą one podgraf pełny, można je wszystkie zgrupować w jednej gałęzi. Jeśli zgrupowano *k* podgałęzi, należy zgrupować *k* kolumn i wierszy w jeden poprzez uśrednianie z pozostałymi. Rząd macierzy zmniejsza się wtedy o *k-1*. Postępujemy tak iteracyjnie, aż rząd macierzy będzie równy jeden. Metoda ta nie podaje żadnej informacji o przodkach zgrupowanych gałęzi. Ilustruje zależności i relacje między gałęziami. Klasycznie ujęty algorytm konstruuje tzw. „ultrametric trees", czyli dla każdej trójki węzłów dwie najdłuższe gałęzie są jednakowej długości. Oznacza to, że wszystkie odgałęzienia zgrupowane w jednym węźle mają taką samą długość.

Metoda *Neighbour joining* [11] jest również metodą iteracyjną konstrukcji drzew filogenetycznych oraz znajdowania długości gałęzi. W każdym kroku dwa najbliższe (według pewnej metryki *d*) węzły są łączone i zastępowane jednym. Wynikowym drzewem jest nieukorzenione drzewo binarne. Utworzone poprzez algorytm drzewo jest addytywne, tzn. posiadające własność (1) $\forall_{x,y,u,v}$:

$$d(x,y)+d(u,v)\le \max(d(x,u)+d(y,v),d(x,v)+d(y,u))$$

Dla konstrukcji wykorzystuje się macierz odległości d_{ij}, dla której $d_{ii}=0$, ale zakładając wektory pomocnicze odległości:

$$M_{ij}=d_{ij}-r_i-r_j$$

Następnie, wybierając takie *i* oraz *j*, dla których M_{ij} jest najmniejsze, wybierany jest nowy węzeł *u*. Odległości od węzła *u* do jego potomków *i, j* wyliczane są następująco:

$$d(u,i)=(d_{ij}+r_i-r_j)/2$$

oraz

$$d(u,j)=(d_{ij}+r_j-r_i)/2$$

Odległości sumują się do d_{ij}, co zgodne jest z założeniem, a długości gałęzi obrazują odległość od pozostałej grupy wierzchołków. Po utworzeniu nowego wierzchołka, podobnie jak w poprzedniej metodzie dokonuje się aktualizacji odległości każdego z pozostałych wierzchołków do nowo powstałego, lecz z pewną modyfikacją:

$$\forall_{k\ne i,k\ne j}\, d_{ku}=(d_{ik}+d_{jk}-d_{ij})/2$$

Algorytm ten może być zastosowany jedynie do macierzy odległości, które spełniają warunek (1). Niestety, nie można oczekiwać, by rzeczywiste sekwencje białkowe czy nukleotydowe spełniały ten warunek.

Testowanie skostruowanych drzew

W pracy zaimplementowane zostały metody bazujące na znajdowaniu odległości między sekwencjami. Wykorzystanie metody szacowania istotności porównanych sekwencji opartej na dystrybucji pozycji identycznych pozwoliło zobrazować i zweryfikować przyjętą metodę. Celem było również stwierdzenie wpływu czynnika rozproszenia na istotność i tworzenie drzew filogenetycznych. Jako porównanie ze stworzoną aplikacją wykorzystano narzędzie ClustalW [6]. ClustalW dopasowuje sekwencje, a także oferuje konstrukcję drzewa filogenetycznego metodą *Neighbour Joining*. Przy konstruowaniu drzew filogenetycznych rozważane były dwie metody. Grupowanie iteracyjne (*Clustering*), czyli poszukiwanie gałęzi, dla których odległość między nimi jest najmniejsza oraz *Neighbour Joining*.

W celu wyznaczenia *odległości między sekwencjami* wykorzystano metodę ramki przesuwającej się wzdłuż sekwencji. Zliczany w ten sposób wynik (*score*) był pomniejszony o kary za przerwy (*gapPenalty*) ustalone przy użyciu ClustalW. Następnie wyliczona wartość podlegała ustandaryzowaniu. Parametr ten nie byłby miarodajny w porównaniu do sekwencji o różnych długościach. Wyliczany jest wynik, jaki można uzyskać, wykorzystując metodę zliczania zgrupowanych identyczności dla 100% identyczności (*100identity*). Ostatecznie odległość obliczana jest jako:

$$1-\frac{score-gapPenalty}{100\,identity}$$

Kara za otwarcie przerwy odpowiada 10-krotności punktacji za identyczność, wydłużenie przerwy odpowiada jednej identyczności. Metoda przyznaje za 6 pozycji identycznych w ramce (maksymalna ilość) premię równą dwum dodatkowym identycznościom, natomiast najmniejsza nagroda za zgrupowaną identyczność wynosi 25% wartości pojedynczej identyczności.

Konsekwencją przyjętego sposobu rozumowania jest zmiana wagi porównywanych sekwencji. Rozważmy dwie hipotetyczne pary sekwencji (*a*,*b*) oraz (*c*,*d*), każda sekwencja ma długość *L*. Przy porównaniu każdej z par mamy *n* identyczności oraz *L-n* pozycji nieidentycznych. Pary te różnią się między sobą rozproszeniem pozycji identycznych wzdłuż sekwencji. W przypadku pary (*a*,*b*) mamy maksymalną liczbę punktów dodatkowych (brak przerw), oznacza to, że wszystkie identyczności muszą być zgrupowane w spójnym obszarze. Natomiast druga para ma tak rozmieszczone pozycje identyczne, że nie otrzymuje żadnych punktów dodatkowych (również brak przerw). Oczywiście jest to możliwe tylko wtedy, gdy $n\le L/3$, ponieważ każde trzy identyczności w dowolnym sześcioelementowym odcinku dają dodatkowe punkty. Wynik dla pary (*a*,*b*), jeśli $n>4$, obliczany jest ze wzoru:

$$W_1=1-\frac{n+(n-5)\cdot 2+\alpha}{L+(L-5)\cdot 2}$$

Gdzie α jest dodatkową punktacją na skraju obszaru spójnego i wynosi około 7-krotnej punktacji za identyczność. Wynik dla pary (*c*,*d*) wynosi:

$$W_2=1-\frac{n}{L+(L-5)\cdot 2}$$

Jeśli oznaczymy procent identyczności $x=\frac{n}{L}$

$$W_1=1-\frac{3n-10+\alpha}{3L-10}=1-\frac{3x-\frac{10+\alpha}{L}}{3-\frac{10}{L}}$$

dla dużych *L* zapis przedstawia się następująco:

$$W_1 \approx 1 - x$$

Tymczasem dla drugiej pary (*c*,*d*)

$$W_2 = 1 - \frac{xL}{3L - 10} = 1 - \frac{x}{3 - \frac{10}{L}} \approx 1 - \frac{x}{3}$$

Oznacza to, że parametr rozproszenia może mieć duże znaczenie dla porównywania sekwencji. Im większy procent identyczności, tym mniejszy jest procentowy udział „premii" za dystrybucję. Przy dużym procencie identyczności właściwie nie ma problemu w oszacowaniu, że sekwencje są homologiczne. Pomimo że granice są płynne, można spróbować oszacować istotność podobieństwa na podstawie udziału procentowego identyczności. Takie oszacowanie przedstawia tab. 1.

Tab. 1. Procent identyczności porównywanych sekwencji białkowych a ich podobieństwo

Procentowy udział identyczności	*Istotność – homologia*
<18%	wątpliwa
18-25%	nie można odrzucić homologii
>25%	prawdopodobna

Metoda punktowania identyczności w powiązaniu z ich rozproszeniem wpływa szczególnie istotnie na pary sekwencji o ilości pozycji identycznych mniejszych niż 30%. Jest to więc ta grupa sekwencji, która budzi duże wątpliwości.

Dwie metody, które zostały zaimplementowane (*Clustering* oraz *Neighbour Joining*), porównane były z drzewami filogenetycznymi generowanymi przez ClustalW. Aby móc porównać drzewa, wprowadzono odległość pomiędzy drzewami. Żeby wyliczyć odległość między drzewami, oba muszą być drzewami ukorzenionymi i posiadać taki sam zbiór odgałęzień. Zdefiniowano odległość między wierzchołkami jako liczbę węzłów na ścieżce z jednego węzła, do drugiego. Długość gałęzi została pominięta, ponieważ najistotniejsze było podobieństwo struktury drzewiastej (filografu). Jeśli dwa odgałęzienia są przyłączone do jednego węzła odległość między nimi wynosi 1. Odległość między węzłem a nim samym wynosi 0. Można zbudować kwadratową macierz odległości między węzłami, przy czym kolejność wyboru liści na kolumny i wiersze musi być taka sama dla wszystkich macierzy odległości wszystkich drzew. Ponieważ macierz jest symetryczna, wystarczy przechowywać tylko wartości macierzy trójkątnej powyżej przekątnej (na przekątnej odległości są zerowe). Odległością między drzewami będzie suma odległości między węzłami:

$$d(A,B) = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(A_{ij} - B_{ij} \right)$$

Odległości między drzewami z obu aplikacji były mniejsze dla algorytmu drugiego. Dlatego też został on wybrany jako odpowiedniejszy do konstruowania drzew filogenetycznych.

Aplikacja

Stworzone narzędzie w celu konstrukcji drzewa filogenetycznego bazuje na grupie sekwencji. Grupa ta może być dobrana bezpośrednio przez użytkownika lub też wygenerowana przez program na podstawie jednej wprowadzonej sekwencji. W tym drugim przypadku, korzystając z usługi FASTA [2], [3], tworzona jest grupa sekwencji podobnych. Grupa ta poddawana jest wzajemnemu dopasowaniu każdej z par sekwencji przy użyciu narzędzia ClustalW [6]. Ponieważ aplikacja korzysta z usług sieciowych, niezbędne jest połączenie z siecią Internet do poprawnego działania aplikacji. Wynik dopasowania poddany jest obliczeniom mającym za zadanie wyznaczyć odległość między sekwencjami uwzględniającą rozproszenie pozycji identycznych wzdłuż porównawanych sekwencji, bazując na algorytmie semihomologii genetycznej [7], [8], [9], [10]. Drzewo filogenetyczne konstruowane jest opisaną wyżej metodą *Neighbour Joining.*

Aplikacja napisana jest w języku Java i przy pisaniu jej korzystano z dostępnych bezpłatnie usług sieciowych. Umożliwia wizualizację utworzonego drzewa w dwóch wersjach: drzewa ukorzenionego oraz nieukorzenionego. Skonstruowane drzewo może zostać zapamiętane w standardowym formacie tekstowym wykorzystywanym przez aplikacje, takie jak np. ClustalW oraz PHYLIP.

Literatura cytowana

1. Karlin S., Altschul S. F.: Methods for assessing the statistical significance of molecular sequence features by using genera,l Proc. Natl. Acad. Sci. USA, 87: 2264-2268.
2. Pearson W. R., Lipman D. J.: Improved Tools for Biological Sequence Comparison, Proc. Natl. Acad. Sci. USA, 85: 2444-2448, 1988.
3. Pearson W. R.: Rapid and sensitive sequence comparison with FASTP and FASTA, Mehods Enzymol., vol. 183, pp. 63-98. 1990
4. Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J.: Basic local alignment search tool, J. Mol. Biol., 215: 403-410, 1990.
5. Altschul S. F., Madden T. L., Schäffer A. A., Zhang J., Zhang Z., Miller W., Lipman D. J.: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res., 25: 3389-3402, 1997.
6. Thompson J. D., Higgins D. G., Gibson T. J.: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice, Nucleic Acids Research, 22: 4673-4680, 1994.
7. Leluk J.: Regularities in mutational variability in selected protein families and the Markovian model of amino acid replacement, Computers & Chemistry, 24: 659-672, 2000.
8. Leluk J.: A New Algorithm for Analysis of the Homology in Protein Structure, Computers & Chemistry, 22: 123-131, 1998
9. Leluk J.: A non-statistical approach to protein mutational variability. BioSystems, 56: 83-93, 2000.
10. Leluk J., Grabiec, M.: Sequence similarity estimation and correlated mutations in selected protein families. I. An approach to protein sequence similarity estimation. Ist Summer School on "Parallel Computing in Biomolecular Simulations", Sept. 1-3, 2001, Gdańsk, Poland; Abstracts L-5, 2001.
11. Saitou N., Nei M.: The Neighbor-Joining Method: A New Method For Reconstructing Phylogenetic Trees, Mol. Biol. Evol. 4(4):406-25, 1987.

BIOINFORMATYK GENOMIK
ALGORYTM WYZNACZANIA NUKLEOTYDOWYCH SEKWENCJI KONSERWATYWNYCH
FARMAKOLOG
PROJEKTOWANIE LEKÓW

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 125-128.

ZMIENNOŚĆ MUTACYJNA W HOMOLOGICZNYCH RODZINACH KINAZ

THE MUTATIONAL VARIABILITY WITHIN HOMOLOGOUS KINASE FAMILIES

JACEK KUŚKA[*], JACEK LELUK[**], BOGDAN LESYNG[**,#]

[*]*Wydział Biologii, Uniwersytet Warszawski, Warszawa, jackk@grid.icm.edu.pl*
[**]*Interdyscyplinarne Centrum Modelowania Komputerowego i Matematycznego, Uniwersytet Warszawski, Warszawa, lulu@icm.edu.pl*
[#]*Zakład Biofizyki, Wydział Fizyki, Uniwersytet Warszawski, Warszawa, B.Lesyng@icm.edu.pl*

Streszczenie. Przeanalizowano cztery rodziny kinaz (heksokinazy, receptorowe kinazy tyrozynowe, niereceptorowe kinazy tyrozynowe-JAK, cAMP-zależne kinazy). Sekwencje pobrano z bazy danych SWISS-PROT. Przeprowadzono dopasowanie sekwencji oraz skonstruowano sekwencje konsensusowe dla każdej rodziny przy użyciu algorytmu semihomologii genetycznej. Do identyfikacji pozycji kontaktujących się wzajemnie w strukturze trzeciorzędowej zastosowano zasoby bazy danych Ligand-Protein Contact and Contacts of Structural. Analiza oraz wizualizacja zmienności dla trzech grup zastała przeprowadzona na podstawie wzorcowej struktury trzeciorzędowej dla każdej grupy, pobranej z Protein Data Bank. Strukturę trzeciorzędową dla grupy JAK wygenerowano w Interdyscyplinarnym Centrum Modelowania Matematycznego i Komputerowego Uniwersytetu Warszawskiego. Przeanalizowano występowanie mutacji sprzężonych w każdej badanej rodzinie enzymów przy użyciu programu Corm, opracowanego również w Interdyscyplinarnym Centrum Modelowania w Warszawie.
Dla każdej przebadanej rodziny stwierdzono, że:

- W poszczególnych pozycjach występowało od 1 do 17 różnych aminokwasów.
- Pozycje tworzące centrum katalityczne wykazują niski stopień zmienności, chociaż pozycje bezpośrednio sąsiadujące z nimi mogą odznaczać się wyższą zmiennością.
- N-końcowy peptyd liderowy charakteryzuje się bardzo wysokim stopniem zmienności.
- Pozycje znajdujące się w bezpośrednim wzajemnym kontakcie posiadają identyczny lub zbliżony poziom zmienności.
- Pozycje wykazujące sprzężenie mutacyjne są odległe od siebie w strukturze pierwszorzędowej, lecz położone blisko siebie w strukturze trzeciorzędowej. To zjawisko potwierdza przyjęty ogólnie mechanizm kompensacji mutacji. Jednak część zaobserwowanych mutacji sprzężonych odnosi się do pozycji znacznie oddalonych od siebie w strukturze trzeciorzędowej. Mechanizm powstawania tego typu sprzężeń mutacyjnych nie jest jeszcze poznany i wymaga dalszych badań.

Wyniki naszych badań generalnie potwierdzają hipotezę dotyczącą korelacji pomiędzy poziomem zmienności pozycji będących w kontakcie. Odnosi się to do twierdzenia, że struktura przestrzenna, jak i funkcja białek są bardziej konserwatywne niż struktura pierwszorzędowa.

Praca wykonana została w ramach grantu MAMBA (Centre of Excellence for Multi-scale Biomolecular Modelling, Bioinformatics and Applications) Projekt nr QLRI-CT-2002-90383.

Słowa kluczowe: kinazy, zmienność mutacyjna, mutacje sprzężone.

Abstract. Four homologous kinase families (hexokinases, RTKs, JAKs, cAMP dependent kinases) were investigated with respect to the mutational variability within each family. The sequences were obtained from Swiss-Prot database. Multiple sequence alignments as well as consensus sequences were created and verified with the aid of the algorithm of genetic semihomology. The application of the Ligand-Protein Contact and Contacts of Structural Units database helped to select the residues which are in mutual contact in the secondary as well as tertiary structures. The analysis and visualization of variability in three analyzed groups was acquired with the use of three-dimensional models obtained from the Protein Data Bank. The tertiary structure of JAKs was predicted at Interdisciplinary Centre for Mathematical and Computational Modelling, Warsaw University. Correlated mutations occurring in each kinase family were identified and investigated with the Corm program developed also at ICM UW. For all four analyzed groups of kinases it was observed that:

- particular positions were occupied by 1 to 17 amino acid residues,
- residues involved in catalytic centre formation reveal relatively low variability level, although residues surrounding the centre may reveal some variability,
- N-terminal leader peptide is highly variable,
- residues that are in mutual contact show identical or very similar level of variability,
- positions involved in correlated mutations may be distant at the level of primary structure but usually stay close in three dimensional structures. This phenomenon confirms the compensation mechanism. However some correlated residues are sited far away form each other both in primary structure and in tertiary structure. The phenomenon of dispersed correlation is still unclear and is the subject of further investigations.

The results of this work generally confirm the hypothesis concerning the correlation mechanism for mutually interacting residues being in close distance. It refers to the statement that protein structure and function which are more significantly conservative than the primary structure itself.

This work was supported by MAMBA (Centre of Excellence for Multi-scale Biomolecular Modelling, Bioinformatics and Applications) Project No. QLRI-CT-2002-90383.

Key words: kinases, mutational variability, correlated mutations

1. Wstęp

Kinazy, inaczej zwane fosfotransferazami, są enzymami katalizującymi reakcję przeniesienia grupy fosforanowej. Donorem grupy fosforanowej jest najczęściej ATP. Kinazy są enzymami, które biorą udział w niezliczonych procesach zachodzących w komórce. Najczęściej są to reakcje w przekazywaniu sygnału, innymi procesami wymagającymi obecności kinaz są: metabolizm węglowodanów (heksokinazy), tłuszczy oraz nukleotydów, witamin i kofaktorów. Wyżej wymienione enzymy są odpowiedzialne za prawidłowy przebieg cyklu komórkowego, wiele z nich jest onkogenami, ich mutacje mogą mieć katastrofalny wpływ na komórkę i na cały organizm. Kinazy charakteryzują się różnorodnością w budowie, posiadają również zróżnicowane substraty, które mogą zostać ufosforylowane. Gatunek *Homo sapiens* posiada 492 poznane kinazy. Stanowi to 1,5% wszystkich zidentyfikowanych białek kodowanych w genomie człowieka. U *Drosophila melanogaster* kinazy zajmują 2,5% kodu całego genomu.

W niniejszej pracy obiektem badań były 4 homologiczne grupy kinaz: heksokinazy, receptorowe kinazy tyrozynowe, kinazy tyrozynowe Janus (JAK) oraz kinazy białkowe zależne od cAMP.

2. Materiały i metody

2.1. Bazy danych sekwencji i struktur białkowych

Sekwencje aminokwasowe kinaz pobierano z baz danych Swiss-Prot oraz Nrbd95, dostępnych pod adresem internetowym http://www.expasy.org/sprot/. Z baz danych pobrano przy użyciu programu BLAST [1], [2] ponad 100 sekwencji wskazujących na pokrewieństwo do poszczególnych badanych rodzin. Do analizy wykorzystano 50 sekwencji o najwyższym stopniu podobieństwa, ponieważ w pozostałych zróżnicowanie było zbyt duże. Pakiet programów BLAST służący do wyszukiwania sekwencji podobnych dostępny jest pod adresem http://www.ncbi.nlm.nih.gov/BLAST/. Struktury trzeciorzędowe pobierano z bazy Protein Data Bank (http://www.rcsb.org/pdb/). Dla każdej z analizowanych grup przyjęto sekwencję oraz strukturę wzorcową, będącą podstawą przeszukiwania baz. Ludzka heksokinaza pochodząca z mózgu o kodzie PDB 1HKB reprezentowała grupę kinaz cukrowych, kinazy receptorowe reprezentowała kinaza o kodzie PDB 1IRK, a kinaza o kodzie PDB 1ATP stanowiła wyjściowy wzorzec dla grupy kinaz białkowych zależnych od cAMP (PKA. Przedstawicielem kinaz tyrozynowych JAK była kinaza oznaczona kodem JAK2. Baza PDB nie dysponuje strukturą tego (ani żadnego innego) przedstawiciela owej grupy enzymów, skorzystano więc w tym przypadku z wygenerowanej teoretycznie stuktury JAK3 w Interdyscyplinarnym Centrum Modelowania Matematycznego i Komputerowego UW, której sekwencja jest bardzo podobna do sekwencji JAK2).

Trójwymiarowe modele wzorcowych białek posłużyły do wizualizacji zmienności oraz do wizualizacji korelacji mutacyjnych w każdej z badanych rodzin.

2.2. Baza kontaktów wewnątrz- i międzycząsteczkowych Ligand-Protein Contacts & Contacts of Structural Units

Użytecznym źródłem informacji potrzebnej do analizy mechanizmów zmienności w odniesieniu do wzajemnego kontaktu zmiennych pozycji jest baza danych Ligand-Protein Contacts & Contacts of Structural Units, dostępna pod adresem http://bip.weizmann.ac.il/oca-bin/lpccsu. Z tej bazy danych pozyskiwano informacje na temat aminokwasów występujących we bezpośrednim kontakcie lub bliskim sąsiedztwie w analizowanym białku. We wspomnianej bazie danych istnieją dane tylko dla białek, których struktury trójwymiarowe są zdeponowane w bazie PDB. Z tego powodu nie przeprowadzano analizy kontaktów dla grupy kinaz tyrozynowych JAK. Na podstawie tych danych skonstruowano mapy kontaktów dla trzech spośród czterech analizowanych grup.

2.3. Algorytm semihomologii genetycznej

Algorytm semihomologii genetycznej [3], [4], [5], [6] został wykorzystany do prawidłowego wzajemnego dopasowania sekwencji homologicznych dla wszystkich czterech grup enzymów oraz do skostruowania sekwencji konsensusowej.

W odróżnieniu od narzędzi bazujących stochastyczne macierze wymiany aminokwasów w sekwencjach homologicznych, algorytm ten opiera się na trójwymiarowym diagramie relacji genetycznych występujących pomiędzy aminokwasami. Dostarcza przy tym bogatej informacji na temat możliwych mechanizmów zmienności i przebiegu zmienności ewolucyjnej.

Obecnie dostępne są dwie aplikacje realizujące analizę podobieństwa sekwencji w oparciu o algorytm semihomologii genetycznej - GEISHA (ICM UW) [7], [8] oraz Semihomology (FQS Poland). Znajdują się one pod adresami: http://bioinfo.icm.edu.pl/geisha/ oraz http://www.fqspl.com.pl/sh/

2.4. Specjalistyczne programy komputerowe

Prowadzone badania wspomagane były przez użycie szeregu aplikacji opracowanych w Interdyscyplinarnym Centrum Modelowania Matematycznego i Komputerowego UW. Wśród wykorzystanych programów były:

- Consensus Constructor – program służący do szczegółowej analizy zestawów dopasowanych wzajemnie sekwencji oraz konstruowania sekwencji konsensusowych na podstawie tych dopasowań. Dostępny jest pod adresem: http://tarawa.icm.edu.pl/lulu/.
- Corm – program służący do wyszukiwania, analizy i opisu mutacji sprzężonych występujących w rodzinach białek homologicznych. Jego najnowsza aktualna wersja dostępna jest pod adresem: http://tarawa.icm.edu.pl/agorecki/corm.
- WebLab ViewerLite 5.0 – program do wizualizacji białkowych struktur trzeciorzędowych. Program jest produktem firmy Accelrys Inc. (http://www.accelrys.com/).

3. Wyniki i dyskusja

3.1. Dopasowanie sekwencji homologicznych

Dla każdej grupy analizowanych enzymów utworzono osobne zestawienie sekwencji. Każde z nich składało się z 50 sekwencji otrzymanych z bazy danych SWISS-PROT lub NRBD95. Są to sekwencje wykazujące największy stopień podobieństwa do użytej sekwencji wzorcowej. Homologiczne sekwencje dopasowano przy zastosowaniu programów realizujących algorytm semihomologii genetycznej (GEISHA i Semihomology). Znaczna część pracy (szczegółowa weryfikacja dopasowań) przeprowadzona została manualnie, również zgodnie z zasadami algorytmu semihomologii genetycznej.

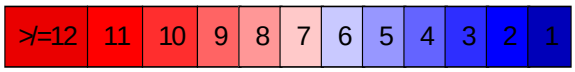
Dla każdej grupy dopasowanych sekwencji homologicznych skonstruowano sekwencję konsensusową, która była zbiorczym podsumowaniem charakteryzującym każdą z rodzin.

Za pozycje konserwatywne uznano takie, które obsadzone były przez ten sam rodzaj aminokwasu przynajmniej w 20 sekwencjach (40%). Przeanalizowano przy tym możliwe relacje genetyczne (pojedyncza tranzycja/transwersja) tych aminokwasów z innymi aminokwasami występującymi w odpowiadających im pozycjach. Również w pozycjach niekonserwatywnych przebadano potencjalne relacje genetyczne między występującymi tam aminokwasami. W zestawieniach dopasowanych sekwencji aminokwasy konserwatywne oraz semihomologiczne wyróżniono w odpowiedni sposób graficzny.

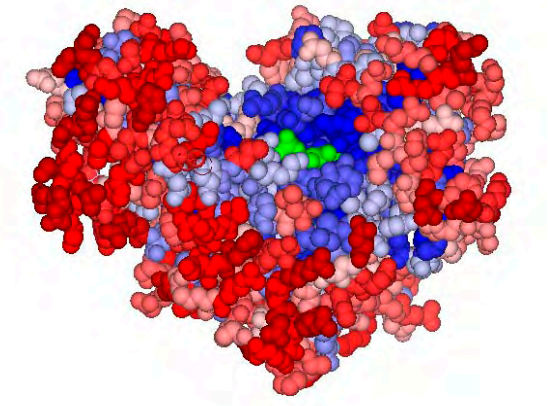
Szczegółowa weryfikacja dopasowania homologicznych sekwencji w oparciu o algorytm semihomologii genetycznej pozwoliła na wykluczenie późniejszych ewentualnych błędów związanych z możliwością uznawania niewłaściwych pozycji jako korespondujących w znaczeniu ewolucyjnym.

3.2. Profile zmienności mutacyjnej

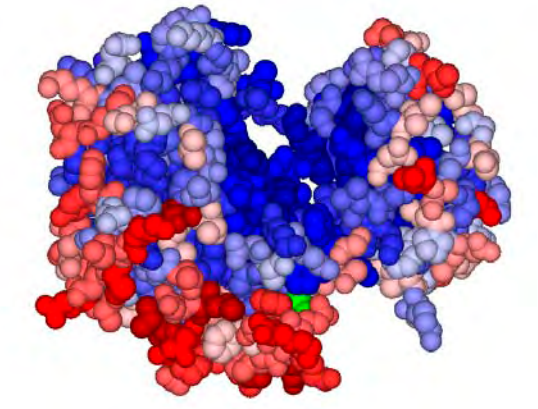
Analiza zmienności mutacyjnej w odpowiadających sobie pozycjach sekwencji homologicznych wykazała, że zakres zmienności w badanych grupach kinaz zawiera się w przedziale od 1 do 17 różnych reszt aminokwasowych. Nałożenie 12-stopniowej skali zmienności (rys. 1) na reprezentacyjne struktury trzeciorzędowe uwidocznił obszary cząsteczki o podobnym – wysokim lub niskim stopniu zróżnicowania (rys. 2-5). Zgodnie z przewidywaniami ogólny profil zmienności wskazywał na wysoką konserwatywność elementów tworzących centrum katalityczne. Jednakże istniały obszary położone bardzo blisko istotnych funkcjonalnie elementów cząsteczki, których stopień zmienności mutacyjnej był wysoki. Również pewnym zaskoczeniem była zmienność wyższa od spodziewanej niektórych pozycji bezpośrednio zaangażowanych w wiązanie ligandów, jonu wapnia, czy grup fosforanowych (rys. 6). Wysoce zmienne fragmenty regionów istotnych funkcjonalnie świadczą o ich odpowiedzialności za wąską specyficzność konkretnego enzymu, wyróżniającą go spośród innych przedstawicieli tej samej rodziny homologicznej. Obszary silnie konserwatywne natomiast odpowiadają cechom wspólnym dla całej rodziny i stanowią element definiujący rodzinę jako grupę białek różniącą się pewnymi cechami od innych rodzin. Niezwykle wysoki stopień zmienności wykazują fragmenty N-końcowe cząsteczek, odgrywające rolę wysoko specyficznych peptydów liderowych.



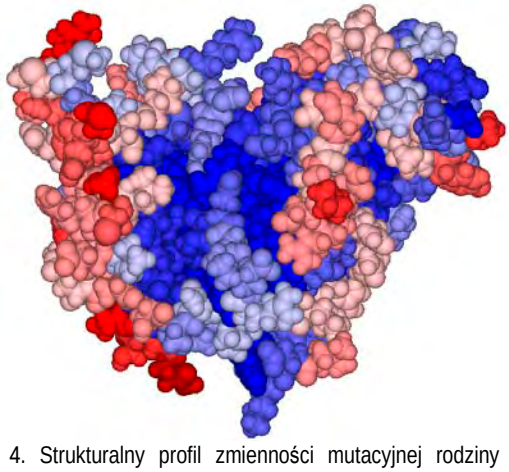
Rys. 1. Skala zmienności – cyfry oraz odpowiadające im kolory oznaczają ilość aminokwasów występujących w danej pozycji w uzgodnionego zestawienia sekwencji



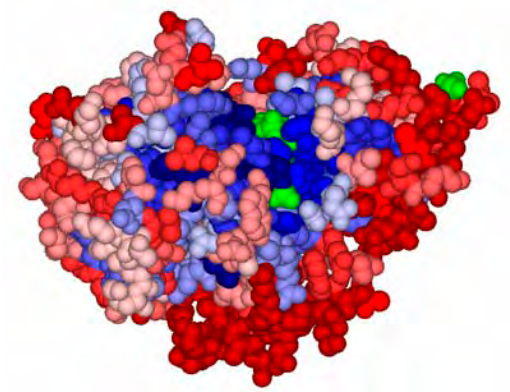
Rys. 2. Strukturalny profil zmienności mutacyjnej rodziny heksokinaz. Skala zmienności nałożona została na strukturę heksokinazy 1HKB. Kolorem zielonym oznaczono substrat



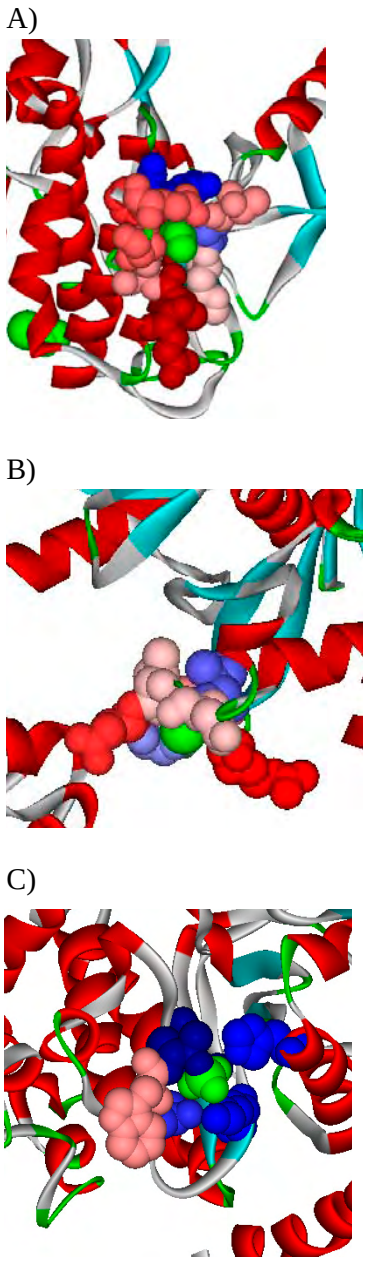
Rys. 3. Strukturalny profil zmienności mutacyjnej rodziny receptorowych kinaz tyrozynowych. Skala zmienności nałożona została na strukturę heksokinazy 1IRK. Kolorem zielonym oznaczono substrat



Rys. 4. Strukturalny profil zmienności mutacyjnej rodziny kinaz tyrozynowych z grupy Janus. Skala zmienności nałożona została na strukturę JAK3 (wygenerowaną w ICM UW)



Rys. 5. Strukturalny profil zmienności mutacyjnej rodziny kinaz cAMP-zależnych. Skala zmienności nałożona została na strukturę 1ATP



Rys. 6. Zmienne pozycje w istotnych funkcjonalnie obszarach kinaz. A) Kinaza 1IRK. Kolor zielony oznacza ligand otoczony resztami o różnym stopniu zmienności; B) Kinaza 1HKB. Wapń (kolor zielony) otoczony resztami o różnym stopniu zmienności; C) Kinaza 1ATP (PKA). Grupa fosforanowa (kolor zielony) otoczona pozycjami o różnym stopniu zmienności

Identyfikacja i szczegółowa charakterystyka regionów zmiennych i konserwatywnych stanowi cenne źródło informacji do wykorzystania podczas projektowania odrębnych grup inhibitorów skierowanych albo przeciwko całej rodzinie homologicznych kinaz, albo specyficznie przeciwko jednemu konkretnemu enzymowi

3.3. Mutacje sprzężone

Analiza wzajemnego położenia pozycji wykazujących sprzężenie mutacyjne pokazała, że w badanych rodzinach kinaz sprzężenia te na ogół odpowiadają przyjętej obecnie hipotezie na temat mechanizmu mutacji sprzężonych [9]. Mianowicie, w ogromnej większości obserwowanych sprzężeń stwierdzono bliskie położenie bądź bezpośredni kontakt między resztami występującymi w tych pozycjach. Stąd wniosek, że w tych grupach białek mutacje sprzężone związane są z utrzymaniem stałej i stabilnej struktury przy zachodzącej – nawet intensywnie – zmienności mutacyjnej. Niemniej jednak, stwierdzono występowanie grup pozycji sprzężonych, oddalonych od siebie znacząco i tworzących tzw. klastry rozproszone. Wzajemne oddziaływanie reszt obsadzających takie pozycje jest wykluczone i takie rodzaje sprzężeń wymagają dalszych badań w celu wyjaśnienia ich powstawania i znaczenia w funkcjonowaniu cząsteczki białkowej.

Literatura cytowana

1. Altschul S. F., Gish W., Miller, W., Myers, E. W., Lipman D. J.: Basic local alignment search tool, J. Mol. Biol., 215: 403-410, 1990.
2. Altschul S. F., Madden T. L., Schäffer A. A., Zhang J., Zhang Z., Miller W., Lipman D. J.: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res., 25: 3389-3402, 1997.
3. Leluk J.: A New Algorithm for Analysis of the Homology in Protein Primary Structure, Computers and Chemistry, Vol. 22, No. 1: 123-131, 1998.
4. Leluk J.: Regularities in mutational variabillity in selected protein families and the Markovian model of amino acid replacement, Computers and Chemistry, 24: 659-672, 2000.
5. Leluk J.: A non-statistical approach to protein mutational variability, BioSystems, 56: 83-93, 2000.
6. Leluk J., Hanus-Lorenz B., Sikorski A. F.: Application of genetic semihomology algorithm to theoretical studies on various protein families, Acta Biochimica Polonica, Vol. 48, No. 1/2001.
7. Górecki A., Leluk J., Lesyng B.: A Java-implementation of a genetic semihomology algorithm (GEISHA), and its applications for analyses of selected protein families, 29th FEBS Congress, 26 June - 1 July 2004, Warsaw, POLAND, Eur. J. Biochem., 271, Supplement, 30, 2004.
8. Górecki A., Leluk J., Lesyng B.: GEISHA – a novel Java application of the genetic semihomology algorithm for comparative studies of protein sequences, The Ninth Annual International Conference on Research in Computational Molecular Biology (RECOMB), MIT Cambridge, MA, USA, Abstracts, in press, 2005.
9. Neher E.: How frequent are correlated changes in families of protein sequences? Proc. Natl. Acad. Sci. USA, 91: 98-102, 1994.

BIOINFORMATYK GENOMIK *MODEL WYZNACZANIA KONSENSUSOWEJ SEKWENCJI NUKLEOTYDOWEJDLA RODZIN BIAŁKOWYCH*	⟷	*BIOINFORMATYK PROTEOMIK* *MODEL KONSENSUSOWEJ STRUKTURY TRZECIORZĘDOWEJ*
MOŻLIWOŚĆ ROZSZERZENIA PROJEKTU NA INNE RODZINY BIAŁEK		MOŻLIWOŚĆ ROZSZERZENIA PROJEKTU NA INNE RODZINY BIAŁEK

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 129-132.

ANALIZA DANYCH Z MIKROMACIERZY DNA METODY, NARZĘDZIA

ANALYSIS OF DNA MICROARRAY DATA METHODS AND TOOLS

HENRYK MACIEJEWSKI[*], ŁUKASZ KONARSKI[*], ANNA JASIŃSKA[**], MARIA DRATH[**]

[*]*Instytut Cybernetyki Technicznej Politechniki Wrocławskiej, henryk.maciejewski@pwr.wroc.pl, lukasz.konarski@pwr.wroc.pl*
[**]*Laboratorium Genetyki Raka, Instytut Chemii Bioorganicznej PAN, Poznań, aniaj@rose.man.poznan.pl, maridrath@yahoo.com*

Streszczenie. Praca poświęcona jest analizie danych z badań ekspresji genów prowadzonych za pomocą mikromacierzy DNA. Mikromacierze DNA są nową techniką wykorzystywaną przez biologów, a coraz powszechniej również przez lekarzy, umożliwiającą równoczesny pomiar poziomu ekspresji wielu (np. kilkudziesięciu tysięcy) genów badanego organizmu. Eksperymenty za pomocą mikromacierzy prowadzą do powstawania dużych zbiorów danych, na podstawie których możliwe jest poszukiwanie zależności pomiędzy aktywnością genów a np. stanami chorobowymi organizmu, badanie funkcji genów, a także budowanie modeli umożliwiających przewidywanie reakcji na terapie i in. Poszukiwanie tego typu zależności, istotnych dla biologów i lekarzy, możliwe jest dzięki zastosowaniu zaawansowanych metod *data mining*, korzystających zarówno z technik statystycznych, jak i rozpoznawania wzorców (*pattern recognition*),a także specjalizowanych narzędzi komputerowych. Specyfika danych mikromacierzowych oraz formułowanych dla nich zadań analitycznych dała nowe impulsy rozwoju metod analizy danych oraz narzędzi komputerowych.
W pracy przedstawiono zadania analizy danych formułowane dla potrzeb analizy danych z badań mikromacierzowych. Szczególną uwagę zwrócono na zadania grupowania w celu poszukiwania grup genów przejawiających podobne profile ekspresji oraz grup podobnych obserwacji (próbek). Pokazano wpływ ustawień zastosowanych algorytmów analitycznych na wyniki, w ten sposób ilustrując trudności interpretacji wyników analizy danych z badań mikromacierzowych. W pracy omówiono również wybrane narzędzia wspomagające analizę danych mikromacierzowych, a także rozwijane ostatnio narzędzia umożliwiające efektywną pracę interdyscyplinarnych zespołów badaczy biologów – statystyków – informatyków, pracujących nad analizą danych z mikromacierzy DNA.

Słowa kluczowe: mikromacierz DNA, data mining, grupowanie

Abstract. The purpose of the work is to present application of data mining techniques for analysis of results from DNA microarray experiments. A challenging problem of finding groups of genes with similar expression patterns and groups of samples with similarly expressed genes is discussed in greater detail.
Software tools that expedite data analysis by building a framework for efficient cooperation of interdisciplinary teams consisting of biologists, statisticians, and data mining specialists are also presented.

Key words: DNA microarray, data mining, clustering

1. Mikromacierz DNA – zasada działania

Mikromacierz DNA umożliwia wyznaczenie poziomu ekspresji bardzo wielu (do kilkudziesięciu tysięcy) genów jednocześnie. Oligonukleotydowa mikromacierz DNA zbudowana jest w następujący sposób: na płytce szklanej naniesione są sondy badające poziom ekspresji poszczególnych genów. Sonda jest łańcuchem nukleotydów o sekwencji charakterystycznej dla określonego genu, umieszczonym w wielu egzemplarzach w określonym polu mikromacierzy. Obecne technologie umożliwiają wykonanie mikromacierzy o następujących przykładowych parametrach: rozmiar pola zajmowanego przez sondę wynosi 11 μm; długość łańcucha nukleotydów umieszczonego w sondzie = 25 (tzw. sondy 25-merowe), liczba sond na płytce – rzędu 10^5.

Przebieg badania aktywności genów za pomocą mikromacierzy jest następujący:

- odpowiednio przygotowana próbka badanej tkanki umieszczana jest na macierzy,
- mRNA zawarte w badanej tkance hybrydyzuje czyli łączy się z komplementarnymi sondami na macierzy. Ilość mRNA, które dołączyło się do poszczególnych sond jest miarą ilości mRNA w próbce, a tym samym miarą aktywności danego genu,

– materiał związany na sondach jest barwiony odpowiednią farbą fluorescencyjną, co umożliwia określenie ilości mRNA, związanego z poszczególnymi sondami. Następnie macierz umieszczana jest w skanerze, który odczytuje poziom fluorescencji poszczególnych sond. Na podstawie wartości odczytanych ze skanera wyliczany jest poziom ekspresji poszczególnych genów. Poziom ten jest wyliczany, a nie bezpośrednio zmierzony, gdyż na każdy gen przypada grupa (np. 11 w macierzach Affymetrix) kopii danej sondy, poza tym oprócz sond mierzących koncentrację mRNA są również sondy mierzące poziom tła – tzw. hybrydyzację niespecyficzną, która jest również uwzględniana przy wyznaczaniu ekspresji genu.

Oprócz macierzy oligonukleotydowych opracowanych przez firmę Affymetrix stosowane są również macierze starszego typu, tzw. macierze cDNA. Podstawową cechą różniącą oba typy macierzy jest gęstość upakowania sond na płytce macierzy – większa w przypadku macierzy oligonukleotydowych. Wynika ona z różnych technologii wykonania macierzy (w przypadku oligonukleotydowych – sondy wykonywane są z zastosowaniem technik fotolitograficznych, zaś w przypadku macierzy cDNA – nanoszone mikropipetą). Badania ekspresji genów przeprowadzane za pomocą macierzy cDNA są badaniami porównawczymi, dającymi w wyniku poziom ekspresji próbki w stosunku do próbki odniesienia.

2. Zadania analizy danych z mikromacierzy DNA

Badanie mikromacierzowe umożliwia uchwycenie aktywności genów w danym stanie organizmu lub tkanki. Badania mikromacierzowe przeprowadza się zatem w celu analizy zależności pomiędzy aktywnością genów a stanem rozumianym jako „zdrowy" – „chory na określoną chorobę" o podłożu genetycznym, a także np. w trakcie rozwoju choroby itp.

Część przykładowego zbioru danych z serii badań mikromacierzowych pokazana jest w tabeli 1. Wiersze tabeli odpowiadają genom (ściślej – sondom reprezentującym geny, kilka różnych sond może reprezentować jeden gen), zaś kolumny – poszczególnym eksperymentom, tzn. badanym próbkom. W analizowanym dalej w pracy przykładzie, próbki pobrane były od kilku osób zdrowych i kilku chorych na syndrom łamliwego chromosomu X – chorobę o podłożu genetycznym.

Tab.1. Postać danych z serii badań mikromacierzowych

Probe	Próbka 1	Próbka 2	Próbka 3
1007_s_at	687.2	709.7	604.5
1053_at	870.6	479.1	613.4
117_at	194.6	326.8	183.7
121_at	1710.8	2217.9	1514.1
1255_g_at	41.2	106.6	75.1
1294_at	984.4	1044.2	745.6
1431_at	93	125.6	41.9

Uwaga: W przedstawionym w tabeli przykładowym wyniku badań ekspresji genów wiersze odpowiadają sondom testującym geny (nazwa sondy macierzy U133A Affymetrix w kolumnie probe), zaś kolumny – poszczególnym próbkom.

Na podstawie serii badań mikromacierzowych definiowane są następujące zadania analizy danych/*data mining*:

– Zadanie selekcji genów najbardziej związanych z przynależnością danych próbek (kolumn w macierzy jak w tab. 1) do pewnych znanych z góry klas (np. chorzy – zdrowi). Większość metod obecnie używanych w tym celu bada korelacje pojedynczych genów ze zmienną określającą przynależność do klasy; jednak prawdopodobnie powinny one zostać rozbudowane o mechanizm wykrywania grup genów, które wspólnie powodują przynależność do danej klasy.
– Budowa modeli predykcyjnych pozwalających na określenie na podstawie wyniku badania mikromacierzowego ryzyka wystąpienia pewnej choroby bądź podatności na pewną terapię.
– Poszukiwanie grup obserwacji o podobnych profilach ekspresji genów – zadanie to wymaga zastosowania algorytmów klasteryzacji po kolumnach macierzy, jak w tab. 1. Mimo tego, że w wielu zastosowaniach jest ono źle uwarunkowane ze względu na wymiarowość macierzy, opisywane są przykłady zrealizowania tego zadania z powodzeniem. W ten sposób wykryto na przykład warianty pewnych chorób nowotworowych nierozróżnialnych na podstawie objawów klinicznych, o zupełnie różnej podatności na terapie.
– Poszukiwanie grup genów o podobnych profilach ekspresji dla różnych próbek. Zadanie to polega na klasteryzacji po wierszach macierzy jak w tab. 1. Wykrycie grup genów o podobnych zmianach w ekspresji, dla np. grupy próbek od osób zdrowych i chorych, ma duże znaczenie dla badania funkcji genów w organizmie – gen o nieznanej roli w pewnych procesach biologicznych, wykazujący profil ekspresji podobny do grupy genów o znanej roli, może być podejrzewany o udział w pewnym procesie, np. chorobowym.
– Badanie, czy pewne geny wykazują ekspresję charakterystyczną dla pewnych chorób.

3. Poszukiwanie genów o podobnych profilach ekspresji

Pokazany przykład obliczeniowy został zrealizowany na grupie 7 badań mikromacierzowych, w których reprezentowane były osoby zdrowe i chore na chorobę genetyczną – syndrom łamliwego chromosomu X. Celem analizy jest znalezienie grup genów o podobnych profilach ekspresji na siedmioelementowym wektorze obserwacji. Badania były przeprowadzone za pomocą mikromacierzy HGU133A firmy Affymetrix, zawierających ok. 22 tysiące sond.

Standardowe [4] podejścia do klasteryzacji opierają się na trzech grupach algorytmów:

– *k*-means,
– klasteryzacja hierarchiczna,
– SOM (samoorganizujące mapy Kohonena).

Metoda zastosowana w omawianym przykładzie wykorzystała klasteryzację hierarchiczną, która w wariancie aglomeratywnym polega na:

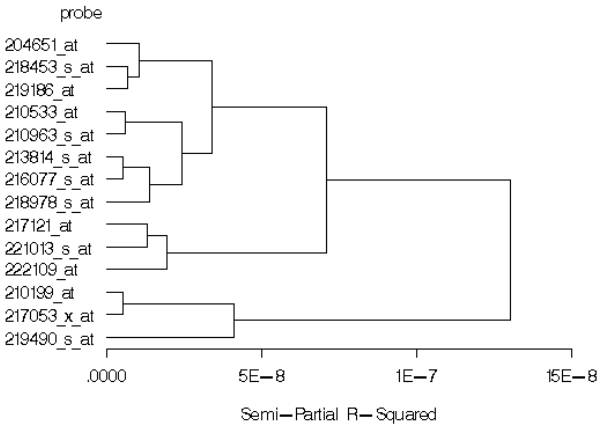
– początkowym potraktowaniu wszystkich obserwacji jako odrębnych jednoelementowych klastrów,
– następnie w kolejnych krokach znajdowane są dwa klastry najbliższe sobie i łączone w jeden klaster,

– proces kończy się po zgrupowaniu wszystkich obserwacji od jednego klastra.

W wyniku klasteryzacji hierarchicznej otrzymywana jest hierarchia klastrów, na podstawie której można uzyskać od jednego klastra do liczby klastrów równej liczbie analizowanych obserwacji.

Wyniki klasteryzacji hierarchicznej najwygodniej jest pokazać w postaci tzw. dendrogramu, tj. wykresu obrazującego hierarchię kolejnych łączonych przez algorytm klastrów, przykład (fragmentu) takiego dendrogramu pokazano na rys. 1.

Wysokość drzewa R2 (rysowana wzdłuż osi poziomej na rys. 1) ma sens odległości między dwoma wcześniejszymi klastrami łączonymi przez algorytm w dane drzewo. Rozcinając drzewo dla ustalonej wartości R2, uzyskuje się podział zbioru punktów (w siedmiowymiarowej przestrzeni) na pewną liczbę klastrów.



Rys. 1. Dendrogram ilustrujący hierarchię kolejno łączonych klastrów

Jednym z podstawowych problemów związanych z analizą danych oraz interpretacją wyników jest fakt, że wyniki klasteryzacji ekspresji genów silnie zależą od przyjętych parametrów algorytmów grupowania oraz wcześniejszego preprocesingu danych (szczególnie normalizacji) [1], [3]. Wprowadzenie normalizacji danych (np. poprzez zapewnienie, żeby każdy z wektorów będących wierszami macierzy jak w tab. 1 miał długość równą 1) powoduje, że szukane będą geny o podobnym profilu ekspresji niezależnie od osiąganego *poziomu* ekspresji.

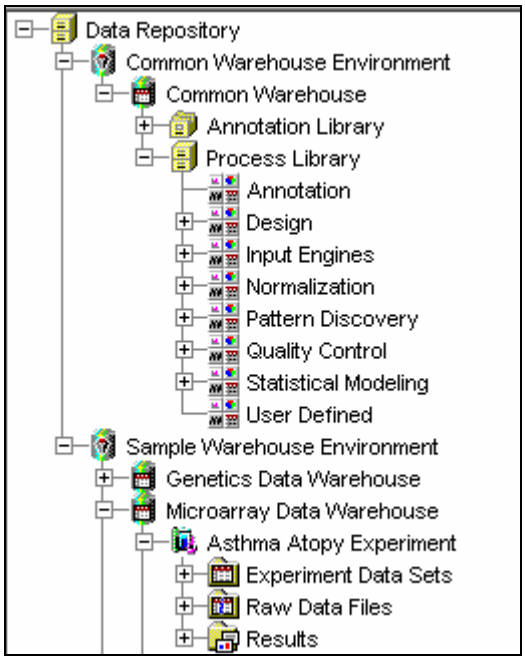
Innym parametrem mającym istotny wpływ na wyniki klasteryzacji jest definicja odległości pomiędzy klastrami. Typowo przyjmowanymi definicjami odległości są np.:

– *Average linkage* – gdzie odległość pomiędzy klastrami jest określana jako średnia odległość par punktów z obu klastrów,
– *Complete linkage* – gdzie odległość między klastrami jest odległością najdalszych punktów z obu klastów,
– Odległość Warda.

Jak pokazano w pracy [2], wpływ przyjętej definicji odległości między klastrami na wyniki klasteryzacji jest istotny. Na przykład, porównując klastry o rozmiarze do kilkudziesięciu genów, formujące się wokół genu FMR1, przy założeniu każdej z powyższych definicji odległości między klastrami, zaobserwowano, że jedynie ok. 30% genów stanowi wspólną część klastrów znalezionych przy założeniu różnych metod. Problem wyboru odległości między klastrami jest problemem otwartym, co ilustruje trudności stosowania metod *unsupervised learning* dla szukania grup genów o podobnych profilach ekspresji.

4. Narzędzia do analizy danych z mikromacierzy DNA – SAS Microarray

SAS Microarray Solution jest nowym narzędziem do analizy i obróbki danych pochodzących z mikromacierzy DNA. Narzędzie jest przeznaczone do pracy w sieci. Posiada wiele możliwości umożliwiających zarządzanie pracą zespołu badawczego. Procedury „obudowane" interfejsem graficznym pozwalają na korzystanie z narzędzia przez zespoły nie mające doświadczenia w języku programistycznym SAS.

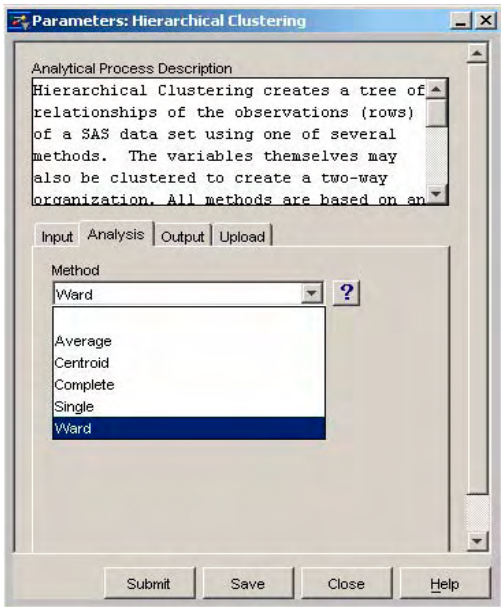


Rys. 2. Ogólny widok aplikacji SAS Microarray wykorzystującej technologię hurtowni danych (data warehouse)

Narzędzie zbudowane jest z wykorzystaniem technologii hurtowni danych (rys. 2) – umożliwia scentralizowane zarządzanie danymi do analizy, wynikami analiz, a także bibliotekami kodów realizującymi różne zadania analityczne. Dane przechowywane są w jednym miejscu na serwerze i są dostępne dla różnych użytkowników. Dostęp możliwy jest dla różnych użytkowników na różnym poziomie, w zależności od nadanych praw dostępu. Narzędzie posiada również duże możliwości w zakresie wersjonowania i prowadzenia historii zmian w projekcie. Wszystko to ma zapewnić sprawną i przejrzystą organizację i zarządzanie projektem.

Drugim, niewątpliwie cennym rozwiązaniem, jakie oferuje SAS Microarray Solution są gotowe skrypty SASowe umożliwiające np. import plików w jednym z formatów generowanych przez skanery do mikromacierzy, a także wykonanie szeregu analiz danych. Lista obsługiwanych formatów zawiera m.in. pliki:

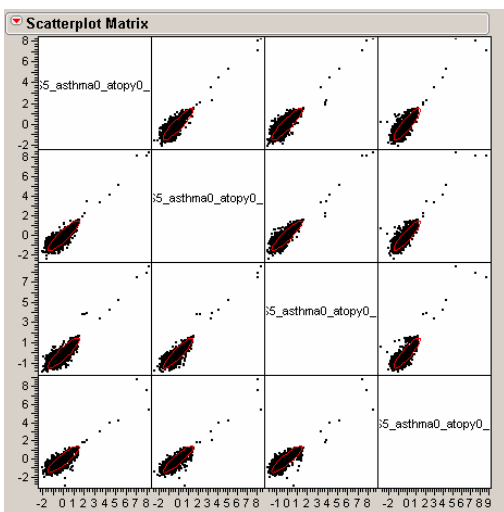
– Affymetriksa
– Agilent
– GenePix
– QuantArray
– ScanAlyze



Rys. 3. Moduł klasteryzacji hierarchicznej – wybór metody liczenia odległości między klastrami

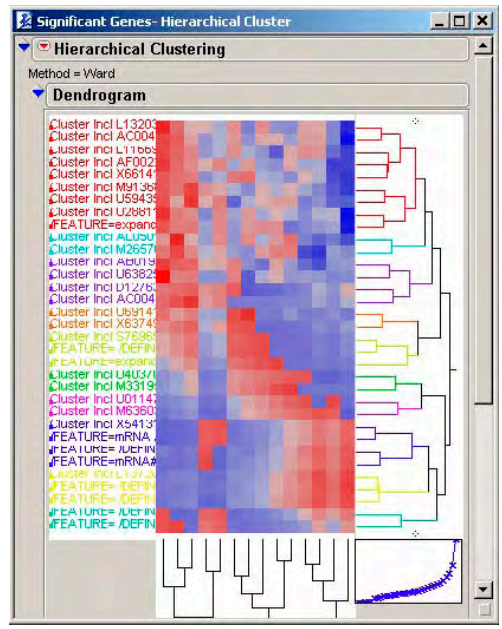
Skrypty te obudowane są czytelnym interfejsem graficznym. Za pomocą interfejsu okienkowego można łatwo i przejrzyście modyfikować parametry procedur, co pozwala uniknąć konieczności opanowania języka programowania SAS. Skrypty te zamieszczone są w bibliotekach i podzielone na kilka kategorii: Design (of Experiments), Input Engines, Normalization, Pattern Discovery, Quality Control, Statistical Modeling oraz User Defined. Rysunek 3 przedstawia interfejs skryptu umożliwiającego klasteryzację hierarchiczną. Okienko interfejsu pozwala na łatwą modyfikację parametrów skryptu, takich jak np. metody liczenia odległości między klastrami (Average, Centroid, Complete, Single, Ward).

Narzędzie umożliwia przeprowadzenie kontroli jakości danych uzyskanych z eksperymentu mikromacierzowego. Przykładowy wynik takiej procedury przedstawiono na rys. 4. Przedstawiony *scatterplot matrix* umożliwia wizualne porównanie wyników uzyskanych z czterech różnych macierzy (porównanie „każdy z każdym"), na których badane próbki odpowiadały tym samym warunkom (np. występowanie pewnej choroby).



Rys. 4. Wyniki procedury kontroli jakości danych z eksperymentu mikromacierzowego – procedura Array Group Correlation

Na rys. 5 przedstawiony został wynik jednej z dostępnych w SAS Microarray Solution procedur analizy danych – klasteryzacji hierarchicznej po wierszach i kolumnach zestawu danych jak na tabeli 1.



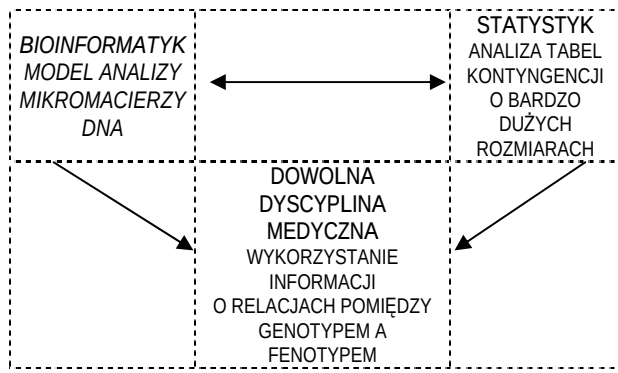
Rys. 5. Przykładowe wyniki klasteryzacji hierarchicznej

4. Podsumowanie

Metody analizy danych z badań mikromacierzowych są obecnie intensywnie rozwijane i trudno jeszcze mówić w tej dziedzinie o „standardowych" algorytmach. Dlatego też istotne jest, żeby narzędzia komputerowe wspomagające analizy danych były otwarte – czyli umożliwiały prostą implementację nowych metod obliczeniowych. Postulat ten spełniają najważniejsze narzędzia zarówno dostępne bezpłatnie (np. pakiet R), jak i komercyjne (takie jak przedstawiony SAS Microarray Solution). Dodatkową funkcjonalnością, jaką oferują narzędzia komercyjne, a niedostępną np. w pakiecie R, jest środowisko pracy grupowej umożliwiające efektywną współpracę interdyscyplinarnych zespołów biologów-genetyków, statystyków i informatyków zaangażowanych w projekty mikromacierzowe.

Literatura cytowana

1. Hoffmann R., Seidl T., Dugas M.: Profound effect of normalization on detection of differently expressed genes in oligonucleotide microarray data analysis, Genome Biology, 2002.
2. Maciejewski H., Jasińska A.: Clustering DNA Microarray Data, 4th International Conference on Computer Recognition Systems, Rydzyna, Advances in Soft Computing, Springer Verlag, 2005.
3. Quackenbush J.: Computational Analysis of Microarray Data. Nature Reviews – Genetics, 2001.
4. Shannon W., Culverhouse R., Duncan J.: Analyzing microarray data using cluster analysis, Pharmacogenomics, 2003.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 133-138.

TWORZENIE I ROZDZIAŁ KOMPLEKSÓW BIAŁEK SUROWICY Z BARWNIKAMI SUPRAMOLEKULARNYMI W ELEKTROFOREZIE DWUKIERUNKOWEJ – PRÓBA OPRACOWANIA KOMPUTEROWEJ TECHNIKI ANALIZY OBRAZU ROZDZIAŁOWEGO

PROTEIN COMPLEXES WITH SUPRAMOLECULAR LIGANDS FORMED AND REVEALED IN TWO-DIMENSIONAL ELECTROPHORESIS. THE SEARCH FOR COMPUTER ANALYSIS OF RESULTS

Bioinformatics

JANINA RYBARSKA[*], PAWEŁ SPÓLNIK[*‡], ANNA DROZD[*], LESZEK KONIECZNY[*], BARBARA PIEKARSKA[*], BARBARA STOPA[*], GRZEGORZ ZEMANEK[*], MARCIN KRÓL[**#], IRENA ROTERMAN-KONIECZNA[**]

[*]*Instytut Biochemii Lekarskiej, Collegium Medicum Uniwersytetu Jagiellońskiego, Kraków*
[**]*Zakład Bioinformatyki i Telemedycyny, Collegium Medicum Uniwersytetu Jagiellońskiego, Kraków*
[#]*Biomolecular Modelling Laboratory, Cancer Research UK, London Research Institute Lincoln's Inn Fields Laboratories, London, UK,* [‡] *pspolnik@cm-uj.krakow.pl*

Streszczenie. Praca przedstawia badania eksperymentalne mające na celu wykrycie zmian w składzie białkowym zmienionej chorobowo surowicy ludzkiej. Dla wykrycia zmian opracowano dwukierunkową technikę elektroforetyczną umożliwiającą tworzenie kompleksów przez barwniki o charakterze ciekłokrystalicznym z białkami. Taśmowe micele barwnika mogą przylegać do szkieletu łańcucha polipeptydowego w obszarach jego ekspozycji, tworząc kompleksy. Warunki elektroforetyczne pozwalają na usunięcie nadmiaru oraz puli słabo związanego z białkami barwnika. Tworzenie kompleksów w elektroforezie pozwala na ujawnienie kompleksów białek z barwnikami, które wędrują szybciej niż białka niezaangażowane w wiązanie barwnika. Nieprawidłowe oraz przejściowo zdestabilizowane natywne białka mogą być podatne na penetrację i wiązanie dużych supramolekularnych ligandów. Do grupy białek wiążących należą białka szpiczakowe oraz serpiny, haptoglobiny i immunoglobuliny, ulegające przegrupowaniu podczas tworzenia kompleksów ze swoimi fizjologicznymi ligandami.
Elektroforeza surowicy prowadzona jest dwuetapowo w prostopadłych względem siebie kierunkach. Pierwszy rozdział jest standardowy. Drugi modyfikowany jest przez kontakt i ewentualną interakcję rozdzielonych uprzednio białek z dodanym barwnikiem. Szybciej wędrujący barwnik, którego punkt nałożenia znajduje się poniżej białek, napotyka w trakcie wędrówki i wyprzedza białko. Białka zdolne do tworzenia kompleksów zabarwiają się w takich warunkach oraz wędrują szybciej od swoich nieskompleksowanych z barwnikiem odpowiedników. Po redukcji barwników dwutioninem sodowym z następowym zabarwieniem błękitem bromofenolowym, można uzyskać kompletny proteinogram rozdzielanej surowicy. Próby opracowania pół-automatycznej i automatycznej analizy uzyskiwanego w elektroforezie rozdziału plam zostały ostatnio podjęte.

Słowa kluczowe: czerwień Kongo, samoasocjacja, serpiny, elektroforeza

Abstract. Two-dimensional agarose electrophoresis was used to create suitable conditions for the formation of protein complexes with supramolecular dyes. Ribbon-shaped dye molecule ligands adhere during complexation to the polypeptide backbone of β-conformation within protein clefts. The electrophoresis helps to remove the dye excess and the dye weakly attached to protein molecules. It simultaneously allows for the exposure of protein-dye complexes migrating faster than proteins not engaged in complexation. Abnormal and transiently destabilized native proteins become susceptible to penetration and binding by large supramolecular dye ligands. This includes many myeloma proteins as well as serpins, haptoglobins and immunoglobulins engaged in their natural complexes.
Electrophoresis of serum proteins is conducted in two steps: the first is a standard electrophoresis while the second (perpendicular run) is modified by the contact and interaction of separated serum proteins (during the first step) with the added dye. The fast migrating dye, for which the starting position is retreated versus to that of proteins, meets and overruns the proteins. The proteins which are susceptible to dye penetration and binding become stained and their migration is accelerated. The complete proteinogram including the binding and non-binding dye proteins may be revealed by standard staining with bromophenol blue after removal of the supramolecular dye. An attempt for semi-automatic or automatic analysis of spots distributed in electrophoresis has been undertaken.

Key words: Congo red, self-assembly, serpins, electrophoresis

1. Wstęp

Praca przedstawia badania eksperymentalne mające na celu wykrycie kompleksów tworzonych przez białka ostrej fazy w surowicy ludzkiej. Kompleksy takie mogą pojawiać się w surowicy w niektórych stanach patologicznych. Opracowanie techniki ujawniającej ilość zaangażowanego w tworzenie kompleksu z naturalnymi ligandem białka pozwoliłoby na właściwą ocenę zagrożenia. Można to uzyskać przez oszacowanie ilości frakcji białka skompleksowanego z ligandem. Pomiar stężenia odpowiedniego białka ostrej fazy, które niekiedy produkowane są w nadmiarze, nieadekwatnie do zapotrzebowania, wydaje się do takiej oceny niewystarczający. Dla wykrycia kompleksów tworzonych przez białka ostrej fazy z naturalnymi ligandami opracowano dwukierunkową technikę elektroforetyczną. Wykorzystuje ona dotychczas słabo poznane zjawisko tworzenia kompleksów przez barwniki bisazowe o charakterze ciekłokrystalicznym z białkami [1], [2], [3], [4]. Zastosowanym w badaniach wskaźnikiem są barwniki z rodziny czerwieni Kongo, które w roztworach wodnych samoasocjują, tworząc taśmowe struktury micelarne zdolne do tworzenia mocnych kompleksów z białkami, pod warunkiem ich niestabilności [5], [6]. Kompleksy te są trwałe, w przeciwieństwie do kompleksów tworzonych przez pojedyncze, małe cząsteczki [7], [8].

Mimo stosowania czerwieni Kongo w naukach medycznych od dawna [9], najczęściej w diagnostycznym teście barwienia tkanek dla wykrywania włókien amyloidowych (w tzw. zmodyfikowanej metodzie Puchtlera [10]), mechanizm jej oddziaływania z białkami nie jest znany. Sugerowano specyficzność wiązania czerwieni Kongo do charakterystycznego dla amyloidu układu łańcuchów polipeptydowych β [11], [12] określanego na podstawie badań dyfrakcji promieni rentgenowskich jako cross-beta [13]. Zaobserwowano jednak, że czerwień Kongo tworzy trwałe kompleksy także z innymi (również rozpuszczalnymi) białkami [14], [15], [16], [17], [18], [19].

Czerwień Kongo jako ligand o supramolekularnym typie oddziaływania z białkami

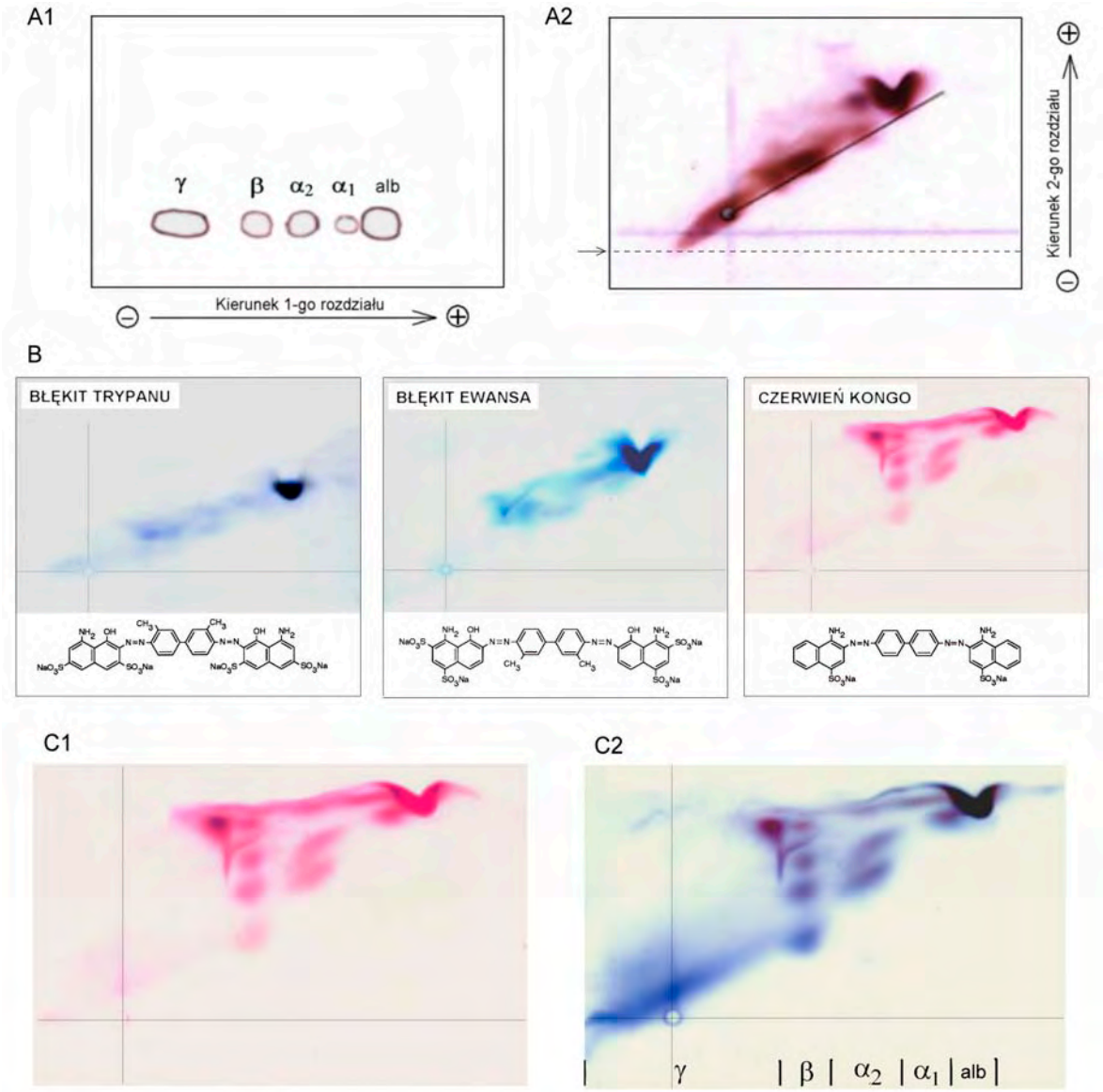
Jedną z teorii wyjaśniających mechanizm interakcji barwników bisazowych z białkami jest założenie, że oddziałują one z nietypowym, bo położonym poza grupą czynną białka miejscem receptorowym i, co najistotniejsze – nie jako pojedyncza cząsteczka a ligand złożony z kilku molekuł uporządkowanych w taśmową strukturę micelarną [3], [20], [21].

Cząsteczki barwników bisazowych o sztywnej, płaskiej cząsteczce o ujemnie naładowanych końcach, oddziałują ze sobą w roztworach wodnych poprzez interakcję pierścieni aromatycznych π, ustawionych prostopadle względem płaszczyzny. Samoasocjując w ten sposób, po osiągnięciu odpowiedniego stężenia, tworzą taśmowe (wstążkowe) micele [22], [23], [24], [25], [26]. Monomery w tego typu układzie nie mogą „schować" centralnych hydrofobowych fragmentów cząsteczki, w wyniku czego niepolarny pas taśmy pozostaje w kontakcie ze środowiskiem [26]. Takie struktury wykazują zdolność adhezyjną, zwłaszcza do polimerów o budowie fibrylarnej (celuloza, amyloid) [27]. Kohezyjne przyłączenie do polimeru daje mocne kompleksy, tym bardziej, że micela taśmowa jest plastyczna i może dobrze dopasować się do jego powierzchni. Oddziaływanie tego typu ma przede wszystkim miejsce poza grupą czynną, choć ligandy z rodziny czerwieni Kongo niekiedy mogą tworzyć kompleksy z białkami w sposób typowy, tj. jako pojedyncze cząsteczki związane w centrach aktywnych białka [17], [28], [29]. Plastyczność miceli barwnika supramolekularnego umożliwia dobre dopasowanie do miejsca receptorowego w białku, a stosunkowo duże siły kohezji między centralnym hydrofobowym pasem miceli barwnika i eksponowanym w miejscu niestabilnym fragmentem beta-płyty zapewniają trwałość kompleksów.

W analizach dwukierunkowych stosowano czerwień Kongo, ponieważ jej szybka wędrówka w polu elektrycznym oraz trwałość tworzonych z białkami kompleksów pozwala na uzyskanie obrazów o najlepszej rozdzielczości (pod względem rozdziału otrzymywanych barwnych plam) spośród testowanych barwników bisazowych (czerwień Kongo, błękit Ewansa i błękit Trypanu) (rys. 1AB). Stwierdzono również, że barwnik ten łatwiej zredukować dwutioninem sodowym do bezbarwnych, dobrze rozpuszczalnych pochodnych, co ułatwia ich usunięcie, a następnie zabarwienie pozostałego na żelu białka błękitem bromofenolowym [31]. Ze względu na wielkość ligandów warunkiem wiązania barwników oddziałujących jako taśma złożona z wielu cząsteczek barwnika jest niestabilność białka o charakterze lokalnym lub globalnym [5], [30], [32], [33]. Zdefiniowanie pojęcia stabilności jest trudne, w praktyce można przyjąć, że w białku istnieją obszary o większej i mniejszej amplitudzie fluktuacji łańcuchów polipeptydowych, a co za tym idzie – zdolność do odkształceń strukturalnych tych obszarów jest różna. Obszary o większej fluktuacji wydają się być łatwiej penetrowalne przez ligandy, które mogą konkurować z łańcuchami polipeptydowymi o miejsce w beta-płycie [7], [33]. Utworzone w ten sposób kompleksy są trwałe, co potwierdzają zarówno badania eksperymentalne, jak obliczenia komputerowe [1], [33].

Warunek niestabilności strukturalnej spełniają niektóre białka monoklonalne obecne w surowicy chorych ze szpiczakiem mnogim (czerwień Kongo mogą wiązać niektóre surowicze monoklonalne immunoglobuliny – IgG i Iga, jak i łańcuchy lekkie obecne w moczu) [4], [5], [30], [31], [32]. Obecność białek monoklonalnych o niestabilnej strukturze może mieć znaczenie diagnostyczne, ponieważ te formy paraprotein zwykle związane są z większym zagrożeniem agregacją i tworzeniem złogów [34], [35], [36].

Innymi białkami, które mogą wiązać w sposób trwały ligandy utworzone przez samoasocjujące barwniki bisazowe, są surowicze inhibitory proteaz serynowych (serpiny) [1], [3], [33]. Modelową cząsteczką tego typu jest α_1 – antyproteinaza, w której istnienie niestabilnej płyty beta jest przystosowaniem do pełnienia funkcji fizjologicznej. Rolą białek z tej rodziny jest tworzenie kompleksów z enzymami proteolitycznymi, które niekiedy mogą pojawić się w surowicy (np. w przebiegu procesów zapalnych). Utworzenie kompleksu enzym-serpina znacząco przyspiesza usuwanie takiej formy z surowicy (czas ten wynosi od kilkunastu godzin dla wolnej formy serpiny do kilku godzin-minut w przypadku kompleksu z enzymem proteolitycznym) [37]. Inne białka z nadrodziny serpin to między innymi: antytrombina, α_1-antychymotrypsyna, α_2-antyplazmina [37], [38]. Wspólny serpinowy motyw strukturalny w postaci pętli oddziałującej z obszarem niestabilnej beta płyty teoretycznie stwarza możliwość wiązania czerwieni Kongo również przez te białka.



Rys. 1. Zasada tworzenia i rozdziału kompleksów białek surowicy z barwnikami bisazowymi w elektroforezie dwukierunkowej. A. Zasada techniki. A1. Schemat dystrybucji frakcji surowicy ludzkiej uzyskanych po pierwszym rozdziale elektroforetycznym. A2. Dystrybucja frakcji po zakończeniu drugiego rozdziału elektroforetycznego. Poziom nanoszenia roztworu barwnika bisazowego zaznaczono linią przerywaną. Linia poprzeczna wskazuje lokalizację białek bez modyfikowania ich mobilności elektroforetycznej barwnikami bisazowymi. B. Dwukierunkowy rozdział tej samej surowicy ludzkiej modyfikowany barwnikami bisazowymi o różnej zdolności do samoasocjacji. Wzory użytych barwników bisazowych przedstawiono w dolnej części obrazu. C. Dwukierunkowy rozdział białek surowicy modyfikowany dodaniem czerwieni Kongo przed drugim rozdziałem. C1. Barwne kompleksy białko-barwnik. C2. Ta sam płytka po redukcji czerwieni Kongo dwutioninem sodu do bezbarwnych pochodnych i następowym zabarwieniu białek błękitem bromofenolowym. Kompleksy widoczne przede wszystkim we frakcjach α i β

Technika elektroforezy dwukierunkowej

Niestabilne białka surowicy mogą tworzyć z ligandami supramolekularnymi kompleksy, które dzięki ładunkowi związanego barwnika można rozróżnić w warunkach elektroforetycznych od białek niezdolnych do oddziaływania z czerwienią Kongo [1], [4]. W celu zbadania właściwości kompleksotwórczych białek przeprowadza się dwukierunkowy rozdział elektroforetyczny surowicy na żelu agarozowym.

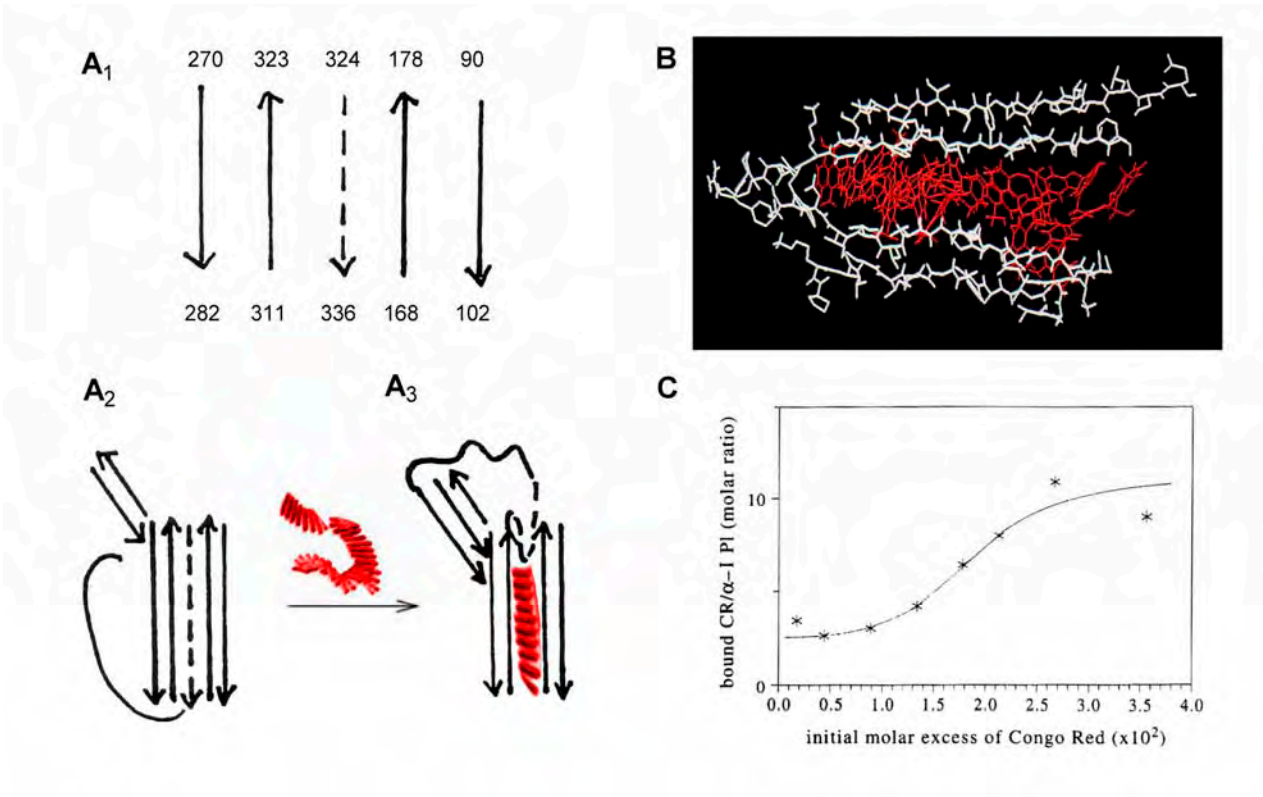
Stosowana metoda elektroforetyczna polega na wykonaniu dwóch prostopadłych względem siebie rozdziałów na tej samej płytce żelu agarozowego (rys. 1A). Pierwszy z rozdziałów jest znanym z klasycznej metody diagnostycznej rozdziałem białek surowicy, w którym otrzymuje się typową dystrybucję frakcji. Drugi prostopadły względem pierwszego rozdział poprzedzony jest naniesieniem barwnika bisazowego (czerwieni Kongo) poniżej pierwszej linii dystrybucji białek. Barwnik bisazowy wędruje szybciej niż białka surowicy. Umożliwia to uzyskanie przejściowego kontaktu i tworzenia kompleksów tylko pod warunkiem dużego powinowactwa. Odbijanie nadmiaru czerwieni Kongo bibułą celulozową, do której związki bisazowe mają wysokie powinowactwo, dodatkowo pozwala rozróżnić trwałe kompleksy białko-barwnik od słabych [1], [4], [31]. Wolną i związaną z białkiem czerwień Kongo można również odróżnić, modyfikując pH. W niskim pH czerwień Kongo jako wskaźnik alkacymetryczny zmienia kolor. Wolny barwnik niebieszczeje szybko, natomiast związane z białkiem cząsteczki miceli są najprawdopodobniej

„chronione" przez białko i w związku z tym obserwuje się znacznie wolniejszy przerzut barwy [1]. Podobnie można interpretować wolniejszą redukcję barwników bisazowych związanych z białkiem podczas działania na płytkę roztworem dwutioninu sodu.

Białka surowicy oddziałujące z czerwienią Kongo

Po przeprowadzeniu rozdziału białek surowicy otrzymuje się na prostokątnej płytce rozmaite układy plam, w których można wyróżnić zidentyfikowane już frakcje białkowe oraz nie do końca poznane kompleksy białek z czerwienią Kongo. Te ostatnie, szczególnie licznie reprezentowane wśród białek o elektroforetycznej ruchliwości beta i alfa, mają do dziś nie określoną wartość diagnostyczną [1], [3], [4] (rys. 1C).

Podejrzewa się zdolność do tworzenia tego typu kompleksów szczególnie przez białka ostrej fazy, które często zawierają duże ilości łańcuchów polipeptydowych o uporządkowaniu beta (rys. 2A).



Rys. 2. Tworzenie kompleksu z micelą czerwieni Kongo przez α_1-antyproteinazę. A. Ujęcie schematyczne budowy i przekształceń β-strukturalnej płyty C serpiny. A1. Numeracja reszt aminokwasowych w łańcuchach polipeptydowych. A2. Forma występująca po hydrolizie wiązania między aminokwasem 336 i 337, na skutek której fragment 323-336 wkleja się do płyty C, tworząc antyrównoległy układ całej płyty. A3. Stabilny kompleks utworzony przez dokowanie miceli czerwieni Kongo w strukturze płyty C. B. Model 7 cząsteczkowej miceli czerwieni Kongo dokowanej do płyty C α_1-antyproteinazy uzyskany na drodze dynamiki molekularnej [A i B wg I. Roterman, Rozprawa habilitacyjna, *Wykorzystanie technik obliczeniowych dla przewidywania zmian strukturalnych związanych z funkcją białka*, Kraków 1995, zmodyfikowano]. C. Liczba cząsteczek czerwieni Kongo związana do α_1-antyproteinazy, wyznaczona eksperymentalnie po inkubacji z różnym wyjściowym nadmiarem czerwieni Kongo względem białka [33]

Z tymi obszarami najpewniej związana jest zdolność do wiązania czerwieni Kongo. Ujawnienie wpływu obecności kompleksów na obraz uzyskiwanego rozdziału można osiągnąć, dodając ligand α_1-antyproteinazy (trypsynę) w ilości nie przekraczającej zdolności wiążącej inhibitora. W wyniku tego obserwuje się powstawanie kompleksów, co powoduje wzrost ilości frakcji szybkowędrującej wywodzącej się z białek o ruchliwości elektroforetycznej α_1 (w blisko 90% jest to α_1-antyproteinaza) [1]. W warunkach eksperymentalnych oszacowano proporcję molową cząsteczek czerwieni Kongo związanych do komercyjnej α_1-antyproteinazy w różnych nadmiarach molowych barwnika ciekłokrystalicznego. Stosowano 20 minutową inkubację w temperaturze pokojowej, a nadmiar niezwiązanego barwnika i jego frakcję słabo zasocjowaną z powierzchnią białka usuwano przez sączenie na kolumnie z sefadeksem G-200 (rys. 2C) [33]. Dane eksperymentalne uzyskane w tym eksperymencie korespondują z danymi obliczeniowymi. Dokowanie 7 cząsteczkowej taśmy barwnika supramolekularnego do układu serpinowej β-płyty jest z punktu widzenia energetycznego korzystniejsze niż obecność na tym miejscu naturalnej pętli polipeptydowej (rys. 2B) [33].

Efekty przegrupowania frakcji białkowych, obserwowane w przedstawionej technice elektroforetycznej, można uzyskać po dodaniu do surowicy hemoglobiny – naturalnego liganda haptoglobin [1].

W literaturze opisano również wpływ barwników bisazowych na ruchliwość elektroforetyczną hemopeksyn, którą badano w technice immunoelektroforezy [14].

Wiązanie czerwieni Kongo potwierdzono również dla lipoprotein frakcji LDL surowicy. Proteinogram, w którym po przeprowadzeniu rozdziału zredukowano czerwień Kongo dwutioninem sodowym, następowo zabarwiono czernią Sudanową. Pozwoliło to na zlokalizowanie plam tłuszczowych, które głównie w zakresie frakcji beta-lipoprotein pokrywają się z uprzednio intensywnie zabarwioną czerwienią Kongo frakcją [dane niepublikowane]. Nie jest jednak jasne, czy wiązanie barwnika ma miejsce w komponencie białkowej czy tłuszczowej frakcji LDL.

Innym białkiem surowicy, którego powinowactwo do czerwieni Kongo jest od dawna znane, jest albumina. Białko to z uwagi na swoje wyjątkowe właściwości transportowe wiąże wiele cząsteczek czerwieni Kongo zarówno w kieszeniach wiążących, jak i poza nimi. W przypadku albuminy mechanizm wiązania barwnika nie jest do końca poznany [1], [14], [39], [40].

Analiza obrazu rozdziałowego

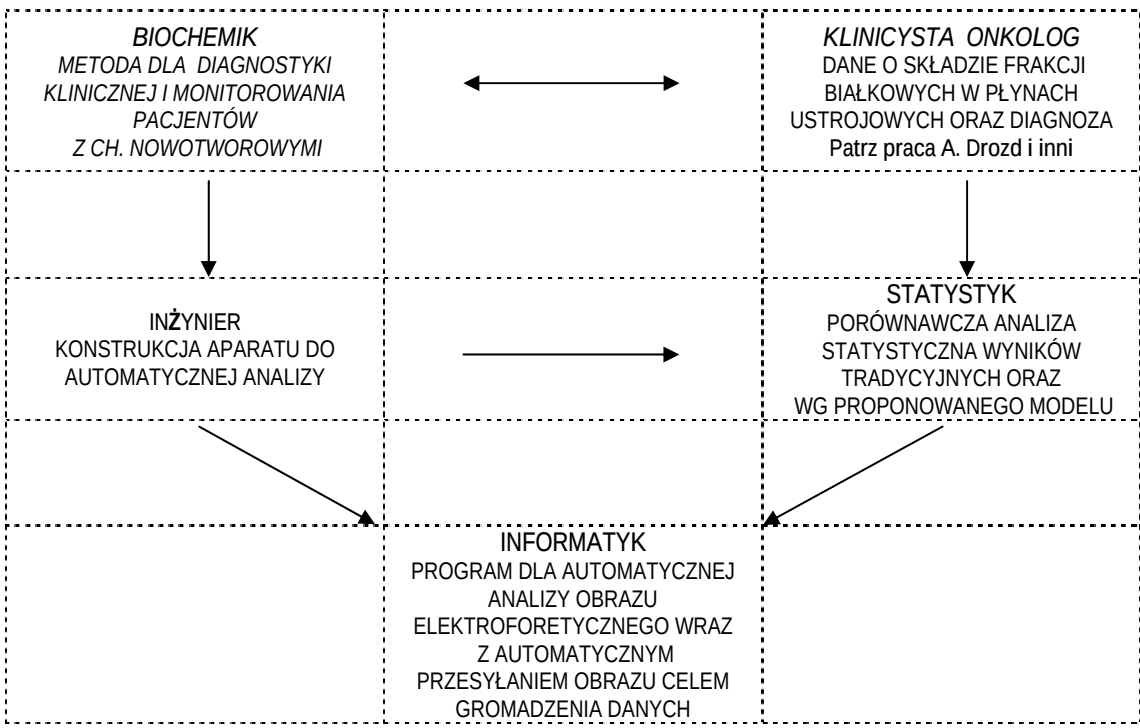
Obecnie brak jest odpowiedniego oprogramowania umożliwiającego analizę uzyskiwanych rozdziałów, zwłaszcza w aspekcie oceny procentowego udziału poszczególnych frakcji. Potrzeba ilościowego oznaczenia tworzonych kompleksów metodą densytometryczną jest krytyczna dla poddania opracowanej metody ocenie klinicznej. Próby opracowania takiego oprogramowania zostały ostatnio podjęte.

Otwartym zagadnieniem pozostaje również wyjaśnienie molekularnego mechanizmu wiązania cząsteczek czerwieni Kongo do białek ostrej fazy. Do dziś znany jest tylko model kompleksu utworzonego przez micele barwnika z α_1-antyproteinazą [33] oraz łańcuchem lekkim λ immunoglobulin (Lλ Loc, kod PDB:4BJL) [5]. Do czasu uzyskania kryształu kompleksu białka z czerwienią Kongo, metody komputerowe pozostają jedyną formą wizualizacji struktury molekularnej tego typu kompleksów.

Literatura cytowana

1. Rybarska I., Konieczny L., Piekarska B., Stopa B., Roterman I.: The detection of specific acute phase serum protein complexes and immune complexes by Congo red binding. J. Physiol. Pharmacol, 46(2): 221-231, 1995.
2. Rybarska J., Konieczny L., Roterman I., Piekarska B.: The effect of azo dyes on the formation of immune complexes, Arch. Immunol. Ther. Exp. (Warsz.), 39(3): 317-327, 1991.
3. Piekarska B., Rybarska J., Stopa B., Zemanek G., Król M., Roterman I., Konieczny L.: Supramolecularity creates nonstandard protein ligands, Acta Biochim. Polon., 46(4): 841-851, 1999.
4. Spólnik P., Piekarska B., Stopa B., Konieczny L., Zemanek G., Rybarska J., Król M., Nowak M., Roterman I.: The structural abnormality of myeloma globulins tested by Congo red binding, Med. Sci. Monit., 9(4): 145-153, 2003.
5. Król M., Roterman I., Piekarska B., Konieczny L., Rybarska J., Stopa B., Spólnik P., Szneler E. An approach to understand the complexation of supramolecular dye Congo red with immunoglobulin L chain lambda. Biopolymers, 77(3): 155-162, 2005.
6. Roterman I., No K. T., Piekarska B., Kaszuba J., Pawlicki R., Rybarska J., Konieczny L.: Bis azo dyes-studies on the mechanism of complex formation with IgG modulated by heating or antigen binding, J. Physiol. Pharmacol., 44(3): 213-232, 1993.
7. Spólnik P., Konieczny L., Piekarska B., Rybarska J., Stopa B., Zemanek G., Drozd A., Król M., Roterman I.: Congo red-derived supramolecular dyes as probes for disclosure of the aggregation tendency of abnormal monoclonal immunoglobulins, Polish Journal of Environmental Studies, w druku.
8. Watt R. M., Voss Jr. E. W.: Solvent perturbation of the fluorescence of fluorescein bound to specific antibody, J. Biol. Chem., 254(5): 1684-1690, 1979.
9. Steensma D. P.: "Congo" red out of Africa?, Arch. Pathol. Lab. Med., 125(2):250-252, 2001.
10. Puchtler H., Sweat F.: Congo red as a stain for fluorescence microscopy of amyloid, J. Histochem. Cytochem., 13(8):693-694, 1965.
11. Glenner G. G., Eanes E. D., Page D. L.: The relation of the properties of Congo red-stained amyloid fibrils to the beta-conformation. J. Histochem. Cytochem., 20(10):821-826, 1972.
12. Glenner G. G.: The bases of the staining of amyloid fibers:their physico-chemical nature and the mechanism of their dye-substrate interaction, Progr. Histochem. Cytochem., 13(3) :1-36,1981.
13. Serpell L. C., Sunde M., Blake C. C. F.: The molecular basis of amyloidosis, Cell Mol. Life Sci., 53(11-12): 871-887, 1997.
14. Emmett M., Cerniglia C. E., Crowle A. J.: Differential serum protein binding of benzidine and benzidine-congener besed dyes and their derivatives, Arch. Toxicol. 57(2): 130-135, 1985.
15. Rybarska J., Piekarska B., Konieczny L., Roterman I.: The formation of soluble heat IgG aggregates for immunological studies, Arch. Immunol. Ther. Exp. (Warsz.), 36(5): 609-622, 1988.
16. Kuciel R., Mazurkiewicz A.: Molten globule as an intermediate on the human prostatic phosphatase folding pathway. Acta Biochim. Polon. 44(4): 645-658, 1997.
17. Khurana R., Uverski V. N., Nielsen L., Fink A.L.: Is Congo red an amyloid-specific dye? J. Biol. Chem., 276(25): 22715-22721, 2001.
18. Zemanek G., Konieczny L., Piekarska B., Rybarska J., Stopa B., Spólnik P., Urbanowicz B., Nowak M., Król M., Roterman I.: Egg yolk platelet proteins from Xenopus laevis are amyloidogenic, Folia Histochem. Cytobiol., 40(3): 311-318, 2002.
19. Zemanek G., Rybarska J., Stopa B., Piekarska B, Spólnik P., Konieczny L., Roterman I., Bugajski A.: Protein distortion-derived mechanism of signal discrimination in monocytes revealed using Congo red to stain activated cells, Folia Histochem. Cytobiol., 41(3): 113-124, 2003.
20. Roterman I., Król., Nowak M. Konieczny L., Rybarska J., Stopa B., Piekarska B., Zemanek G.: Why Congo red binding is specific for amyloid proteins-model studies and a computer analysis approach, Med. Sci. Monit., 7(4): 771-784, 2001.
21. Skowronek M., Roterman I., Konieczny L., Stopa B., Rybarska J., Piekarska B.: Why do Congo red,Evans blue, and Trypan blue differ in their complexation properties. J. Comput. Chem., 21(8): 656-667, 2000.
22. Wasilewskaya A. C., Generalova E. W., Sonin A. C.: Chromonic mesophases, Usp. Khim., 58: 1574-1596, 1989.
23. Attwood T. K., Lydon J. E., Hall C., Tiddy G. J. T.: The distinction between chromonic and amphiphilic lyotropic mesophases, Liquid Crystals, 7: 657-668., 1990.
24. Skowronek M., Roterman I., Konieczny L., Stopa B., Rybarska J., Piekarska B., Górecki A., Król M.: The conformational characteristics of Congo red, Evans blue and Trypan blue, Comput. Chemistry, 24(3-4)|: 429-450, 2000.
25. Skowronek M., Stopa B., Konieczny L., Rybarska J., Piekarska B., Szneler E., Bakalarski G., Roterman I.: Self-assembly of Congo red-a theoretical and experimental approach to identify its supramolecular organization in water and salt solution, Biopolymers, 46(5): 267-281., 1998.
26. Stopa B., Górny M., Konieczny L. Piekarska B., Rybarska J., Skowronek M., Roterman I.: Supramolecular ligands: monomer structure and protein ligation capability, Biochimie, 80(12): 963-968., 1998.

27. Woodcock S., Henrissat B., Sugiyama J.: Docking of Congo red to the surface of crystalline cellulose using molecular mechanics, Biopolymers, 36(2): 201-210., 1995.
28. Gabriel J. L., Plaut G. W. E.: Specific site inhibitors of NAD-specific isocitrate dehydrogenase from bovine heart, J. Biol. Chem., 257(14): 8021-8029., 1982.
29. Edwards R. A., Woody R. W.: Spectroscopic studies of Cibacron blue and Congo red bound to dehydrogenases and kinases. Evaluation of dyes as probes of the dinucleotide fold, Biochemistry, 18(23): 5197-5204., 1979.
30. Spólnik P., Konieczny L., Piekarska B., Rybarska J., Stopa B., Zemanek G., Król M., Roterman I.: Instability of monoclonal myeloma protein may be identified as susceptibility to penetration and binding by newly synthesized Congo red derivatives, Biochimie, 86(6): 397-401., 2004.
31. Spólnik P.: Opracowanie testu oceny stabilności białek monoklonalnych osocza. Praca doktorska wykonana w Instytucie Biochemii Lekarskiej CM UJ, 2004.
32. Piekarska B., Konieczny L., Rybarska J., Stopa B., Zemanek G., Szneler E., Król M., Nowak M., Roterman I.: Heat-induced formation of a specific binding site for self-assembled Congo red in the V domain of immunoglobulin L chain, Biopolymers, 59(6): 446-456. 2001.
33. Roterman I., Rybarska J., Konieczny L. Skowronek M., Stopa B., Piekarska B., Bakalarski G.: Congo red bound to α_1-proteinase inhibitor as a model of supramolecular ligand and protein complex, Computers Chem., 22(1): 61-70. 1998.
34. Stevens F. J., Argon Y.: Pathogenic light chains and the B-cell repertoire. Immunol. Today, 20(10):451-457. 1999.
35. Dealwis C., Wall J.: Towards understanding the structure-function relationship of human amyloid disease, Curr. Drug Targets, 5(2): 159-171. 2004.
36. Solomon A., Weiss D. T., Kattine AK.: Nephrotoxic potential of Bence-Jones proteins, N. Eng. J. Med., 324(26): 1845-1851. 1991.
37. Wells MJ., Sheffield WP., Blajchmann MA.: The clearance of thrombin-antithrombin and related serpin-enzyme complexes from circulation:role of various hepatocyte receptors. Thromb. Haemost., 81(3): 325-337. 1999.
38. Silverman G. A., Bird P. I., Carrel R. W., Church FC., Coughlin PB., Gettins P. G. W., Irving J. A., Lomas D. A., Luke C. J., Moyer R. W., Pemberton P. A., Remolad-O`Donnell E., Salvesen G. S., Travis J., Whisstock J. C. The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. J. Biol. Chem. 276(36): 33293-33296. 2001.
39. Rybarska J., Piekarska B., Stopa B., Spólnik P., Zemanek G., Konieczny L., Roterman I.: In vivo accumulation of self-assembling dye Congo red in area marked by specific immune complexes: possible relevance to chemotherapy, Folia Histochem. Cytobiol., 42(4): 101-110. 2004.
40. Kragh-Hansen U.: Structure and ligand binding properties of human serum albumin, Dan. Med. Bull., 37(1): 57-84. 1990.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 139-142.

TWORZENIE SYSTEMU WSPOMAGAJĄCEGO ORGANIZOWANIE KONFERENCJI MEDYCZNYCH ZGODNIE Z MODELEM KASKADOWYM CYKLU ŻYCIA OPROGRAMOWANIA

THE SYSTEM BASED ON WATERFALL MODEL HELPING MEDICAL CONFERENCES ORGANIZATION

MICHAŁ CAŁKA[*], PRZEMYSŁAW GUZIK[**], JAN KNIAT[*], GRZEGORZ KONIECZKA[*], RADOSŁAW LUBOCH[*], KAROL SAWCZYK[*], EDWARD JACEK GORZELAŃCZYK[***]

[*]*Instytut Informatyki, Politechnika Poznańska, Poznań,*
[**]*Katedra i Klinika Intensywnej Terapii Kardiologicznej, Akademia Medyczna im. K. Marcinkowskiego w Poznaniu*
[***]*Zakład Neuropsychologii i Genetyki Behawioralnej, Akademia Bydgoska, Bydgoszcz, medsystem@medsystem.com.pl*

Streszczenie. Przedstawiono internetowy system Ordalia, który wspomaga organizację konferencji medycznych. System umożliwia obsługę wielu konferencji jednocześnie. Użytkownicy systemu to: organizatorzy, autorzy artykułów naukowych i recenzenci. Rolę głównego recenzenta pełni system. Zaprezentowano funkcje oferowane dla różnego typu użytkowników w postaci formularzy ekranowych. Przedstawiono krótki przegląd użytych metod i narzędzi programowania oraz możliwości rozwoju systemu.

Słowa kluczowe: wspomagane komputerowo zarządzanie, aplikacje internetowe, organizacja konferencji medycznych, model kaskadowy

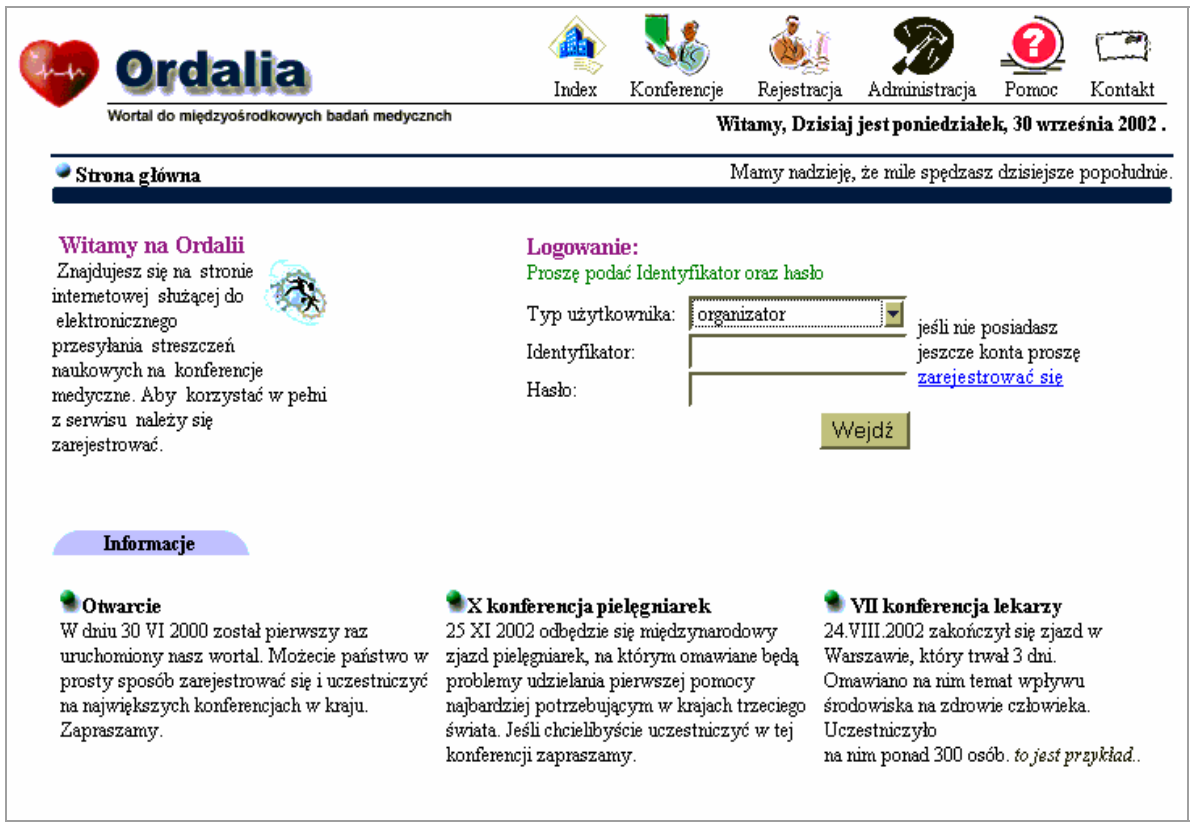
Abstract. The paper presents the Ordalia internet-based system which helps to organize medical conferences. The system offers the possibility to run many conferences at the same time. The users of the system are: organizers, authors of scientific papers, and reviewers. The role of the general supervisor is played by the system administrator. In the paper, the functionality, offered to different types of users is briefly presented by describing appropriate screen forms. The short review of used methods and programming tools is also included in the final section of the paper. The summary stresses the benefits of the Ordalia system and the possibilities of its future development.

Key words: computer aided management, internet applications, medical conferences organization, waterfall model

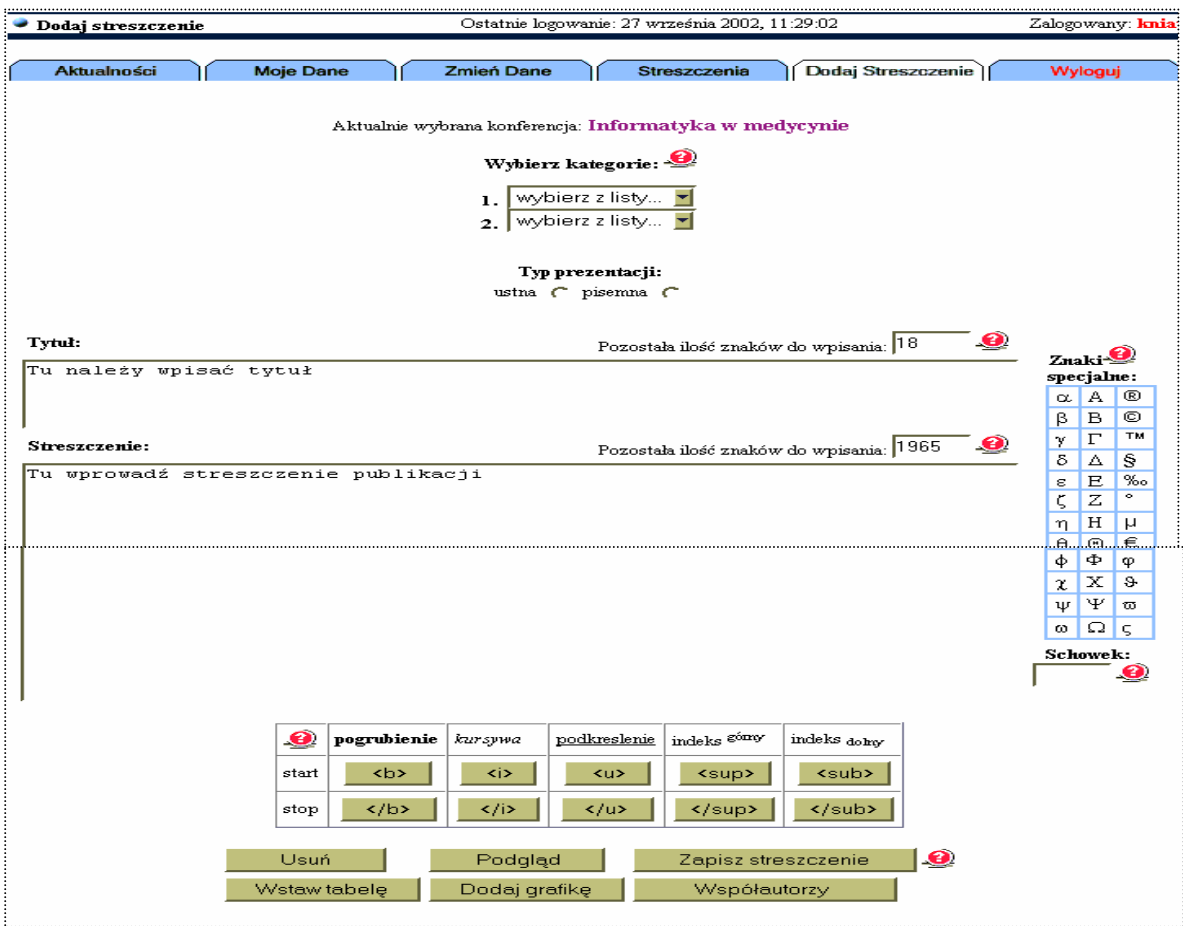
1. Wstęp

Przygotowanie medycznych konferencji naukowych wymaga od organizatorów intensywnej, wielomiesięcznej pracy oraz nierzadko znacznych nakładów finansowych. Wiele towarzystw naukowych wykorzystuje od pewnego czasu w swojej pracy specjalne formularze, służące do zgłaszania streszczeń, rejestracji konferencyjnej, wymiany informacji między organizatorami a recenzentami i osobami zgłaszającymi się na konferencje. Z roku na rok wzrasta popularność internetowych formularzy zgłoszeniowych wśród uczestników konferencji medycznych. Zapotrzebowanie na utworzenie podobnego formularza zostało sformułowane przez Komisję Informatyczną Sekcji Elektrokardiologii Nieinwazyjnej przy Polskim Towarzystwie Kardiologicznym (PTK).

W celu realizacji tego zadania powstał zespół złożony z pracowników i studentów Politechniki Poznańskiej, a także przedstawicieli środowiska medycznego związanych z działalnością Sekcji Elektrokardiologii Nieinwazyjnej PTK. Zaprojektowano i zrealizowano od podstaw własny, autorski system Ordalia, który umożliwia wspomaganie organizacji wielu konferencji jednocześnie. Przewidywani użytkownicy systemu to: organizatorzy konferencji, autorzy prac naukowych, recenzenci, administratorzy systemu.



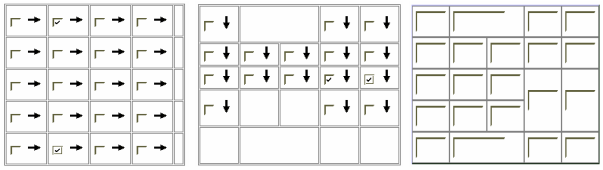
Rys. 1. Strona główna systemu Ordalia



Rys. 2. Formularz wprowadzania streszczeń

2. Projektowanie systemu Ordalia

Przystępując do projektowania formularza, autorzy zapoznali się z istniejącymi już systemami wspomagającymi organizowanie różnych konferencji medycznych w kraju i za granicą. Po dokładnej analizie różnych cech ocenianych formularzy zdecydowano się na wykorzystanie rozwiązań umożliwiających użytkownikom wysyłanie streszczeń zawierających poza tekstem m.in. różne grafiki, litery alfabetu greckiego, indeksy górne czy dolne, a także tabele, w których można łączyć komórki.



Rys. 3. Łączenie pól tabeli: kolejno w poziomie i w pionie oraz postać końcowa

3. Korzystanie z systemu Ordalia

Poniżej przedstawiono w skrócie sposób korzystania z systemu Ordalia przez autorów zgłaszających prace do prezentacji na konferencji.

Po ogłoszeniu konferencji w środowisku osób zainteresowanych, autorzy mogą nadsyłać streszczenia swoich prac do dnia zamknięcia formularza zgłoszeniowego. Wprowadzenie treści streszczenia do systemu nie wymaga korzystania z zewnętrznego edytora tekstu, gdyż edytor udostępniany jest na stronie konferencji wygenerowanej przez system Ordalia. O odrzuceniu bądź przyjęciu pracy do prezentacji autor może być poinformowany zarówno drogą elektroniczną, jak i poprzez SMS oraz listownie.

Współpraca z systemem Ordalia rozpoczyna się od wyświetlenia strony głównej zawierającej pola zgłoszenia (rys.1). Nowi użytkownicy proszeni są o wypełnienie formularza rejestracji i ustaleniu swojego identyfikatora i hasła. Formularz ten zawiera podstawowe pytania o dane osobowe, nazwę i adres instytucji, numery telefonów, faksu oraz adres poczty elektronicznej. Istnieje możliwość podania odrębnego adresu do korespondencji. Osoby uprzednio korzystające z systemu mogą podać swój wcześniejszy identyfikator i hasło. W każdej chwili można przejrzeć i zmienić swoje dane osobowe – wybranie tej opcji powoduje ponowne wyświetlenie formularza rejestracji.

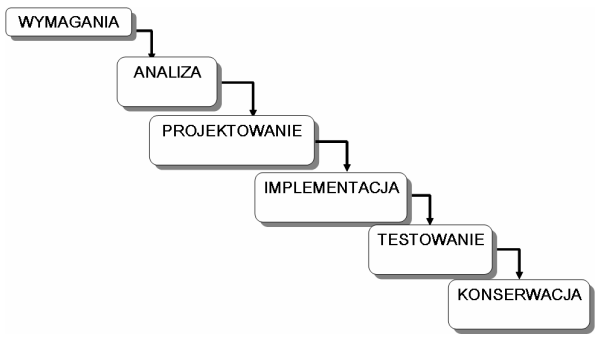
Po zgłoszeniu, użytkownikowi wyświetla się odpowiedni formularz.

Po zgłoszeniu, autorowi wyświetlana jest lista konferencji, na które może wysłać streszczenie. Jest to lista wszystkich konferencji zarejestrowanych aktualnie w systemie Ordalia – autor może zapoznać się z opisem każdej z nich. Po wybraniu opcji dodawania streszczeń, autor wskazuje konferencję, na którą jego streszczenie ma być przesłane i przechodzi do formularza wprowadzania streszczenia (rys. 2).

Korzystając z formularza przeznaczonego do wprowadzania streszczeń autor wybiera najpierw kategorie tematyczne i formę prezentacji. Następnie można wpisać nazwę streszczenia i jego treść. Edytor streszczeń udostępniany przez system Ordalia pozwala na wprowadzanie tekstu (akceptowane są litery specyficzne dla języka polskiego) i realizowanie typowych operacji edycyjnych (usuwanie, kopiowanie, wklejanie). Tekst może być również kopiowany z innego dokumentu. Ponadto edytor umożliwia wprowadzanie liter greckich (małych i dużych) i znaków specjalnych. W tym wypadku symboliczne oznaczenie wybranego znaku kopiowane jest do pola o nazwie Schowek. Ze Schowka należy ten znak przeciągnąć myszką w odpowiednie miejsce tekstu. Dodatkowo istnieje możliwość pogrubiania, pochylania (kursywa) i podkreślania tekstu indeksów, a także wprowadzania indeksów dolnych i górnych. W tym przypadku obowiązuje zasada obejmowania specjalnymi znacznikami fragmentu tekstu, którego charakter ma zostać zmieniony (na przykład pogrubiony). Znaczniki te można generować za pomocą odpowiednich przycisków i schowka. Edytor umożliwia również wprowadzanie rysunków zapisanych w kilku popularnych formatach, tj. jpg, gif, bmp. Po wskazaniu pliku zawierającego rysunek, w tekście streszczenia pojawia się odnośnik do tego pliku. Dopuszczalna wielkość rysunku w pikselach jest określana przez organizatorów. Przy pomocy systemu Ordalia można wprowadzać tabele o dowolnej liczbie kolumn i wierszy, a ponadto łączyć pola tabeli zarówno w poziomie, jak i w pionie (rys. 3)

Do tekstu streszczenia wprowadzane jest kodowe oznaczenie tabeli. Edytor udostępniany przez system Ordalia nie posiada cechy WYSIWYG (*What You See Is What You Get*) – w tekście pojawiają się różnego rodzaju znaczniki opisujące charakter pewnych fragmentów. Postać docelową można obejrzeć po zapisaniu streszczenia i wybraniu opcji Podgląd. Po zapoznaniu się z pełnym wyglądem streszczenia można powrócić do jego edycji lub przekazać je do recenzji (opcja Wyślij). Po upłynięciu ostatecznego terminu nadsyłania streszczenie dalsza ich edycja jest automatycznie zablokowana i streszczeń, zostają przekazane do recenzji.

System Ordalia był realizowany zgodnie z modelem kaskadowym (*waterfall model*) cyklu życia oprogramowania (rys. 4).



Rys. 4. Model kaskadowy cyklu życia oprogramowania

Model ten składa się z następujących faz:

- określenie wymagań – definiowane są cele oraz szczegółowe wymagania wobec tworzonego systemu,
- analiza,
- projektowanie – powstaje szczegółowy projekt systemu spełniającego ustalone wcześniej wymagania,
- implementacja – projekt zostaje zaimplementowany w konkretnym środowisku programistycznym,

- testowanie – następuje integracja poszczególnych modułów połączona z testowaniem poszczególnych podsystemów oraz całego oprogramowania,
- konserwacja – oprogramowanie jest wykorzystywane przez użytkowników, a producent wykonuje modyfikacje polegające na usuwaniu błędów, zmianach i rozszerzaniu funkcji systemu.

W fazie określania wymagań bardzo istotne okazały się uwagi przedstawicieli środowiska medycznego, którzy pełnili rolę klientów specyfikujących cechy i funkcje, jakie system powinien posiadać. Podczas fazy projektowania dokonano wyboru narzędzi programistycznych. Postanowiono wykorzystać głównie oprogramowanie darmowe w celu zminimalizowania kosztów stworzenia systemu. Drugą ważną przesłanką przy podejmowaniu decyzji wyboru narzędzi programistycznych była ich efektywność działania, niezawodność oraz łatwość konserwacji gotowego systemu. Uwzględniając wszystkie powyższe aspekty doboru odpowiednich narzędzi, postanowiono wykorzystać:

– język HTML,
– kaskadowe arkusze stylów,
– język skryptowy JavaScript,
– język programowania PHP,
– bazę danych MySQL.

Zastosowanie tych właśnie narzędzi umożliwiło stworzenie produktu nowoczesnego, zgodnego z obecnymi trendami, a zarazem użytecznego i wychodzącego naprzeciw zapotrzebowaniu polskiego środowiska medycznego.

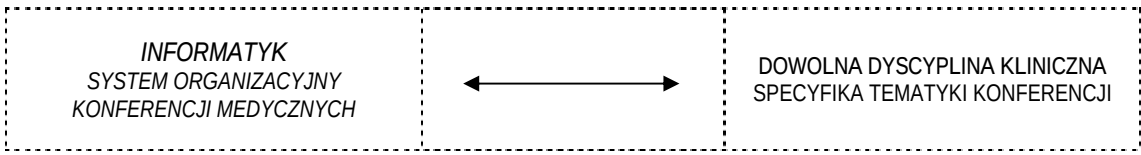
System Ordalia jest przewidziany do instalowania i użytkowania pod nadzorem systemu operacyjnego FreeBSD (darmowa odmiana systemu UNIX) z serwerem Apache. Użytkownicy końcowi korzystać natomiast mogą z dowolnej przeglądarki stron WWW, uruchomionej pod nadzorem dowolnego systemu operacyjnego. Niestety, stosowane powszechnie przeglądarki nie są ze sobą całkowicie zgodne, co uwidacznia się szczególnie przy realizacji bardziej zaawansowanych funkcji. System Ordalia dostosowano do najpopularniejszej przeglądarki, tj. do Internet Explorer firmy Microsoft.

4. Podsumowanie

System Ordalia ułatwia przygotowanie medycznych konferencji naukowych. Organizator takiej konferencji może na bieżąco przydzielać streszczenia recenzentom, obserwować postępy procesu recenzowania i kształtować na tej podstawie program konferencji. Autorzy mają do dyspozycji edytor, który umożliwia szybkie wpisywanie i przesyłanie tekstów. System Ordalia umożliwia przygotowanie wielu różnych konferencji w tym samym czasie.

Literatura cytowana

1. Bradenbaugh J.: JavaScript Receptury, Wydawnictwo Helion, Warszawa 2001.
2. Lemay L.: HTML 4, Wydawnictwo Helion, Warszawa 1998
3. Jaszkiewicz A.: Inżynieria Oprogramowania, Wydawnictwo Helion, Warszawa 1997.
4. Ladanyi H.: SQL Księga eksperta –Wydawnictwo Helion, 2000.
5. Converse T., Park J.:, PHP 4 Biblia, Wydawnictwo Helion, Warszawa 2001.
6. Tatschiller T., Gerken T.:, PHP Aplikacje, Wydawnictwo Robomatic, Warszawa 2001.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 143-146.

ANALIZA BIOMECHANICZNA SZTUCZNEGO KRĄŻKA MIĘDZYKRĘGOWEGO

BIOMECHANICAL ANALYSIS OF ARTIFICIAL LUMBAR DISC

PAWEŁ BORKOWSKI

Instytut Techniki Lotniczej i Mechaniki Stosowanej Politechniki Warszawskiej, Warszawa, pbork@itlims.meil.pw.edu.pl

Streszczenie. Praca zawiera założenia projektowe, biomechanikę i analizę nowego rozwiązania sztucznego krążka w kręgosłupie lędźwiowym. Proteza składała się z dwóch metalowych płyt i wkładki polietylenowej. Do analizy zastosowano metodę elementów skończonych. W protezie wyznaczono przemieszczenia i naprężenia dla przypadku ściskania z jednoczesnym skręceniem o kąt równy zakresowi ruchu.

Słowa kluczowe: inżynieria biomedyczna, kręgosłup lędźwiowy, implanty, MES

Abstract. The work presents brief foredesign, biomechanics and analysis of the new artificial disc in the lumbar spine. The prosthesis consisted of two metal plates and the polyethylene inlay. Displacements and stresses for compression and torsion in the prosthesis were determined for assumed range of motion by using the finite element method.

Key words: biomedical engineering, lumbar spine, implants, FEM

1. Wprowadzenie

Cechą współczesnego stylu życia człowieka jest zbyt częste przebywanie w pozycji siedzącej. Wymuszona sylwetka siedząca ma niekorzystny wpływ na obciążenia kręgosłupa lędźwiowego, powodując zmiany zwyrodnieniowe oraz zespoły bólowe już u osób w młodym wieku. Szacuje się, że problem ten dotyka od 40 do 80% ludności. W dawnych czasach średni okres życia ludzkiego był znacznie krótszy. Człowiek umierał, zanim w kręgosłupie zdążyły pojawić się zaawansowane zmiany zwyrodnieniowe. W dzisiejszych czasach długie godziny spędzane w pracy przed komputerem, a w domu przed telewizorem powodują osłabienie i rozluźnienie mięśni grzbietu i brzucha, które w normalnych warunkach stabilizują kręgosłup. Często procesowi temu towarzyszy nadwaga. Wynika stąd, że w schorzeniach układu kostno-szkieletowego tryb życia jest głównym czynnikiem chorobotwórczym.

Obecnie leczeniu operacyjnemu poddawanych jest kilka procent pacjentów z bólami kręgosłupa. Bezwzględnym wskazaniem do operacji są zaburzenia neurologiczne. Od lat w praktyce klinicznej stosuje się metodę stabilizacji. Segment ruchowy kręgosłupa składający się z dwóch kręgów i zdegenerowanego krążka międzykręgowego unieruchamia się za pomocą stabilizatora [2]. Powoduje to przyspieszony proces zużycia krążków w sąsiednich segmentach na skutek zwiększenia zakresów ruchu potrzebnych do kompensacji usztywnienia [7].

Wielu doświadczonych chirurgów od lat stosujących metodę stabilizacji uważa protezy krążków międzykręgowych za rozwiązanie dyskusyjne. Zarzuty dotyczą niepełnego odtworzenia funkcji naturalnego krążka oraz powikłań pooperacyjnych. Główne z nich polegają na wypadaniu implantu z miejsca wszczepienia lub zapadaniu do wnętrza kręgów. Pojawiają się również infekcje i odczyny zapalne spowodowane reakcją organizmu na implant i produkty zużycia. Z drugiej strony wykonanie operacji stabilizacji u osób w młodym wieku lub czynnych zawodowo w dłuższym okresie czasu również przynosi niekorzystne skutki związane z wykonywaniem kolejnych operacji łączenia sąsiednich segmentów ruchowych.

Inżynieria biomedyczna ma stosunkowo niewielkie doświadczenie w badaniach związanych z implantami kręgosłupa w porównaniu z badaniami protez stawu biodrowego, kolanowego czy ramiennego, chociaż protezy SB Charite (LINK) (DePuy Spine, Raynham, MA) przez ostatnie 20 lat wszczepiono ponad 6500 pacjentom [6]. W Stanach Zjednoczonych trwa obecnie rozpoczęty w roku 2000 program IDE (Investigational Device Exemption), mający za zadanie włączenie operacji wszczepienia sztucznego krążka do praktyki klinicznej. Programem objętych jest kilka rozwiązań (między innymi ProDisc II (Spine Solution/Synthes, Paoli, PA) i Maverick Total Disc Replacement (Medtronic Sofamor Danek, Inc., Memphis, TN)). Pacjenci są dzieleni losowo na dwie grupy. Większej liczbie osób wszczepia się protezy, natomiast u pozostałych wykonuje się stabilizację. Wyniki w określonych odstępach czasu porównuje się za pomocą

diagnostyki obrazowej (radiologia, tomografia komputerowa, rezonans magnetyczny) oraz kwestionariuszy wypełnianych przez pacjentów (VAS, ODI, SF-36, LBOS). Z dotychczas przeprowadzonych badań wynika, że przy określonych wskazaniach przedoperacyjnych implanty okazały się rozwiązaniem bardziej korzystnym od połączenia kręgów zarówno w krótkim [6, 15], jak i w dłuższym okresie czasu [10]. Wszczepienie protezy w porównaniu do operacji stabilizacji trwało średnio dwa razy krócej, utrata krwi podczas operacji była dwukrotnie mniejsza, a okres hospitalizacji był krótszy o jeden dzień. W obu typach operacji stan samopoczucia pacjentów od poziomu określanego według kwestionariuszy jako kalectwo poprawił się w ciągu 2 lat do poziomu odczuwanego jako umiarkowana niepełnosprawność. Stany zapalne i infekcje wystąpiły w równym stopniu u niewielkiej części pacjentów z obu grup. Pacjenci po operacji stabilizacji mieli zakładane gorsety usztywniające na okres 3 miesięcy, natomiast w przypadku wszczepienia protezy okres ten wynosił 2 tygodnie. Podobnie średni czas powrotu do pracy wynosił odpowiednio 16 i 8 tygodni. Sztuczne krążki okazały się przez to konkurencyjne w stosunku do metody tradycyjnej.

2. Wymagania stawiane protezom

Przy omawianiu wymagań stawianych protezom można wyróżnić kryteria kliniczne i biomechaniczne [9]. Kryteria kliniczne dotyczą złagodzenia bólu, odtworzenia funkcji połączenia stawowego, uniknięcia powikłań pooperacyjnych, skrócenia okresu rekonwalescencji, łatwej implantacji oraz bezpiecznej rewizji.

Kryteria biomechaniczne polegają na zapewnieniu odpowiedniej ruchliwości w segmencie, tłumieniu nagłych przyrostów obciążeń, zmniejszeniu obciążenia sąsiednich segmentów ruchowych, zapewnieniu stabilności biomechanicznej i właściwego przylegania protezy do kości.

Złagodzenie bólu osiąga się przez usunięcie zdegenerowanego krążka jako czynnika zapalnego i przywrócenie odpowiedniego napięcia mięśni i więzadeł. Funkcje połączenia stawowego odtwarza się przez umożliwienie ruchu i przywrócenie właściwego odstępu między kręgami oraz średnicy kanału kręgowego. Można uniknąć powikłań zapalnych przez sterylizację implantu i dobór materiałów o nietoksycznych produktach zużycia.

Nieprawidłowe umiejscowienie protezy w kręgosłupie podczas operacji wszczepienia może spowodować skutki znacznie gorsze od uzyskiwanych po stabilizacji, takie jak utrzymywanie się bólu, rozwój procesu degeneracji wyrostków stawowych oraz przewlekłe stany zapalne. Czasami w takich przypadkach implant należy usunąć. Operacja taka zwykle wiąże się z dużym ryzykiem, związanym między innymi z możliwością uszkodzenia naczyń krwionośnych, takich jak żyły i tętnice [8]. Dlatego przy konstruowaniu należy również uwzględnić sposób usunięcia protezy spomiędzy kręgów.

Poprawnie dobrane zakresy ruchu protezy mogą powstrzymać procesy degeneracyjne w sąsiednich segmentach. Zapewnienie odpowiednich zakresów kątowych oraz położeń chwilowych środków obrotu dotyczy ruchów ściskania, rozciągania, skłonu, prostowania, zgięcia bocznego i skręcania [14].

Tłumienie nagłych przyrostów obciążeń wymaga zastosowania odpornej na zmęczenie, podatnej wkładki. W przeszłości podejmowane były próby wszczepienia protez AcroFlex (DePuy Spine, Inc., Raynham, MA) z podatną wkładką wykonaną z gumy poliolefinowej. Materiał ten wykazywał we wcześniejszych testach laboratoryjnych dużą wytrzymałość zmęczeniową. Pomimo pomyślnych wyników badań po wszczepieniu we wkładce dochodziło do pęknięć [5]. Obecnie stosowane protezy praktycznie nie tłumią obciążeń i dlatego zadanie to muszą spełnić pozostałe krążki międzykręgowe.

Zapewnienie biomechanicznej stabilności polega na ograniczeniu nadmiernych przemieszczeń powierzchni stawowych, więzadeł i mięśni prowadzących do naderwania tkanek miękkich czy zwichnięcia stawów.

Zamocowanie protezy do kręgów w czasie wszczepiania odbywa się za pomocą elementów mocujących (kolce, grzebienie). Późniejszy przerost kości na powierzchni implantu zależy od właściwości elektrycznych materiału protezy i warstwy porowatej napawanej na zewnętrznej powierzchni płyt (np. tytan lub hydroksyapatyt).

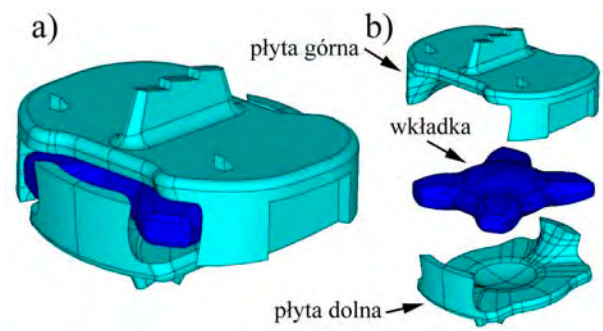
Przy doborze materiałów w obecnie stosowanych implantach skorzystano z wcześniejszych doświadczeń uzyskanych z badań protez innych stawów [12]. Płyty mocowane do kręgów wykonuje się głównie ze stopów kobaltowo-chromowych. Wkładki pośredniczące w przenoszeniu obciążeń i zapewniające ruch wytwarza się z wielkocząsteczkowego polietylenu. Dzięki rozwojowi inżynierii materiałowej i opracowaniu lepszych, bardziej biozgodnych stopów powraca również tendencja do wszczepiania protez wykonanych całkowicie z metalu, w których ilość produktów zużycia jest wielokrotnie mniejsza niż w przypadku polietylenu [11]. Zasadność takiego postępowania potwierdzają pomyślne wyniki badań na zwierzętach [4].

3. Projekt wstępny nowej protezy

Proteza nowego typu miała spełnić wymagania dotyczące wszystkich zakresów ruchu z wyjątkiem przemieszczeń w kierunku pionowym (tab. 1). Składała się z dwóch płyt ze stopu kobaltowo-chromowego i polietylenowej wkładki (rys. 1).

Tab. 1. Zakresy ruchów kątowych nowej protezy

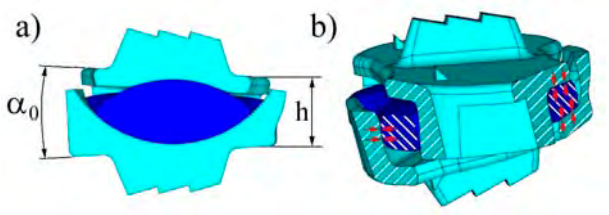
zakres ruchu	wartość
skłon	13°
prostowanie	7°
zginanie boczne	±10°
skręcanie osiowe	±5.5°



Rys. 1. Nowa proteza krążka: a) złożenie, b) części składowe

W płytach znajdowały się sferyczne zagłębienia o promieniu 17mm. Promienie w soczewkowej części wkładki miały tę samą wartość. Wysokość protezy w tylnej części wynosiła *h*=10mm, kąt lordozy α_0=6°, długość i szerokość odpowiednio 30mm i 39mm (rys. 2).

Występy znajdujące się w dolnej i górnej płycie oraz wypustki we wkładce naprzemiennie ograniczały ruchy wkładki, na przykład dla skrajnej wartości zakresu ruchu przy prostowaniu kręgosłupa tylna część wkładki była ściskana pomiędzy płytami górną i dolną i jednocześnie przedni występ dolnej płyty zabezpieczał wkładkę przed wypadnięciem (rys. 2.b). Podobnie odbywało się dla ruchów skłonu i zgięcia bocznego. Płyty górną i dolną specjalnie ukształtowano, aby uniknąć kontaktu pomiędzy płytami i zminimalizować zjawisko karbu w rejonach o skokowej zmianie grubości.



Rys. 2. Wyjaśnienie zasady działania nowego krążka: a) przekrój płaszczyzną pionową, b) sposób blokowania protezy dla skrajnego zakresu ruchu prostowania (powierzchnie kontaktu zaznaczono strzałkami)

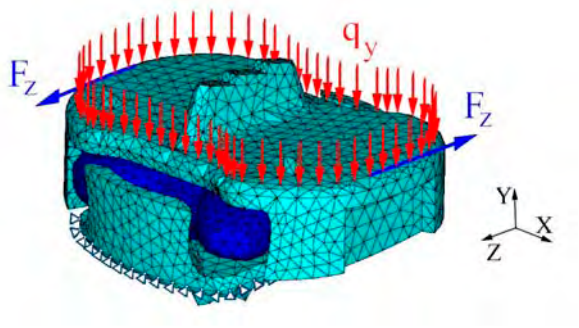
Przy omawianiu wymagań stawianych protezom wspomniano o zapewnieniu prawidłowego położenia chwilowego środka obrotu. Przy rozpatrywaniu poszczególnych ruchów jest to pojęcie w znacznym stopniu umowne ze względu na sprzężenie ruchów w płaszczyźnie czołowej i strzałkowej z ruchami skrętnymi kręgosłupa [14]. W naturalnym krążku punkt ten zmienia swoje położenie w zależności nie tylko od rodzaju, ale również fazy ruchu. Położenie środka obrotu zależy również od degeneracji krążka i wyrostków stawowych. W omawianym rozwiązaniu położenie tego punktu zmieniało się nieznacznie przy zginaniu w obu płaszczyznach i było ustalone dla ruchów skrętnych. Występy w płytach służyły do odebrania swobody przemieszczeń wkładki w płaszczyźnie poziomej (według rys. 2a), umożliwiając przenoszenie sił tnących.

4. Analiza MES

Model numeryczny zbudowano w systemie ANSYS w celu wyznaczenia stanu przemieszczeń i naprężeń w częściach składowych protezy (rys. 3). Przyjęto liniowe, izotropowe modele materiałów sztucznego krążka (tab. 2). Na powierzchniach pomiędzy wkładką a płytami znajdowały się elementy kontaktowe ze współczynnikiem tarcia μ=0,05.

Tab. 2. Właściwości materiałów sztucznego krążka

materiał	moduł Younga [MPa]	liczba Poissona
Polietylen UHMWP	1200	0,35
Stop CoCrMo	$2 \cdot 10^5$	0,3

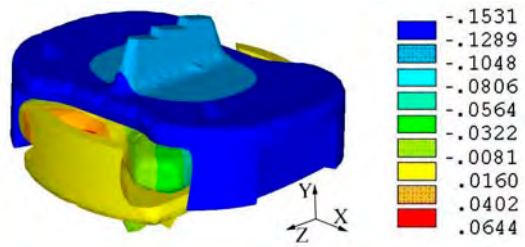


Rys. 3. Model numeryczny protezy z zaznaczonym obciążeniem i utwierdzeniem dla przypadku ściskania ze skręcaniem

Analizowano przypadek ściskania ze skręcaniem, niekorzystny ze względu na ścinanie wkładki. Płyty górną i dolną wstępnie obrócono o kąt równy zakresowi skręcania osiowego w celu zniesienia luzu. Utwierdzono krawędź zewnętrznej powierzchni dolnej płyty. W górnej płycie wybrano linie należące do krawędzi powierzchni zewnętrznej równoległe do płaszczyzny symetrii. Węzły należące do tych linii obciążono po prawej i lewej stronie siłami o przeciwnych zwrotach skierowanymi wzdłuż linii. Ich wypadkowe F_z o wartości ±83,78 N tworzyły parę sił równoważną momentowi skręcającemu 3,1 Nm. Oprócz tego krawędź zewnętrznej powierzchni górnej płyty obciążono równomiernym wydatkiem sił normalnych ściskających q_y o wypadkowej -3000 N. We wszystkich węzłach tej krawędzi nadano więzy równości przemieszczeń w kierunku normalnym do powierzchni. Przedstawiony powyżej sposób obciążenia i utwierdzenia przyjęto na podstawie wcześniejszych analiz dotyczących współpracy sztucznego krążka z kręgami [3].

5. Wyniki

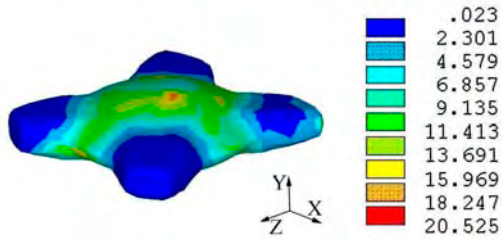
Zmiana kształtu wkładki polietylenowej pod wpływem działającego obciążenia w rozpatrywanym przypadku nie miała znaczącego wpływu na zakresy ruchu protezy. Przemieszczenia w kierunku pionowym były niewielkie i zależały głównie od podatności wkładki (rys. 4). Zmniejszenie wysokości wkładki wynosiło 0,12 mm, co stanowiło 1,3% wysokości. Obciążone wypustki ulegały niewielkim przemieszczeniom pionowym w kierunku przeciwnym do działającego obciążenia (ok. 0,07 mm).



Rys.4. Przemieszczenia pionowe [mm] w protezie dla siły ściskającej 3000N i momentu skręcającego 3,1 Nm.

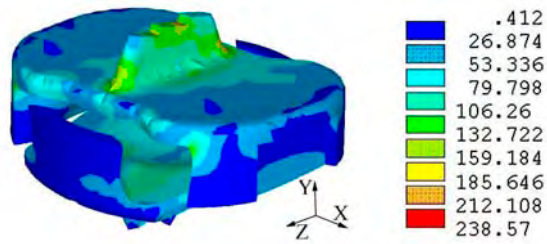
Strefy kontaktu wkładki z płytami wystąpiły w części soczewkowej i przy nasadzie wypustek. We wkładce dominowały naprężenia ściskające. Najbardziej obciążony był obszar pod powierzchnią i wewnątrz soczewkowej części wkładki (rys.5). Naprężenia zredukowane wyznaczone według

hipotezy Hubera-Misesa-Hencky'ego dochodziły tam do granicy plastyczności polietylenu, wynoszącej ok. 22 MPa [1].



Rys. 5. Naprężenia zredukowane HMH [MPa] we wkładce dla siły ściskającej 3000 N i momentu skręcającego 3,1 Nm

W płytach metalowych uzyskano korzystne rozkłady naprężeń zredukowanych (rys. 6). Największe koncentracje wystąpiły przy nasadzie grzebieni mocujących. Wartości tych naprężeń były o połowę mniejsze od granicy plastyczności, wynoszącej dla stopu kobaltowo-chromowego 450 MPa [13].



Rys. 6. Naprężenia zredukowane HMH [MPa] w płytach dla siły ściskającej 3000 N i momentu skręcającego 3,1 Nm

6. Podsumowanie

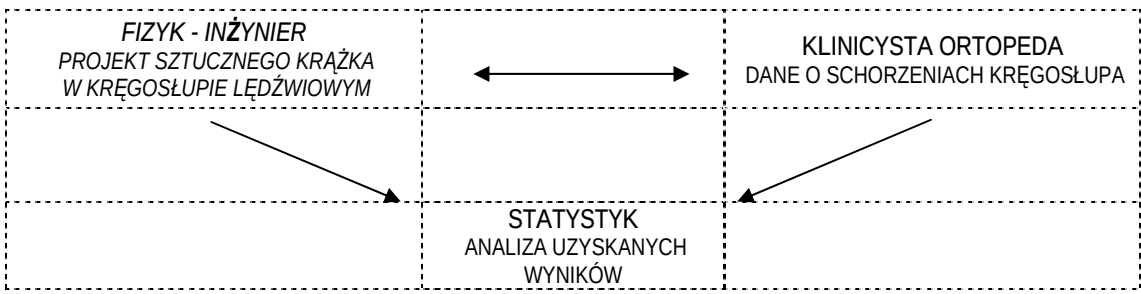
W badanej protezie zakresy ruchu można było dobrać w ograniczony sposób. Zakres skręcania osiowego zależał w pewnym stopniu od zakresów zginania w obu płaszczyznach, co wynikało z przemieszczeń poprzecznych wkładki podczas ruchów kątowych protezy.

Obliczenia numeryczne pozwalają na wczesne wyeliminowanie niepożądanych efektów w postaci koncentracji naprężeń czy nadmiernego obciążenia poszczególnych części implantu. Znaczne wartości naprężeń uzyskane we wkładce polietylenowej były podobne do uzyskanych przy wcześniejszych analizach istniejących rozwiązań SB Charite i ProDisc II [3]. Natomiast koncentracje naprężeń w płytach metalowych były o wiele mniejsze w nowym rozwiązaniu. W badaniach klinicznych potwierdzono występowanie deformacji plastycznych w polietylenowej wkładce [13]. Ze zjawiskiem tym wiązała się zmiana wysokości protezy i zakresów ruchu, które powodowały powrót bólu, stany zapalne, zaburzenie biomechaniki segmentu ruchowego i konieczność rewizji.

Plastyczna deformacja wkładki polietylenowej będzie uwzględniona w modelu numerycznym w kolejnych obliczeniach za pomocą nieliniowego modelu materiałowego. Badany będzie również model z metalową wkładką wykonaną ze stopu kobaltowo-chromowego.

Literatura cytowana

1. Bajaria S. H., Bellare A.: Deformation, Morphology, and Wear Behavior of Polyethylene Used in Orthopedic Implants, Medical Plastics and Biomaterials, http://www.devicelink.com/mpb/index.html
2. Będziński R.: Biomechanika inżynierska- zagadnienia wybrane, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 1997.
3. Borkowski P., Kędzior K., Krzesiński G., Skalski K. R., Wymysłowski P., Zagrajek T.: Numerical investigation of a new type of artificial lumbar disc, Journal of Theoretical and Applied Mechanics, 42, 2, pp. 253-268,
4. Chang B-S., Brown P. R., Sieber A., Valdevit A., Tateno K., Kostuik J. P.: Evaluation of the biological response of wear debris, The Spine Journal, 4: 239S–244S, 2004.
5. Fraser R. D., Ross E. R., Lowery G. L., Freeman B. J., Dolan M.: AcroFlex design and results, The Spine Journal, 4: 245S–251S, 2004.
6. Guyer R., McAfee P., Hochschuler S., Blumenthal S., Fedder I. L., Ohnmeiss D. Cunningham B. W.: Prospective randomized study of the Charite´ artificial disc: data from two investigational centers, The Spine Journal, 4: 252S–259S 2004.
7. Hilibrand A. S., Robbins M.: Adjacent segment degenration and adjacent segment disease: the consequences of spinal fusion? The Spine Journal, 4: 190S–194S 2004.
8. Kostuik J. P.: Complications and surgical revision for failed disc arthroplasty, The Spine Journal, 4: 289S–291S, 2004.
9. Lee C. K., Goel V. K.: Artificial disc prosthesis: design concepts and criteria, The Spine Journal, 4: 209S–218S, 2004.
10. Marnay T. Lumbar disc replacement: 7 to 11-year results with Prodisc, The Spine Journal, 2: 94S, 2002.
11. Mathews H. H., LeHuec J. Ch., Friesem T., Zdeblick T., Eisermann L.: Design rationale and biomechanics of Maverick Total Disc arthroplasty with early clinical results, The Spine Journal, 4: 268S–275S, 2004.
12. Santos E. G., Polly D. W., Mehbod A. A., Saleh K. J.; Disc arthroplasty: lessons learned from total joint arthroplasty, The Spine Journal, 4: 182S–189S, 2004.
13. Taksali S., Grauer J. N., Vaccaro A. R.: Material considerations for intervertebral disc replacement implants, The Spine Journal, 4: 231S–238S, 2004.
14. White A.A., Panjabi M.M.(eds): Clinical Biomechanic of the Spine. ed 2. Lippincot-Williams&Wilkins, 1990.
15. Zigler J. E.: Lumbar spine arthroplasty using the ProDisc II, The Spine Journal, 4: 260S–267S, 2004.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 147-156.

MODELOWANIE SZTUCZNYCH UKŁADÓW MIĘŚNIOWYCH Z AKTUATORAMI PNEUMATYCZNYMI

MODELLING OF MUSCULOTENDON SYSTEMS USING PNEUMATIC ACTUATORS

RYSZARD DINDORF

Zakład Mechatroniki, Politechnika Świętokrzyska, Kielce, dindorf@tu.kielce.pl

Streszczenie. Praca dotyczy możliwości zastosowania sztucznych mięśni w inżynierii biomedycznej i inżynierii rehabilitacyjnej. Przedstawiono najnowsze osiągnięcia w technologii wytwarzania i technice sterowania sztucznymi mięśniami. Zamieszczono różne rozwiązania kinematyczne sztucznego układu mięśniowo–ścięgnowego z aktuatorami pneumatycznymi. Podano przykłady zastosowania mięśni pneumatycznych i hydraulicznych w zaopatrzeniu ortopedycznym oraz egzoszkieletonach. Dla mięśni pneumatycznych typu „Fluidic Muscle" przedstawiono modele statyczne i charakterystyki izobaryczne, izotoniczne, izometryczne, a dla pneumatycznych aktuatorów mięśniowych (PMA), nazywanych „McKibben", przedstawiono modele i charakterystyki dynamiczne. Zaproponowano także model obliczeniowy egzoszieletonów z pneumatycznymi aktuatorami mięśniowymi w konfiguracji z mięśniem zginaczem – bicepsem i mięśniem prostownikiem – tricepsem.

Słowa kluczowe: układ mięśniowo-ścięgnowy, sztuczne mięśnie, pneumatyczne aktuatory mięśniowe

Abstract. The paper deals with the application of artificial muscle in biomedical engineering and rehabilitations engineering. The latest developments in production technology and control technology concerning artificial muscles and some kinematic solutions for an artificial musculotendon system with pneumatic actuator are included. The examples of using pneumatics muscle actuators and hydraulic muscle actuators in prostheses and orthotics devices as well in exoskeletons are presented. For pneumatic muscle of the "Fluidic Muscle" type statistic models, isobaric, isotonic and isometric characteristics are given while for pneumatic muscle acutator (PMA) also known as "McKibben" dynamic models and response characteristic are described. The computational model for exoskeletons with pneumatic muscle actuators in flexor muscle – biceps configuration and extensor muscle – triceps configuration is put forth.

Key words: musculotendon actuator system, artificial muscle, pneumatic muscle actuator

1. Wprowadzenie

Z danych statystycznych wynika, że coraz więcej osób wymaga wspomagania podczas poruszania się. Metody leczenia narządów ruchu obejmują zarówno metody zachowawcze z zastosowaniem aparatów ortopedycznych (ortez, protez, egzoszkieletonów), jak i metody operacyjne polegające na przeszczepianiu ścięgien i mięśni, wszczepianiu endoprotez, implantów i stymulatorów mięśni oraz zastosowaniu sztucznych aktuatorów mięśniowych.

Mięśnie biologiczne (szkieletowe, gładkie, sercowy) są naturalnymi aktuatorami czynnościowymi przetwarzającymi energię chemiczną na mechaniczną i cieplną. Mięśnie umożliwiają ruch organizmu i pracę jego narządów [1]. Wiele procesów fizjologicznych, np. oddychanie, krążenie, trawienie, mimika, zależy od pracy poszczególnych grup mięśniowych. Każda grupa mięśniowa ma inne charakterystyki, określane w układach: siła mięśniowa – długość sarkomeru, napięcie mięśniowe – względna długość mięśnia, potencjał elektryczny – napięcie mięśniowe. Mięsień zbudowany jest z tkanki łącznej włóknistej, którą można nazwać elastomerem biologicznym wykazującym równocześnie cechy sprężyste i plastyczne. Właściwości lepkosprężyste mięśni wynikają z ich włóknistej budowy polimerowej, dlatego mięśnie szkieletowe są bardziej rozciągliwe od innych grup mięśniowych. Mięśnie szkieletowe stanowią główny składnik masy ciała człowieka, występują w postaci oddzielnych jednostek anatomicznych o zróżnicowanej wielkości oraz kształcie. Większość ze 150 różnych mięśni szkieletowych jest przyczepiona do kośćca dwoma końcami za pomocą ścięgien.

Obecnie w praktyce medycznej i rehabilitacyjnej przy usprawnianiu narządu ruchu dąży się do wykorzystania wysokich technologii (Hi-Tech) w zakresie zastosowania nowych bioelementów, do których można obecnie zaliczyć sztuczne mięśnie (*Arificial Muscle*) [6]. Na całym świecie prowadzone są zaawansowane badania dotyczące sztucznych mięśni wykonanych z różnych biomateriałów, np. naukowcy z Uniwersytetu Nowy Meksyk (USA) proponują wszczepienie sztucznego mięśnia wykonanego z podwójnego pasemka „Bi-Strips", zasilanego cewką ze złotego drutu przymocowanego do gałki ocznej (zewnętrznej twardówki). Skracanie lub wydłużanie soczewki oka umożliwia wówczas sztuczny

mięsień pobudzany zewnątrz elektromagnesem. Sztuczne mięśnie mogą być stosowane do usprawniania chorych po przebytych chorobach (udar i uraz mózgu, porażenie mózgowe, uszkodzenie rdzenia kręgosłupa, uszkodzenie obwodowego układu nerwowego, choroby przebiegające z zanikiem mięśni) lub po amputacji kończyn. Sztuczne układy (aktuatory) mięśniowe mogą być pomocne w usprawnieniu ludzi starszych, chorych oraz niepełnosprawnych. Sztuczne układy mięśniowe stosowane np. w protezach kończyn dolnych mogą zwiększyć ich ruchliwość, poprawić symetrię chodu, a także zredukować energię zasilania do poziomu metabolizmu człowieka. Sztuczne mięśnie mogą być wykonane z różnych biomateriałów pobudzanych elektrycznie, magnetycznie, termicznie, optycznie, chemicznie, pneumatycznie oraz hydraulicznie.

Większość prac naukowo-badawczych koncentruje się obecnie na następujących sztucznych mięśniach – aktuatorach mięśniowych [9]:

– Sztuczne mięśnie zbudowane z nanorurek NAM (*Nanotube Artificial Muscle*), wynalezione w Instytucie Maxa Plancka (Stuttgart), produkowane są przez firmę AlliedSignal Inc. Wzdłuż węglowych nanorurek ułożone są krążki zbudowane z 14 atomowego węgla. Submikroskopowy węgiel ukształtowany w postaci nanorurki wypełniony jest elektrolitem. Po doprowadzeniu prądu sztuczny muskuł NAM kurczy się lub wydłuża. Z miliona nanorurek buduje się włókna mięśni o dowolnej długości, które przypominają strukturę mięśnia naturalnego.

– Sztuczne biomięśnie (*BioArificial Muscle*) zbudowane są z kurczliwego polimeru wykonanego z włókna poliakrylonitrylu (PAN), poliakrylonitrylu – polipyrolu (PAN-PPY) oraz polialkoholu winylowego lub polikwasu akrylowego (PVA-PAA) o średnicy 2,5 μm. Charakterystyka mięśnia siła-wydłużenie jest liniowa, w ciągu 20 ms mięśnie polimerowe kurczą się o 20%. Sztuczne mięśnie polimerowe w szybkości działania są najbardziej podobne do działania ludzkich mięśni, ale ich siła jednostkowa (na 1cm^2) jest dwa razy większa od siły mięśnia naturalnego.

– Sztuczne biomięśnie zbudowane z kompozytów polimerowych IPMC (*Ionic Polymer-Metal Composite*) i IPCC (*Ionic Polymer Conductor Composites*) wykazują zdolność odkształcenia pod wpływem prądu o małym napięciu, są lekkie, łatwo sterowalne, mają naturalne właściwości tłumienia drgań oraz są bardzo elastyczne.

– Sztuczne biomięśnie polimerowe żelowe wypełnione są płynem żelowym wrażliwym na zmianę liczby pH – IPG (*Ionic Polymeric Gels*) lub na zmianę prąd elektrycznego – EAPG (*Electro-Active Polymer Gel*) i CPG (*Contractile Polymer Gel*).

– Sztuczne mięśnie zbudowane z polimerów elektroaktywnych EAPAM (*Electroactive Polimers Artificial Muscles*), elektrostrykcyjnych EPAM (*Electrostrictive Polymer Artificial Muscles*) i dielektrycznych DEAMA (*Dielectric Elastomer Artificial Muscle Actuators*). Działanie tych mięśni polega na pobudzeniu prądowym dwóch elektrod, w wyniku tego powstaje różnica potencjałów elektrycznych wywołująca pole magnetyczne, które wpływa na zmianę wymiarów geometrycznych – zmniejszenie grubości i wydłużenie – filmu polimerowego lub elastomerowego. W filmie polimerowym i elastomerowym powstaje efekt ciśnieniowy Maxwella, zależny od podatności materiału.

– Sztuczne mięśnie wykorzystujące efekt piezoelektryczny stosowane są w mikrosystemach elektromechanicznych MEMS (*Micro Eletro Mechnical System*). Energia dostarczana do elementu piezoelektrycznego powoduje jego odkształcenie, wówczas piezoelement połączony z ruchomym elementem tworzy mikroukład mięśniowy.

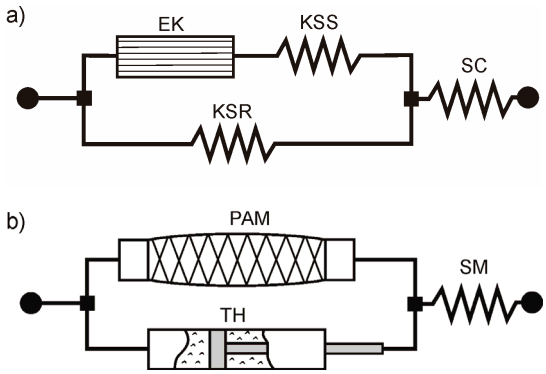
– Sztuczne mięśnie w kształcie sprężyny lub spirali z drutu wykonanego z materiałów o właściwościach z tzw. „pamięcią kształtu". W materiałach z „pamięcią kształtu" wykorzystuje się stopy Ti-Ni (SMA – *Shape Memory Alloy*) oraz stopy Ti-Ni i silikonu (SMAS – *Shape Memory Alloy–Silicone*).

– Sztuczne mięśnie płynowe wypełniane cieczą – sztuczne mięśnie hydrauliczne HAM (*Hydraulic Artificial Muscle*) lub gazem – sztuczne mieśnie pneumatyczne PAM (*Pneumatic Artificial Muscle*) wykonuje są z elastycznej przepony w kształcie rurki (balonu, mieszka lub worka), wykonanej z gumy, lateksu lub silikonu, oplecionej elastyczną siatką z rozciągliwego i elastycznego włókna silikonowego. Siatka mocowana na końcach mięśnia tworzy rodzaj sztucznych ścięgien. Właściwości mięśni płynowych uzasadniają ich porównanie do biologicznych mięśni szkieletowych. W mięśniach pneumatycznych jako czynnik roboczy wykorzystuje się zwykle powietrze, a w mięśniach hydraulicznych ciecze o własnościach elektroreologicznych i magnetoreologicznych. Skurcz tych mięśni zależy od ciśnienia płynu sterowanego zaworami [8].

Do wykonania sztucznych mięśni poszukuje się nowych adaptacyjnych materiałów o właściwościach elektroaktywnych (np. elektroaktywne polimery, elektroaktywna ceramika), piezoelektrycznych (np. piezopolimery) oraz elektro- i magentostrykcyjnych.

2. Modele kinematyczne sztucznego układu mięśniowego z aktuatorami płynowymi

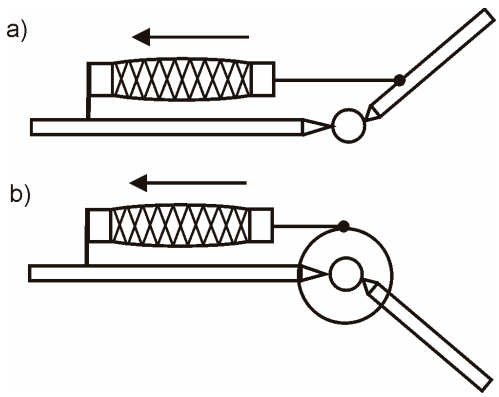
Budowę mięśni szkieletowych przedstawia się za pomocą modelu mięśniowego (mięśniowo-ścięgnowego) nazywanego modelem Hilla, w którym jeden z przyczepów jest nieruchomy (*punctum fixtum*), a drugi ruchomy (*punctum mobile*). Model Hilla porównano ze sztucznym modelem mięśniowo-ścięgnowym, w którym zastosowano pneumatyczny aktuator mięśniowy (rys. 1). W modelu Hilla występuje element kurczliwy EK, komponent sprężysty równoległy KSR, komponent sprężysty szeregowy KSS oraz ścięgno SC (rys. 1a). Element kurczliwy zbudowany jest z włókna mięśniowego zawierającego włókienka mięśniowe, z których każde podzielone jest przez krążki Z na sarkometry. Komponenty sprężyste równoległe, utworzone m.in. przez błonę włókna mięśniowego (sarkolemę) i tkankę łączną (powięzie), zapobiegają rozchodzeniu się włókna przy rozciąganiu biernym. Taki model mięśniowo–ścięgnowy może naśladować pracę grupy mięśniowej: mięśnia trójgłowego łydki (*triceps surae*), ścięgna Achillesa oraz mięśnia brzuchatego łydki (*gastrocnemius*), które odgrywają znaczącą rolę podczas chodzenia. W sztucznym modelu mięśniowo-cięgnowym (rys. 1b) zastosowano element wykonawczy – pneumatyczny aktuator mięśniowy, element tłumiący – tłumik hydrauliczny oraz element sprężysty mechaniczny – włókna elastyczne [10].



Rys. 1. Schemat modelu mięśniowego: a) model analogowy Hilla: EK – element kurczliwy, KSR – komponent sprężysty równoległy, KSS – komponent sprężysty szeregowy, SC – ścięgno, b) model sztuczny: PAM – pneumatyczny aktuator mięśniowy, TH – tłumik hydrauliczny, SM – sprężysty element mechaniczny (AI Lab. MIT USA)

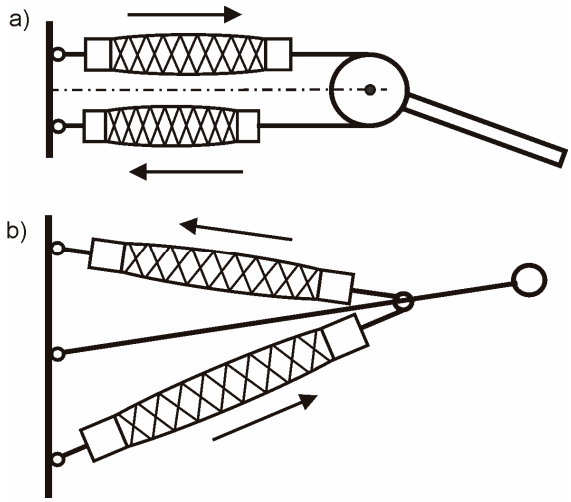
Modele kości długich stanowią dźwignie jednostronne lub dwustronne, do których przyczepione są sztuczne mięśnie rozciągane lub kurczące się. Każda dźwignia ma punkt podparcia oraz punkt przyłożenia sztucznego mięśnia. W przypadku zastosowania mięśni pneumatycznych, ich działanie polega na napełnieniu sprężonym powietrzem, w wyniku czego mięsień proporcjonalnie do ciśnienia pęcznieje i kurczy się. Powstałe w mięśniu pneumatycznym naprężenia wynikające z jego dużej sztywności wzdłużnej odpowiadają zewnętrznemu obciążeniu osiowemu. Przez regulację ciśnienia można zmieniać stopień skurczu mięśnia oraz siłę ciągnącą mięśnia. Siła ciągnąca mięśnia pneumatycznego w odniesieniu do jego przekroju poprzecznego jest dużo większa od siły mięśnia biologicznego. Znanych jest wiele sztucznych mięśni pneumatycznych (*Pneumatic Artificial Muscle*): Bladder Muscles (*Fluidic Muscle, Pneumatic Muscle Aktuator „McKibben"*, *Sleeved Bladder Muscle*), Platted Muscles (*Platted PAM*, *Flexator, Fiber Muscle*, *RPM*), Netted Muscles (*Yarlott Muscle*, *ROMAC*, *Kukoij Muscle*), Embedded Muscles (*Baldwin Muscle*, *Morin Muscle*, *Paynter Hyperboloid Muscle*, *Paynter Knitted Muscle*, *UPAM*, *Kleinwachter torsion device*).

Schemat kinematyczny dźwigni (sztucznego ramienia i stawu) poruszanej pojedynczym mięśniem pneumatycznym o kształcie elastycznego przewodu gumowego (np. *Fluidic Muscle*, *McKibben*, *Kukolj Muscle*), pracującym jako zginacz (*flexor*) i prostownik (*extensor*) przedstawiono na rys. 2 [5].

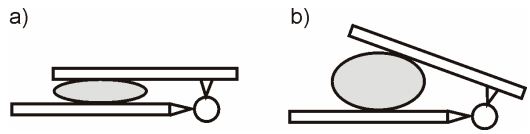


Rys. 2. Schemat kinematyczny dźwigni z pojedynczym mięśniem pneumatycznym: a) zginacz, b) prostownik

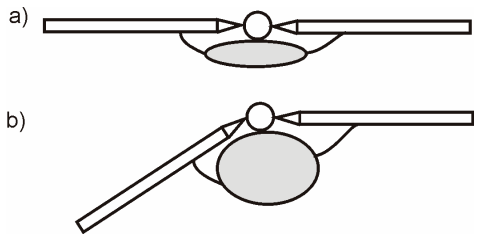
Dwa schematy kinematyczne dźwigni (sztucznego ramienia i stawu) poruszanej dwoma przeciwstawnie działającymi pneumatycznymi aktuatorami mięśniowymi zamieszczono na rys. 3 [4]. Tego typu układ dwumięśniowy BMDS (*Bi-Muscular Driving System*) jest porównywany do pracy dwóch antagonistycznych mięśni biologicznych. Schematy kinematyczne dźwigni rozpieranej i ściąganej mięśniami w kształcie mieszka lub worka (*Plated PAM*, *Bag Muscle*, *ROMAC*, *Yarlott Muscle*, *Baldwin Muscle*) przedstawiono na rys. 4 i 5.



Rys. 3. Schematy kinematyczne dźwigni z dwoma przeciwstawnie działającymi mięśniami pneumatycznymi



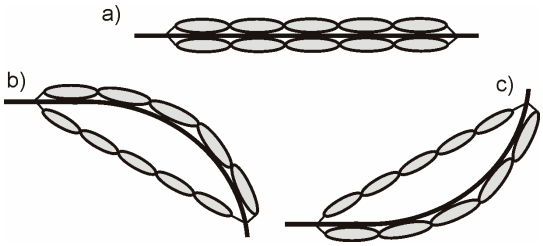
Rys. 4. Schemat kinematyczny dźwigni rozpieranej mięśniem pneumatycznym: a) w stanie początkowym, b) w stanie napełnienia



Rys. 5. Schemat kinematyczny dźwigni ściąganej mięśniem pneumatycznym: a) w stanie początkowym, b) w stanie napełnienia

Na rys. 6 przedstawiono schemat kinematyczny elastycznego ramienia z foliowymi mięśniami płynowymi w stanie spoczynku oraz w stanach skurczu. Mięśnie foliowe – folie polimerowe zgrzewane laserem, wypełnione płynem o dużej rozszerzalności objętościowej – połączone są szeregowo z dwóch stron elastycznego ramienia. Mięśnie foliowe mogą być wypełnione powietrzem (mięsień pneumatyczny) lub płynem (mięsień hydrauliczny). W mięśniach hydraulicznych stosuje się płyny elektroreologiczne ER. Elastyczne ramie odchyla się proporcjonalnie do skurczu mięśni foliowych (maksymalny kąt zgięcia powyżej 90°). Stopień skurczu h i siła skurczu F (F_{max} = 0,5 N) mięśni foliowych zależą od ciśnienia płynu (powietrza, cieczy) sterowanego zaworami. Mięśnie

foliowe znalazły zastosowanie w technice medycznej, np. terapii laserowej i endoskopii [17].



Rys. 6. Schemat kinematyczny elastycznego ramienia z mięśniami foliowymi w stanie spoczynku (a) oraz w stanach skurczu (b) i (c)

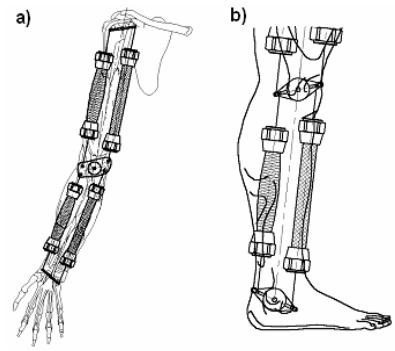
3. Zastosowanie płynowych aktuatorów mięśniowych w protezach kończyn

Tradycyjne protezy kończyn wykonane z drewna, tworzywa sztucznego i metalu, o budowie jednolitej lub szkieletowej zastępuje się nowoczesnymi materiałami z włókna węglowego, tworzyw termoplastycznych i lekkich metali. Prowadzone są także doświadczenia nad nowymi neuroprotezami (ang. *Stand/Go*, niem. *Steh auf und Geh*). W praktyce rehabilitacyjnej dąży się w coraz większym stopniu do wszczepiania sztucznych mięśni. Sztuczne mięśnie (aktuatory mięśniowe) mogą zastąpić biologiczne mięśnie w pracy naturalnych organów lub wspomagać ruch człowieka za pomocą protez, ortez lub egzoszkieletonów [2], [3].

Przy projektowaniu i wykonywaniu protez kończyn wykorzystuje się zasady sterowania ruchem układu mięśniowego, polegające na współdziałaniu mechanizmów fizjologicznych z mechanizmami protez i ich układem sterującym. Zastosowanie protezy lub innych aparatów ortopedycznych wiąże się z rozpoznaniem możliwości i potrzeb ruchowych każdego pacjenta. Proteza jest uproszczonym modelem fizycznym utraconej kończyny, której działanie jest bardziej lub mniej zbliżone do funkcjonowania kończyny naturalnej. Proteza powinna być tak wykonana, aby umożliwiała sprawne poruszanie się przy minimalnym zużyciu energii, najlepiej na poziomie metabolicznym. Tradycyjne napędy elektryczne, pneumatyczne i hydrauliczne wypierane są przez sztuczne mięśnie, które umożliwiają zmniejszenie ciężaru i wymiarów protezy oraz zminimalizowanie energii potrzebnej do ich poruszania i sterowania. Dzięki rozwojowi elektroniki i sterowania mikroprocesorowego obserwuje się znaczny rozwój techniki płynowej w skali zminiaturyzowanej i mikro (mikropneumatyki i mikrohydrauliki) w aparatach ortotycznych i protetycznych oraz urządzeniach rehabilitacyjnych. Aktuatory płynowe znajdują już powszechne zastosowanie w protezach i ortezach kończyn górnych i dolnych oraz egzoszkieletonach. Znane są protezy kolana: C-LEG, Seattle Power Knee. Budowa egzoszkieletonu przypomina swoim kształtem szkielet ręki, nogi lub całego człowieka, jak aparat SKIL MATE.

W wielu ośrodków naukowo-badawczych dąży się do stworzenie sztucznej ręki (protezy ręki) o funkcjonalności i podobieństwie do ruchu ręki ludzkiej. Trudności ze stworzeniem sztucznej ręki wynikają z tego, iż ludzka ręka składa się z 18 członów oraz 17 połączeń ruchomych, ma 22 stopni swobody i 23 stopnie ruchliwości. Konstrukcja sztucznej ręki powinna być dostosowana do ciężaru i wymiarów naturalnej ręki. Proteza ręka musi być łatwo sterowalna przy małym zużyciu energii, nie może także emitować nadmiernego hałasu. Konstrukcja protezy ręki i system jej sterowania muszą umożliwiać chwytanie dowolnego przedmiotu z dużą dokładnością, dynamiką i odpowiednią siłą, w sposób zsynchronizowany przez niezależny ruch palców i kciuka. Sztuczne ręce (chwytaki w kształcie ludzkiej ręki) stosowane jako biochwytaki w robotach antropomorficznych i humanoidalnych mają od trzech do pięciu palców. Od 1999 roku Unia Europejska finansuje projekt SYNERAGH (*Systems Neuroscience and Engineering Rresearch for Anthropomorphie Grasping and Handling*), który jest realizowany przez pięć jednostek badawczych z różnych krajów. Projekt ten dotyczy planowania procesów w chwytakach robotów humanoidalnych z uwzględnieniem zagadnień technicznych, biomechanicznych, neurologicznych i fizjologicznych. Projekt SYNERAGH nie rozwiązał jeszcze takich problemów, jak realizacja receptorów skóry za pomocą sensorów dotyku oraz algorytmy sterowania chwytem ludzkiej ręki. Opracowane w tym projekcie hipotezy i teorie neurofizjologiczne wykorzystano w programie komputerowym GSL (*Grasping Simulation Language*), który został zastosowany do planowania zadań manipulacyjnych sztucznego ramienia z wykorzystaniem architektury sieci neuronowej. Program GSL zastosowano do sterowania ruchem palców, ręki, ramienia, koordynacji ruchu ręka-ramię oraz informacji o ruchu ręki. W programie GSL wykorzystuje się cztery sztuczne sieci neuronowe, które korespondują z czterema ośrodkami (polami) mózgu człowieka, odpowiedzialnymi za czynności: ruchowe (motoryczne), czuciowe (somatyczne), wzrokowe i przejściowe pomiędzy czuciowymi a wzrokowymi oraz czynności wyższe (związane ze sferą inteligencji).

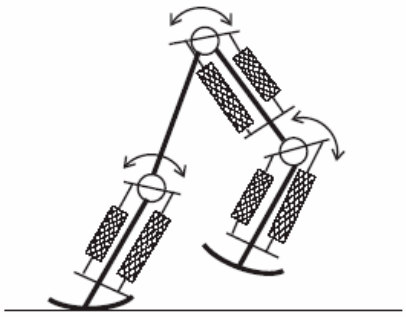
Zastosowanie niektórych typów pneumatycznych aktuatorów mięśniowych w protezach kończyn jest ograniczone dostępnością ich wymiarów oraz koniecznością sterowania sprężonym powietrzem, dostarczanym ze sprężarki lub butli. Dokładne pozycjonowanie sztucznego układu mięśniowego wymaga stosowania dodatkowych systemów pomiarowych lub wizyjnych. Wiele przykładów wykorzystania mięśni pneumatycznych PAM McKibben firmy Bridgestone Rubber Company (Japania), jako elementu wykonawczego – aktuatora w protezach ręki, nogi i dłoni można znaleźć w licznej literaturze, np. [7]. Efektywne wykorzystanie mięśni pneumatycznych jest możliwe tylko przez zastosowanie układu dwóch mięśni działających przeciwstawnie, mięśni anatgonistycznych (np. bicesps – tricesps). W protezach jeden z mięśni nazywa się zginaczem – przywodzicielem, a drugi prostownikiem – odwodzicielem. Układ mięśniowy ramienia i nogi porównywalny do pracy dwóch mięśni biologicznych przedstawiono na rys. 7.



Rys. 7. Układ mięśniowy ręki (a) i nogi (b) poruszany dwoma przeciwstawnie działającymi aktuatorami pneumatycznymi

Częstotliwość pracy takiego układu mięśniowego wynosi do 5 Hz przy dokładności regulacji położenia 1%. Dokładność kąta obrotu protezy zależy od stosunku ciśnień panujących w obu mięśniach. Obrót protezy jest ograniczony dopuszczalnym stopniem skrócenie mięśni pneumatycznych.

Na rys. 8 przedstawiono schemat kinematyczny dwunożnego układu mięśniowego, który składa się z trzech stawów: biodrowego i dwóch kolanowych. Każdy staw obrotowy o jednym stopniu swobody połączony jest z dwoma pneumatycznymi aktuatorami mięśniowymi.



Rys. 8. Dwunożny układ mięśniowy z aktuatorami pneumatycznymi (HITACHI Medical Corporation)

Widok protezy nogi z pneumatycznymi sztucznymi mięśniami „McKibben": wykonanej w Centrum Medyczno-Rehabilitacyjnym w Seattle (USA) przedstawiono na rys. 9. Mięśnie zostały zintegrowane z systemem pomiarowym i sterującym. Taki system napędowo–sterujący umożliwił poprawę symetrii chodu i zredukowanie energii potrzebnej do ruchu protezy. Protezę testowano za pomocą przyrządu Bionix (MTS Systemy Corp. Minnesota, U.S.). Mierzono siłę przy różnej szybkości i różnej aktywności ruchowej pacjenta. Przy izometrycznym napięciu mięśnia maksymalna siła wynosiła 1600 N. Podczas fazy wydłużania siła zwiększała się do 2000 N przy maksymalnej prędkości 300 mm/s. Podczas fazy skurczu siła zmniejszała się wraz ze wzrostem prędkości skurczu. Siła była zerowa przy maksymalnej prędkości skurczu. Otrzymane charakterystyki muskułu sztucznego w układzie trójwymiarowym: siła muskułu – długość muskułu – prędkość skurczu były podobne do charakterystyki mięśnia trójgłowego łydki (*triceps surae*) i ścięgna Achillesa. Przeprowadzone badania potwierdziły, że sztuczny model mięśniowo–ścięgnowy z mięśniem McKibben jako elementem wykonawczym, dobrze imituje funkcję mięśnia trójgłowego łydki i ścięgna Achillesa.

Rys. 9. Widok protezy kończyny dolnej ze sztucznymi mięśniami pneumatycznymi „McKibben" [14]

Aktualnie prowadzone badania zmierzają do wykonania systemu sterowania protezy napędzanej sztucznym układem mięśniowo–ścięgnowym z aktuatorami pneumatycznymi, przy wykorzystaniu sztucznych sieci neuronowych SSN. Schemat takiego układu sterowania zamieszczono na rys. 10.

Rys. 10. Schemat systemu sterowania segmentu protezy ze sztucznym układem mięśniowo–ścięgnowym: SSN – sztuczne sieci neuronowe, F – siła, L – długość. v – prędkość

Widok ortezy nogi ze sztucznymi muskułami pneumatycznymi „McKibben" wykonanej w Centrum Protetyki na Uniwersytecie Michigan (USA) przedstawiono na rys. 11 [15]. Orteza uruchamiana jest dwoma sztucznymi mięśniami pneumatycznymi przymocowywanymi do kostki i stopy ortezy. Jeden jest mięśniem zginaczem grzbietowym (*dorsiflexor*), a drugi mięśniem zginaczem stopy (*plantaflexior*). Podczas ruchu ortezy pierwszy kurczy się, a drugi rozciąga. Orteza wyposażona została w przetwornik siły, który służy do pomiaru napięcia skurczu mięśnia i sterowania cyklem pracy sztucznych mięśni pneumatycznych. Przeprowadzone testy potwierdziły, że mechaniczne właściwości ortezy są bardzo podobne do działania ludzkich mięśni. Po przesterowaniu zaworu stan szczytowego napięcia – skurczu mięśnia – osiągano w czasie 69 ms, po tym następował stan odprężenia – rozkurcz mięśnia w czasie 69 ms. Dla porównania – trójgłowy mięsień łydki (*triceps surae*) ma odpowiednio cykl pracy 101 ms i 94 ms, natomiast mięsień piszczelowy ma cykl pracy 99 ms i 87 ms.

Rys. 11. Widok ortezy kończyny dolnej ze sztucznymi mięśniami pneumatycznymi „McKibben" [13]

4. Modele statyczne pneumatycznych aktuatorów mięśniowych

Mięśnie pneumatyczne zbudowane są z kurczliwej membrany i zakończone odpowiednimi elementami złącznymi. Po podaniu sprężonego powietrza następuje zwiększenie ich objętości i jednoczesne skrócenie, w wyniku czego zostaje wytworzona siła ciągnąca w osi muskułu. Powstałe w mięśniu pneumatycznym naprężenia odpowiadają zewnętrznemu obciążeniu osiowemu. Tak generowana siła zależy od ciśnienia panującego wewnątrz mięśnia, długości początkowej, stopnia skrócenia oraz jego własności materiałowych. W fazie początkowej skrócenia mięsień przenosi największą

siłę, która maleje do zera przy maksymalnym jego skróceniu. Poprzez regulację ciśnienia można kontrolować przenoszoną siłę i stopień skrócenia mięśnia pneumatycznego.

Pneumatyczne aktuatory mięśniowe charakteryzuje się następującymi wskaźnikami [4], [12]:

– wskaźnik mocy P mięśnia do jego masy m:

$$K_{Pm} = \frac{P}{m} \quad \frac{\mathrm{kW}}{\mathrm{kg}} \tag{1}$$

– wskaźnik mocy P mięśnia do jego objętości V:

$$K_{PV} = \frac{P}{V} \quad \frac{\mathrm{W}}{\mathrm{cm}^3} \tag{2}$$

– wskaźnik siły F mięśnia do jego przekroju A poprzecznego:

$$K_{FA} = \frac{F}{A} \quad \frac{\mathrm{N}}{\mathrm{cm}^2} \tag{3}$$

Ważnymi parametrami sztucznych mięśni są także [4]:

– sztywność K określana stosunkiem siły F do skrócenia ΔL mięśnia:

$$K = \frac{F}{\Delta L} \quad \frac{\mathrm{N}}{\mathrm{m}} \tag{4}$$

– sztywność względna K_F, jako stosunek sztywności K mięśnia do jego siły F:

$$K_F = \frac{K}{F} \quad \frac{1}{\mathrm{m}} \tag{5}$$

– sztywność względna K_p jako stosunek sztywności K mięśnia do ciśnienia p powietrza:

$$K_p = \frac{K}{p} \quad \mathrm{m} \tag{6}$$

Teoretyczny zakresu pracy mięśni pneumatycznych ograniczony jest parametrami:

$$\begin{cases} F = F_{\max} & \text{dla } h = h_{\max w} \\ F = F_0 & \text{dla } h = 0 \\ F = 0 & \text{dla } h = h_{\max s} \end{cases} \tag{7}$$

Stopień skrócenia – stopień skurczu h mięśnia pneumatycznego określa się zależnością:

$$h = \frac{L_n - L}{L_n} = \frac{\Delta L}{L_n} \tag{9}$$

gdzie: L – długość mięśnia, L_n – długość nominalna mięśnia, ΔL – skrócenie mięśnia.

Długość mięśnia wynika więc z zależności (9):

$$L = L_n(1-h) \tag{10}$$

Teoretyczna wartość siły mięśnia pneumatycznego $F(p,h)$ wynika z jego analizy geometrycznej i energetycznej [4], [5]:

$$F(p,h) = p\,D^2\left[\alpha(1-h)^n - \beta\right] \tag{11}$$

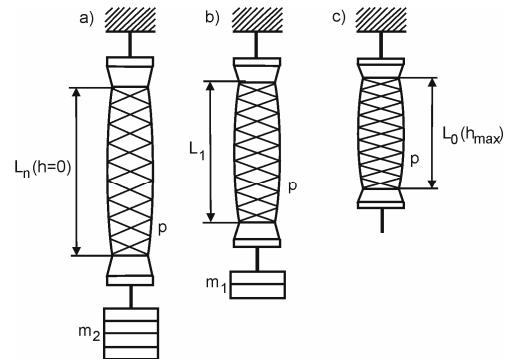
gdzie: D – średnica wewnętrzna mięśnia, p – ciśnienie wewnątrz mięśnia, α,β – stałe mięśnia zależne od właściwości materiałowych mięśnia, n – wykładnik potęgi.

Wzór (11) przy stałych współczynnikach α i β oraz wykładniku $n = 2$ sprawdza się w modelowaniu statycznym mięśni pneumatycznych „McKibben". W przypadku modelowania statycznego mięśni pneumatycznych Fluidic Muscle typu MAS firmy FESTO współczynniki materiałowe α i β oraz wykładnik n występujące we wzorze (11) można w przybliżeniu określić za pomocą programu komputerowego *MuscleSim* lub na podstawie badań doświadczalnych.

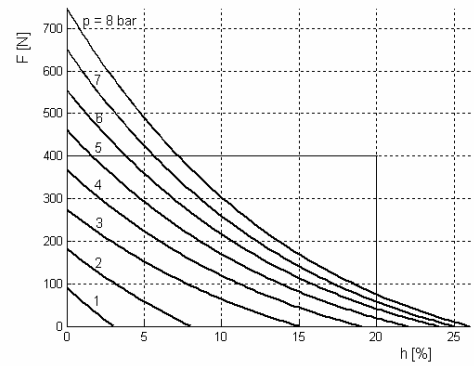
Dla mięśnia pneumatycznego Fluidic Muscle typu MAS-10-300 (o średnicy wewnętrznej D = 10 mm i długości początkowej L_o = 300 mm) wyznaczono podstawowe charakterystyki statyczne: izobaryczne, izotoniczne i izotermiczne.

4.1. Charakterystyki izobaryczne

Zasadę wyznaczania charakterystyk izobarycznych (dla p = const) mięśnia pneumatycznego ilustruje rys. 12. Przy stałym ciśnieniu p mierzy się stopień skrócenia h mięśnia w zależności od jego obciążenia masowego F. Charakterystyki izobaryczne $F(h)$ przedstawiające zależności siły F od jego stopnia skrócenia h, wyznaczone dla stałych wartości ciśnienia p = const wewnątrz muskułu za pomocą wzoru (11) przedstawiono na rys. 13. Na rysunku zaznaczono dopuszczalny zakres pracy mięśnia MAS-10-300, który ograniczony jest siłą ciągnącą F = 400 N, stopniem skrócenia h = 20 % i maksymalnym ciśnieniem p_{max} = 8 bar.



Rys. 12. Zasada wyznaczania charakterystyk izobarycznych (p = const) mięśni pneumatycznych



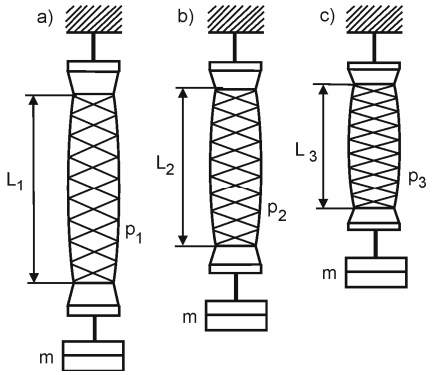
Rys. 13. Charakterystyki izobaryczne mięśnia pneumatycznego typu MAS-10-30

4.2. Charakterystyki izotoniczne

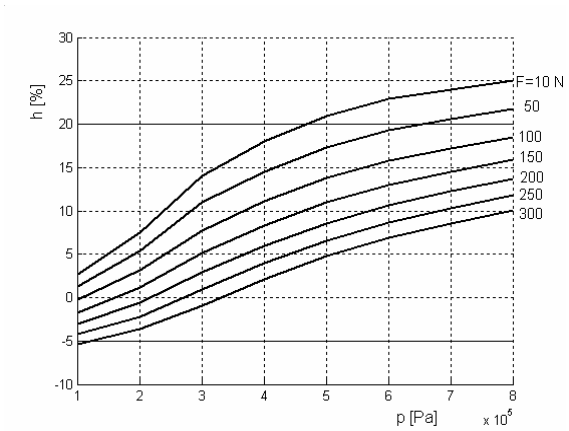
Zasadę wyznaczania charakterystyk izotonicznych (dla F = const) mięśnia pneumatycznego ilustruje rys. 14. Przy stałym obciążeniu F mięśnia mierzy się stopień jego skrócenia h w zależności od ciśnienia p. Charakterystyki izotoniczne $h(p)$ przedstawiające stopień skrócenia mięśnia h w funkcji ciśnienia p dla stałej siły ciągnącej F = const, wyznaczone ze wzoru:

$$h = 1 - \left[\frac{1}{\alpha(p)}\left(\frac{F}{p\,D^2} + \beta(p)\right)\right]^{1/n} \tag{12}$$

przedstawiono na rys. 15. Charakterystyki te ograniczone są maksymalnym dopuszczalnym względnym skróceniem h = 20% i maksymalnym dopuszczalnym względnym wydłużeniem h = –5%.



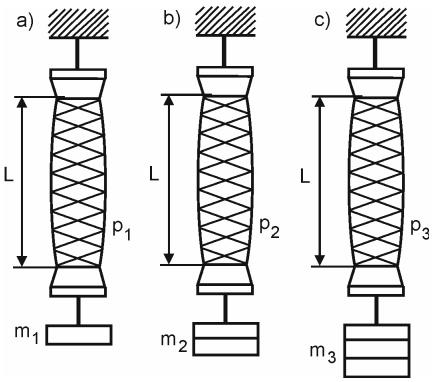
Rys. 14. Zasada wyznaczania charakterystyk izotonicznych (F = const) mięśni pneumatycznych



Rys. 15. Charakterystyki izotoniczne mięśnia pneumatycznego typu MAS-10-30

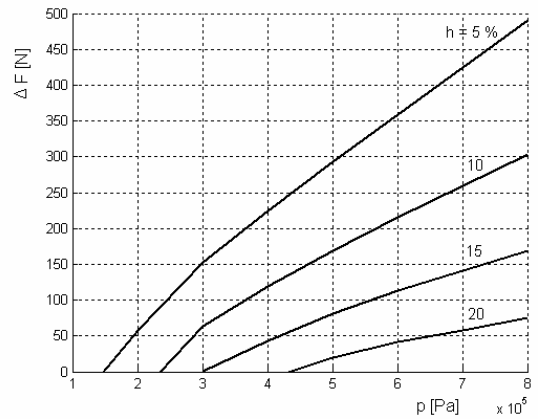
4.3. Charakterystyki izotometryczne

Zasadę wyznaczania charakterystyk izometrycznych (dla h = const) mięśnia pneumatycznego ilustruje rys. 16. Przy stałym stopniu skrócenia h mięśnia mierzy się ciśnienie p w zależności od obciążenie F mięśnia.



Rys. 16. Zasada wyznaczania charakterystyk izometrycznych (h = const) mięśni pneumatycznych

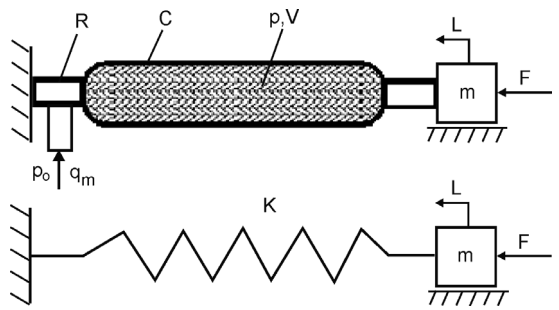
Charakterystyki izometryczne przedstawiające przyrost siły mięśnia ΔF w funkcji ciśnienia p, wyznaczone dla h = const za pomocą wzoru (11), przedstawiono na rys. 17.



Rys. 17. Charakterystyki izometryczne mięśnia pneumatycznego typu MAS-10-30

5. Modele dynamiczne pneumatycznych aktuatorów mięśniowych

Analiza dynamiczna pneumatycznych akuatorów mięśniowych „McKibben" oparta została na modelu dynamicznym przedstawionym na rys.18 [6].



Rys. 18. Model dynamiczny mięśnia pneumatycznego

W modelu dynamicznym pneumatycznego aktuatora mięśniowego uwzględniono przepływ masowy przez przyłącze mięśnia, równanie ciśnienia i równanie sił działających na mięsień.

– Równanie ciągłości przepływu przez przyłącze mięśnia:

$$q_m = \frac{1}{R_p}\left(p_o - p\right) \qquad (13)$$

gdzie: q_m – strumień masowy, p_o – ciśnienie zasilania, p – ciśnienie panujące wewnątrz mięśnia, R_p – opór pneumatyczny.

Na podstawie I zasady termodynamiki napisano równanie ciśnienia w mięśniu pneumatycznym:

$$\dot{p} = \frac{\kappa}{V}\left[R\,T\,q_m - p\,\dot{V}\right] = \frac{\kappa}{V}\left[R\,T\;\;q_m - \frac{dV}{dL}\,p\,\dot{L}\right] \qquad (14)$$

Po uwzględnieniu przepływu masowego i pochodnej *dV/dL* otrzymano następujące równanie [6]:

$$\dot{p} = \frac{1}{R_p\, C_p}\left(p_o - p\right) - \kappa\, \frac{\left(b^2 - 3\, L^2\right)}{L\left(b^2 - L^2\right)}\, p\, \dot{L} \qquad (15)$$

gdzie: C_p – pojemność pneumatyczna mięśnia,

$$C_p = \frac{V(L)}{\kappa\, R\, T} \qquad (16)$$

R – stała gazowa, T – temperatura gazu, κ – wykładnik adiabaty, V – objętość mięśnia,

$$V = \frac{L\left(b^2 - L^2\right)}{4\,\pi\, n^2} \qquad (17)$$

L – długość mięśnia, b – długość splotu mięśnia, n – ilość splotów mięśnia.

Równanie (15) można przedstawić w postaci uproszczonej jako filtr dolnoprzepustowy z pulsacją $\omega = 1/(R_p\, C)$:

$$\dot{p} = \omega\, \left(p_o - p\right) \qquad (18)$$

– Równanie sił działających na mięsień pneumatyczny według modelu dynamicznego z rys. 17 ma postać:

$$m\, \ddot{L} = F - K\, L - F_t \qquad (19)$$

gdzie: F – siła ciągnąca wzdłużna, m – masa obciążenia, K – sztywność mięśnia,

$$K = \frac{dF}{d\,L} \approx \frac{3\, p\, L}{2\,\pi\, n^2} \qquad (20)$$

F_f – siła tarcia,

$$F_t = F_w + F_v \qquad (21)$$

F_w – siła tarcia wewnętrznego, F_v – siła tarcia lepkiego,

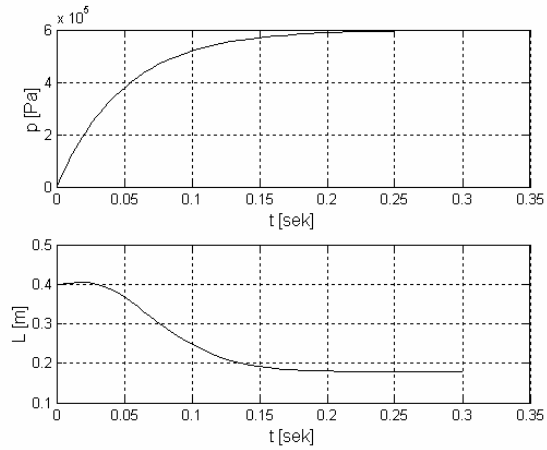
$$F_v = f\, \dot{L} \qquad (22)$$

f – współczynnik tarcia lepkiego.

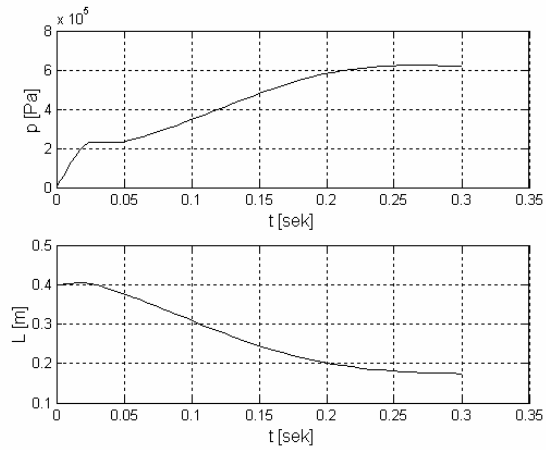
Po uwzględnieniu (20), (21) i (22) w równaniu (19) ostatecznie otrzymano:

$$m\, \ddot{L} + f\, \dot{L} + F_w\, \mathrm{sgn}(\dot{L}) + \frac{3}{2\,\pi\, n^2}\, p\, L^2 = F \qquad (23)$$

Z charakterystyk statycznych dobrano parametry początkowe mięśnia. Podczas symulacji przyjęto wartości parametrów: m = 20 kg, F = 1000 N i F_w = 2,5 N. Charakterystyki dynamiczne ciśnienia $p(t)$ i długości $L(t)$ mięśnia pneumatycznego dlamodeli I według równań (18) i (23) przedstawiono na rys. 19, a dla modelu II według równań (15) i (23) przedstawiono na rys. 20.



Rys. 19. Charakterystyki ciśnienia $p(t)$ i długości $L(t)$ mięśnia dla modelu I



Rys. 20. Charakterystyki ciśnienia $p(t)$ i długości $L(t)$ mięśnia dla modelu II

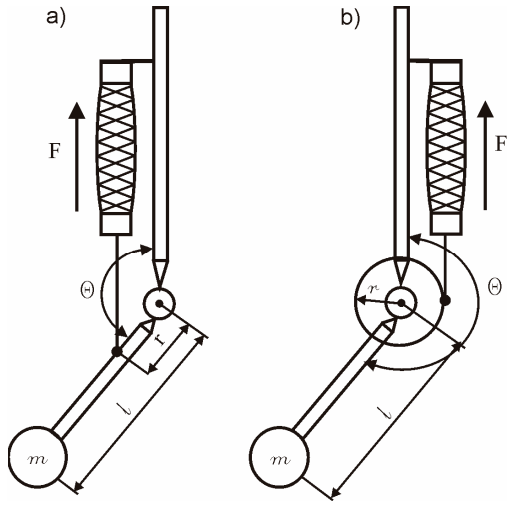
6. Model egzoszkieletonu z pneumatycznymi aktuatorami mięśniowymi

Na rys. 21 przedstawiono schematy obliczeniowe egzoszkieletonów z pneumatycznym aktuatorem mięśniowym w konfiguracji z mięśniem zginaczem – bicepsem i mięśniem prostownikiem – tricepsem. W dalszej części pracy przedstawiono przykładowy model obliczeniowy egzoszkieletonu w konfiguracji z mięśniem zginaczem – bicepsem. Na podstawie zależności geometrycznych przedstawionych na schemacie rys. 20a napisano nieliniowe równania dynamiczne egzoszkieletonu i przeprowadzono jego symulację cyfrową.

– Równanie momentów działających na ramię egzoszkieletonu:

$$I\, \ddot{\theta} + f\, \dot{\theta} + F\, r\, \sin\theta = mgl\, \sin\theta \qquad (23)$$

gdzie: I – moment bezwładności $I = m\, l^2$, m – masa obciążająca, l – długość ramienia, f – współczynnik tarcia lepkiego w przegubie, F – siła mięśnia, r – promień przyczepu mięśnia, θ – kąt ugięcia ramienia.



Rys. 21. Schemat obliczeniowy egzoszkieletonów z pneumatycznym aktuatorem mięśniowym w konfiguracji z mięśniem zginaczem – bicepsem (a) i mięśniem prostownikiem – tricepsem (b)

– Równanie siły pneumatycznego aktuatora mięśniowego:

$$F = K\ L = a\ p\ L^2 \tag{24}$$

gdzie: K – sztywność mięśnia wg (20), p – ciśnienie wewnątrz mięśnia, a – stała mięśnia.

$$a = \frac{2}{3\,\pi\ n^2} \tag{25}$$

L – długość mięśnia,

$$L = L_o - \Delta L \tag{26}$$

L_o – długość początkowa mięśnia, ΔL – skrócenie mięśnia,

$$\Delta L = r(1 - \cos\theta) \tag{27}$$

Z równania (27) wynikają warunki kinematyczne dla analizowanego ramienia egzoszkieletonu:

$\Delta L = 0$ dla $\theta = 0^o$ – ramię wyprostowane,

$\Delta L = r$ dla $\theta = 90^o$ – ramię zgięte pod kątem prostym,

$\Delta L = 2r$ dla $\theta = 180^o$ – ramię całkowicie zgięte.

Na podstawie tych warunków, dla przyjętego maksymalnego skrócenie mięśnia $\Delta L_{\max} = 2\ r = 0{,}2\ L_o$, wynika zależność $r = 0{,}1\ L_o$.

Po uwzględnieniu zależności kinematycznych, równanie ruchu ramienia egzoszkieletonu przyjmie postać:

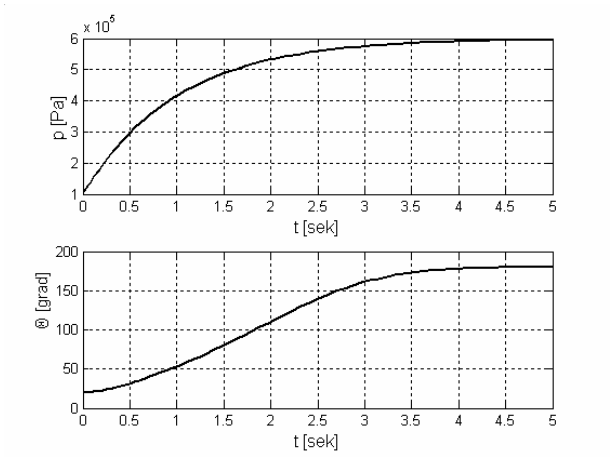
$$I\ \ddot{\theta} + f\ \dot{\theta} + a\ p\ \left[L_o - r(1-\cos\theta)\right]^2 r\sin\theta = mgl\ \sin\theta \tag{28}$$

– Równanie ciśnienia członu sterującego typu PT1:

$$\dot{p} = \frac{1}{T}\left(K\ p_o - p\right) \tag{29}$$

gdzie: p_o – ciśnienie sterowane zaworem ciśnieniowym, K – współczynnik wzmocnienia, $T = R\ C$ – stała czasowa.

Przyjęte warunki początkowe wynikają z charakterystyk statycznych pneumatycznego aktuatora mięśniowego. Dla parametrów: $L_o = 0{,}5$ m, $l = 1$ m, $m = 50$ kg oraz $\theta_0 = 20^\circ$ przeprowadzono symulację ruchu ramienia egzoszkieletonu w konfiguracji z bicespsem. Przykładowe charakterystyki zmiany ciśnienia $p(t)$ oraz kąta obrotu $\theta(t)$ ramienia tego egzoszkieletonu przedstawiono na rys. 22.



Rys. 22. Charakterystyki ciśnienia $p(t)$ i kąta obrotu $\theta(t)$ ramienia egzoszkieletonu

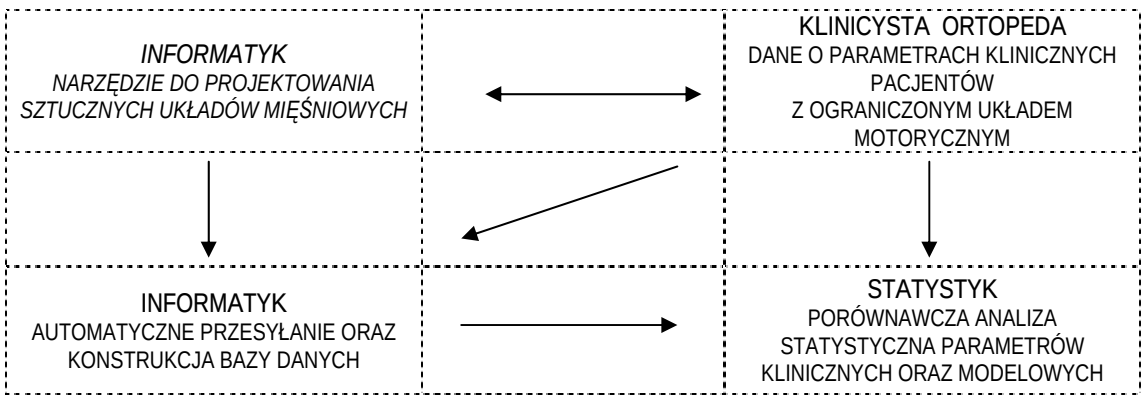
Przedstawimy model dynamiczny egzoszkieletonu z pneumatycznym aktuatorem mięśniowym pracującym jak mięsień zginacz – bicesps oraz uzyskane wyniki symulacyjne zostaną zweryfikowane doświadczalnie na stanowisku badawczym, budowanym w Zakładzie Mechatroniki Politechniki Świętokrzyskiej [12].

7. Podsumowanie

W pracy przedstawiono możliwości zastosowania sztucznych mięśni w inżynierii medycznej i rehabilitacyjnej. Scharakteryzowano najnowsze osiągnięcia i kierunki rozwoju sztucznych mięśni, zwłaszcza w zakresie technologii wytwarzania oraz techniki sterowania. Szczególnie analizowano możliwości zastosowania różnych sztucznych mięśni płynowych (pneumatycznych i hydraulicznych) w zaopatrzeniu ortopedycznym. Zamieszczono przykładowe rozwiązania konstrukcyjne protezy i ortezy nogi z pneumatycznymi aktuatorami mięśniowymi, które umożliwiają ich poprawny, symetryczny i stabilny ruch. W sterowaniu układów mięśniowo-ścięgnowych wykorzystuje się sztuczne sieci neuronowe, które mogą współdziałać z ośrodkowym systemem nerwowym. Systemy kinematyczne sztucznego układu mięśniowego (mięśniowo– ścięgnowego), oparte na pneumatycznych aktuatorach mięśniowych, umożliwiają rozwój nie tylko protez kończyn dolnych i górnych, ale także egzoszkieletonów ramion i nóg, sztucznych rąk, robotów antropomorficznych oraz różnych aparatów ortopedycznych i urządzeń rehabilitacyjnych [11], [16]. Parametry i wskaźniki energetyczne określane dla sztucznych mięśni przewyższają znacznie parametry mięśni biologicznych, a przy tym zużywają minimalną ilość energii. Niektóre rodzaje sztucznych mięśni mogą mieć zapotrzebowanie energetyczne na poziomie metabolizmu człowieka. Z protezami kończyn nowej generacji pacjenci mogą prowadzić aktywne życie zawodowe, a także uprawiać ćwiczenia rekreacyjne oraz sportowe. Sztuczny układ mięśniowy z różnymi aktuatorami (pneumatycznymi i innymi) daje ludziom starszym, chorym i niepełnosprawnym nadzieję na poprawę warunków życia.

Literatura cytowana

1. Dindorf R., Wołkow J.: Układy płynowe w inżynierii medycznej, Ossolineum, Wrocław 1999.
2. Dindorf R.: Zastosowanie elementów płynowych w protezach kończyn., Seminarium Naukowe Mechanika w Medycynie, Rzeszów 8-9 września 2000.
3. Dindorf R., Wołkow J.: Elementy hydrauliczne i pneumatyczne w protetyce ortopedycznej, Hydraulika i Pneumatyka, 5, 2001.
4. Dindorf R.: Muskuły pneumatyczne. Charakterystyki statyczne, Pneumatyka, 4, 2003.
5. Dindorf R.: Static and dynamic models of pneumatic muscle actuator, Proc. The 18th International Conference on Hydraulics and Pneumatics – ICHP'2003, Prague (Czech Republic), September 30 – October 1, 2003.
6. Dindorf R.: Muskuły pneumatyczne. Charakterystyki dynamiczne, Pneumatyka, 5, 2003.
7. Dindorf R.: Model i charakterystyki muskułów pneumatycznych, Pomiary Automatyka Robotyka, 2, 2004.
8. Dindorf R., Wołkow J.: Zastosowanie muskułów pneumatycznych w inżynierii medycznej, III Krakowskie Warsztaty Inżynierii Medycznej, Kraków 13-14 maja 2004.
9. Dindorf R., Wołkow J.: Sztuczne muskuły jako nowe bioelementy, III Krakowskie Warsztaty Inżynierii Medycznej, Kraków, 13-14 maja 2004.
10. Dindorf R.: Rozwój zaopatrzenia ortopedycznego z elementami płynowymi, Pomiary Automatyka Robotyka, 6, 2004.
11. Dindorf R.: Rozwój i zastosowanie manipulatorów i robotów rehabilitacyjnych, Pomiary Automatyka Robotyka, 11, 2004.
12. Dindorf R.: Stanowisko do badania muskułów pneumatycznych typu MAS, Pneumatyka, 5, 2004.
13. Ferris D.: A robotic exoskeleton for gait rehabilitation, Dept. of Movement Science and Dept. of Biomedical Engineering, The University of Michigan Orthotics and Prosthetics Center, Michigan 2002.
14. Klute G. K., Czerniecki J. M., Hannaford,B.: Powering lower limb prosthetics with muscle-like actuators, The 1st National Meeting of the Veterans Affairs Rehabilitation Research and Development Service, Washington, D. C., October 1-3, 1998.
15. Lilly J.: Adaptive tracking for pneumatic muscle actuators in biceps and tricesps configurations, IEEE Transactions Neural Systems and Rehebilitation Engineering, 3, 2003.
16. Rice C. L., Cunningham D. A., Taylor A. W., Paterson D. H.: Comparison of the histochemical and contractile properties of human triceps surae, Journal of Appl. Physiology, No.58, 2000.
17. Schulz S., ua: Hydraulische Muskeln als Antriebseinhaeit für endoluminale Systeme in Medizin, Forschungszentrum Karlsruhe, FZKA 6080, 1998.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 157-163.

ZINTEGROWANE SYMULATORY CHODU W ZASTOSOWANIU DO SYNTEZY ALGORYTMÓW STEROWANIA

APPLICATION OF INTEGRATED SIMULATORS OF WALKING IN SYNTHESIS OF CONTROL ALGORITHMS

PIOTR KACZMAREK, ANDRZEJ KASIŃSKI

Instytut Automatyki i Inżynierii Informatycznej, Politechnika Poznańska, Poznań, ul. Piotrowo3a, e-mail: Piotr.Kaczmarek@put.poznan.pl

Streszczenie. Tworzenie efektywnych algorytmów sterowania układami biomechanicznymi wymaga uwzględnienia wielu specyficznych właściwości dynamiki i kinematyki tych układów. Artykuł opisuje metody i narzędzia wykorzystane do stworzenia modelu biomechanicznego kończyn dolnych dorosłego mężczyzny. Szczególną uwagę poświęcono wpływowi mięśni oraz naturalnych ograniczeń na dynamikę całego układu oraz niezbędnym uproszczeniom przyjętym w fazie modelowania. W drugiej części przedstawione zostało zastosowanie modelu jako symulatora dla algorytmów stabilizacji postawy oraz algorytmów sterowania chodem dwunożnych robotów kroczących. Dane z symulacji zestawione z danymi z eksperymentów fizjologicznych pozwalają określić przydatność tych algorytmów w zastosowaniu do sterowania ruchem w układzie biomechanicznym, jak również możliwości ich modyfikacji i rozwijania w celu wykorzystania w systemach funkcjonalnej elektrostymulacji (FES).

Słowa kluczowe: model biomechaniczny, symulacja, FES

Astract. In order to create effective and robust control algorithms of biomechanical systems motion, the specific properties of the dynamics as well as kinematics have to be taken into account. This paper describes methods and tools which support biomechanical modelling. These software tools have been used to develop a model of lower extremities of an adult male. Discussed are the features of the model i.e. an influence which muscles and natural movement limitations of knee and ankle joints have on its dynamics. The second part of the paper presents an application of the model to simulate and to test balance and walking control strategies. A comparison of the simulation results to experimental evidence enables the evaluation of usefulness of these strategies in application as controllers of the biomechanics system movement, in particular to a functional electrical stimulation (FES) task.

Key words: biomechanical model, simulation, FES

1. Wprowadzenie

Struktura i działanie układów biologicznych często staje się pierwowzorem dla rozwiązań technicznych. Takie podejście jest uzasadnione przede wszystkim faktem, że układy biologiczne ewoluują tworząc pewne optymalne rozwiązania wielu różnorodnych problemów od lokomocji po metody komunikacji, widzenie, rozpoznawanie, kojarzenie itp. W celu przeniesienia tych rozwiązań na grunt techniczny niezbędne jest poznanie i możliwie dokładne opisanie działania układów biologicznych. Wiąże się to z koniecznością stworzenia modelu, który w sposób spójny reprezentowałyby określoną funkcję układu biologicznego. Analiza działania takiego modelu pozwala na wyróżnienie pewnych cech i mechanizmów istotnych z punktu widzenia realizacji określonego zadania. Takie podejście ma miejsce przy budowie robotów kroczących. Większość tych konstrukcji (dwu i wielonożnych) posiada strukturę kinematyczną nóg, która ma swój pierwowzór w świecie organizmów żywych. Również część algorytmów sterowania, sekwencje ruchu nóg robotów wielonożnych zostały stworzone na podstawie obserwacji przebiegu lokomocji u zwierząt. Modelowanie prócz celu aplikacyjnego ma przede wszystkim cel poznawczy. Tworzenie dokładnych, zweryfikowanych doświadczalnie modeli pozwala nie tylko odtwarzać zachowania rzeczywistych obiektów, lecz również tłumaczyć i pogłębiać wiedzę o fizjologii i naturze tych zjawisk. Badania, w których wykorzystuje się takie, modele dążą w dwóch kierunkach. Pierwszy z nich koncentruje się na wyjaśnianiu naturalnych mechanizmów sterowania ruchem, synchronizacji pobudzeń mięśni w trakcie realizacji określonego ruchu. Modele tego typu, jak np. hipoteza punktu równowagi (model λ) [5], opisują zjawiska fizjologiczne (takie jak skurcz mięśnia, odpowiedzi receptorów prioproceptywnych na zmianę długości mięśnia) wyłącznie na poziomie behawioralnym, redukując np. opis przebiegu siły skurczu mięśnia do pewnego parametru nazywanego aktywnością, zaś sygnał pobudzenia z centralnego układu nerwowego do dwuelementowego wektora sterowań [6]. Mimo tych uproszczeń wyniki symulacji modelu λ wykazują dużą

zbieżność z danymi eksperymentalnymi w zakresie obserwowanej aktywności mięśnia oraz synchronizacji grup mięśni podczas ruchu w obrębie wielu stawów. Drugą grupę modeli stanowią realistyczne modele mięśni oraz mniejszych struktur biomechanicznych. Odpowiedniość zachowania rzeczywistego układu i modelu sięga znacznie głębiej niż to miało miejsce w modelach behawioralnych. Uwzględnienie procesów biochemicznych i dokładne odwzorowanie struktury mikroskopowej umożliwia obserwację pewnych efektów fizjologicznych [10] czy też patologicznych, tj. zmiany dynamiki skurczu mięśnia przy zmianie pobudzenia, występowanie procesu rekrutacji (gdy model opisuje mięsień jako układ złożony z wielu jednostek ruchowych), skutków odnerwienia czy też zmian występujących podczas uszkodzenia wybranych dróg przewodnictwa nerwowego. Modele takie z uwagi na złożoność obliczeniową nie są wykorzystywane jako elementy modeli biomechanicznych obejmujących wiele stawów i grup mięśniowych. Trzecia grupa zastosowań dotyczy modeli biomechanicznych tworzonych dla celów szeroko pojętej rehabilitacji [11]. Badania prowadzone w tym zakresie koncentrują się przede wszystkim na zadaniu poszukiwania metod zmniejszania dysfunkcji ruchowej u pacjentów z upośledzeniem funkcji motorycznych. Niniejszy artykuł opisuje model, który można zaklasyfikować do trzeciej grupy przedstawionego podziału.

Model ten będzie służył do testowania i rozwijania algorytmów sterowania, które będą mogły być następnie zastosowane do sterowania układami rzeczywistymi. Planuje się, że ruch takiego układu może być wywoływany bądź przez egzoszkielet, bądź też za pomocą stymulacji elektrycznej mięśni i nerwów obwodowych. Funkcjonalna elektrostymulacja (FES), której celem jest przywrócenie wybranej funkcji ruchowej, wymaga uwzględnienia w modelu dodatkowych czynników, tj. ograniczeń związanych z cechami skurczu mięśnia stymulowanego oraz zubożonej informacji na temat stanu układu. Analiza wyników symulacji z punktu widzenia spełniania wymienionych ograniczeń pozwoli na wstępną ocenę stosowalności danego algorytmu dla sterowania ruchem w układzie biologicznym.

W dalszej części zostaną omówione narzędzia wykorzystywane do tworzenia modeli, następnie model biomechaniczny oraz dwa wybrane algorytmy sterowania ruchem, dedykowane pierwotnie do zastosowań dla robotów kroczących. Następnie wyniki symulacji tych algorytmów zostaną przeanalizowane w aspekcie możliwości ich aplikacji do sterowania układami biomechanicznymi.

2. Narzędzia modelowania

Z punktu widzenia modelowania układów biomechanicznych środowisko dedykowane dla tworzenia takich obiektów powinno umożliwiać:

- budowę łańcucha kinematycznego,
- możliwość nałożenia więzów na zakres ruchu poszczególnych par kinematycznych oraz możliwość określenia położenia węzła takiej pary jako funkcji kąta,
- uwzględnienie dynamiki poszczególnych ogniw tego łańcucha,
- uwzględnienie pasywnego (w trakcie rozciągania) oraz aktywnego (skurcz) oddziaływania mięśni,
- uwzględnienie interakcji z otoczeniem (tj.: oddziaływania pomiędzy obiektem a podłożem).

Tab. 1. Zestawienie cech narzędzi wspomagających tworzenia modeli biomechanicznych.

	SIMM	MATLAB	ODE
Interface	Okienkowy +język skryptowy	Simulink +język skryptowy	Biblioteka dla C, C++, Pascal
Model kinematyczny	Tak	Tak**	Tak
Model dynamiki	Tak*	Tak**	Tak
Mięśnie	Tak	Nie	Nie
Interakcje ze środowiskiem	Nie	Nie	Tak
Wizualizacja	Tak	Tak	Nie
Szybkość	Średnia	Niska	Duża
Cena	40000$	Licencja edukacyjna	Licencja GNU

* funkcja dostępna z wykorzystaniem pakietu dynamics pipelining

** funkcja dostępna przy wykorzystaniu SIMULINK oraz SimMechanics Toolbox

Wśród dostępnych narzędzi do tworzenia modeli biomechanicznych niewątpliwie dominującą rolę odgrywa pakiet SIMM, Musculographics Motion Analysis Corp. Umożliwia on interaktywne tworzenie modeli układów biomechanicznych z uwzględnieniem pasywnych, jak i aktywnych właściwości mięśni oraz cech anatomicznych układu, tj. zmiany orientacji i położenia osi obrotu w stawie wraz ze zmianą kąta. Ponadto uwzględnia zmianę punktu zetknięcia ścięgna z kością w trakcie ruchu, co ma istotny wpływ na wartość ramienia siły podczas skurczu mięśnia.

Środowisko takie jak, MATLAB, Mathworks Inc. nie posiada narzędzi dla pełnego wspomagania procesu tworzenia modeli biomechanicznych. Dzięki zastosowaniu SimMechanics Toolbox możliwe jest tworzenie łańcuchów kinematycznych, symulacja ich dynamiki oraz wizualizacja. W celu uwzględnienia oddziaływania mięśni konieczne jest stworzenie ich modelu.

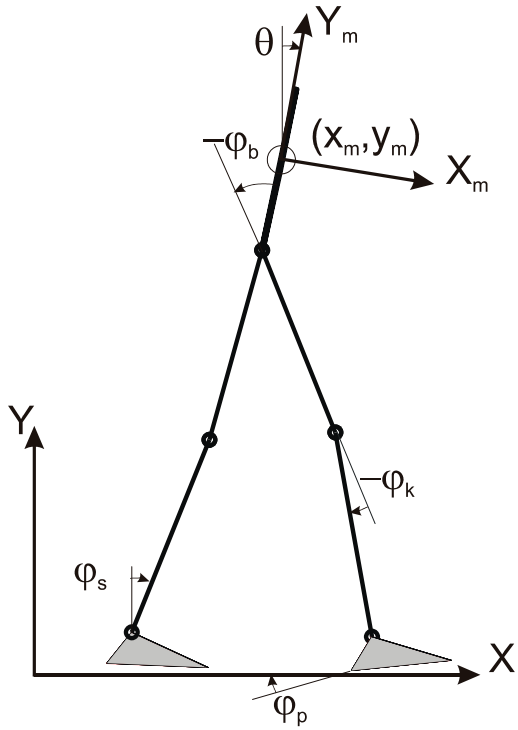
Biblioteka ODE (Open Dynamics Engine) zawiera funkcje pozwalające na tworzenie łańcuchów kinematycznych oraz zdefiniowanie ich cech dynamicznych. Ponadto możliwe jest tworzenie złożonego modelu otoczenia. Biblioteka wspomaga wykrywanie kolizji między obiektami tego otoczenia, definiowanie interakcji (tj. tarcie, lepkość itp). Dzięki temu możliwa jest symulacja obiektów przemieszczających się. Ta właściwość wyróżnia ODE spośród poprzednio wymienionych narzędzi, które wymagają określenia nieruchomej bazy łańcucha kinematycznego, a symulacja chodu wymaga dodatkowych zabiegów pozwalających na obejście tego ograniczenia. W przedmiotowej pracy do budowy i symulacji modelu wykorzystano pakiet MATLAB oraz bibliotekę ODE. Pakiet SIMM pomimo niewątpliwych zalet nie był wykorzystywany z uwagi na cenę licencji.

3. Model

3.1. Model kinematyki i dynamiki

Model złożony jest z siedmiu ogniw reprezentujących kości (stopy, podudzia, uda, tułowia) oraz 6 węzłów reprezentujących stawy (skokowy, kolanowy, biodrowy). Wartości parametrów dynamicznych są średnimi wartościami wyznaczonymi dla dorosłego mężczyzny [11]. Parametry geome-

tryczne zostały określone na podstawie modelu człowieka stworzonego w SIMM Musculographics Corp. 2002. (tab. 2, rys. 1)



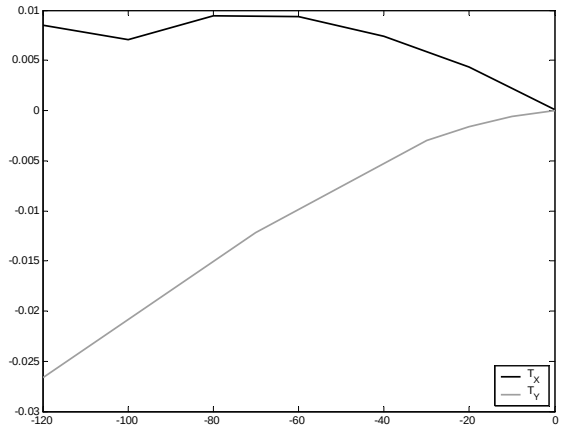
Rys. 1. Schemat łańcucha kinematycznego modelu. φ_s, φ_k, φ_b są kątami mierzonymi odpowiednio w stawie skokowym, kolanowym i biodrowym, φ_p jest kątem między stopą a podłożem, θ jest pochyleniem tułowia w globalnym układzie współrzędnych, zaś (x_m, y_m) jest położeniem środka masy tułowia w globalnym układzie współrzędnych

Tab. 2. Wartości parametrów dynamicznych modelu

	Wymiary (szacunkowe) [m]	Masa [kg]	momenty bezwładności [I_x; I_y; I_z] [kg*m^2]
stopa	^XL_s= 0.175 ^YL_s= 0.062	1.10	[0.002;0.008;0.009]
podudzie	L_p=0.435	3.75	[0.019;0.065;0.065]
udo	L_u=0.410	7.58	[0.080;0.126;0.126]
tułów	L_t=0.343	51.22	[0.764;3.407; 3.297]

Wymiary poszczególnych ogniw są orientacyjne z uwagi na zmianę odległości miedzy środkami stawów podczas ruchu.

W układzie rzeczywistym ruch w obrębie danego stawu odbywa się wokół pewnej osi obrotu, jednak jej położenie ulega zmianie wraz z kątem. W modelu uwzględniono ten fakt w odniesieniu do stawu kolanowego (rys. 2), zakładając że zmienia się wyłącznie położenie osi obrotu, zaś orientacja pozostaje stała. Właściwość ta jest implementowana w modelu poprzez opisanie stawu jako pary kinematycznej 3 klasy (obrotowo-przesuwnej), gdzie wartość przesunięcia jest funkcją kąta obrotu.



Rys. 2. Zmiana położenia osi obrotu w stawie kolanowym w funkcji kąta. T_x T_y – translacja osi obrotu w kierunku x i y

3.2. Mięśnie i aktywacja układu

Przyjęto, że w tworzonym modelu mięśnie będą traktowane wyłącznie jako elementy pasywne. Każdy mięsień opisywany jest modelem trójelementowym Hilla [3], [7], [11] (rys. 3). Elementy tego modelu odpowiadają poszczególnym częściom mięśnia. Równoległe połączenie tłumika *c* oraz nieliniowego elementu sprężystego k_2 reprezentuje brzusiec mięśnia, zaś elementy sprężyste k_1 odpowiadają ścięgnom.

Współczynnik sprężystości k_2 jest wyznaczany z zależności empirycznej [7]:

$$k_2 = 1.97 \cdot (\frac{L}{L_0} - 0{,}8)^3 * \frac{P_0}{L}, \quad (1)$$

gdzie: P jest siłą pasywną generowaną przez mięsień o długości L, zaś L_0 jest optymalną długością mięśnia, a P_0 jest siłą pasywną mierzoną przy tej długości.

Lokalizacja przyczepu danego mięśnia została określona na podstawie danych anatomicznych [5] oraz modelu stworzonego w SIMM. Umożliwia to wyznaczenie aktualnej długości mięśnia, a tym samym siły pasywnej generowanej przez mięsień podczas ruchu w stawie.

Żadne z wykorzystanych narzędzi modelowania nie umożliwia bezpośredniego włączenia w łańcuch kinematyczny elementów nie będących bryłami sztywnymi lub węzłami. Oddziaływanie na układ możliwe jest wyłącznie za pomocą momentów siły. Konieczne jest więc przeliczenie siły oddziaływania danego mięśnia na moment siły:

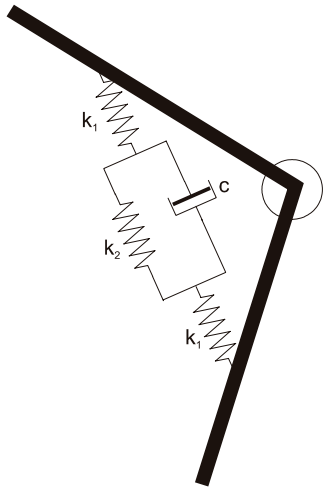
$$\tau = \vec{P} \times \vec{r}, \quad (2)$$

gdzie: P jest siłą reakcji mięśnia, zaś r ramieniem siły zależnym od kąta zgięcia w stawie.

W przedmiotowym modelu przyjęto, że ruch aktywowany jest zewnętrznym momentem działającym w każdym z węzłów (stawów). Wartości momentów wyznaczane są na podstawie wybranego, aktualnie testowanego algorytmu sterowania układem. Tym samym pomija się właściwości skurczu mięśnia (tj. czas narastania siły, jej zmienność w czasie). Ten aspekt zostanie szerzej omówiony w Dyskusji.

3.3. Naturalne ograniczenia

Każdy układ biomechaniczny posiada ograniczony zakres ruchomości w obrębie danego stawu. Część z tych ograniczeń jest efektem pasywnego oddziaływania mięśni. Dzięki uwzględnieniu nieliniowej wartości współczynnika sprężystości brzuśca (k_2) model uwzględnia to ograniczenie.



Rys. 3. Model mięśnia. k_1 – współczynnik sprężystości ścięgien, k_2, c – współczynnik sprężystości i lepkości modelu brzuśca mięśnia

Drugą grupę stanowią ograniczenia wynikające z ukształtowania anatomicznego kości i stawów. W modelu uwzględniono dwa z nich:

- ograniczenie ruchomości stawu kolanowego, które jest typu więzu jednostronnego $\varphi_k \leq 0$
- ograniczenie ruchomości w stawie skokowym: $\varphi_s \leq 10^0$

Więzy te zostały uwzględnione w modelu przez wprowadzenie dodatkowych elementów lepko-sprężystych, które generują moment reakcji:

$$\tau_r = \begin{cases} 0 : \varphi < \varphi_{\max} \\ -k(\varphi - \varphi_{\max}) - c \cdot \dot{\varphi} : \varphi \geq \varphi_{\max} \end{cases}, \tag{3}$$

gdzie: φ jest kątem mierzonym w stawie, $\varphi_{\max}$ jest zakresem ruchomości, zaś k i c współczynnikami sprężystości i lepkości ograniczenia.

3.4. Siły reakcji podłoża

Chwila kontaktu z podłożem jest wykrywana, gdy położenie pięty lub palców osiągnie poziom gruntu: $y=y_p$, gdzie y_p jest poziomem gruntu. Siła reakcji gruntu jest wyznaczana z zależności [4]:

$$F_y = -E_y \cdot \Delta y^3 - D_y \cdot \dot{y} \cdot \Delta y^2 \tag{4}$$

gdzie: F_y jest siłą reakcji w kierunku pionowym, E oraz D są modułami sprężystości i lepkości gruntu, zaś $\Delta y = y_p - y$

Dodatkowo w celu uniknięcia sklejania się stopy z podłożem zakłada się $F_y>0$.

W osi poziomej zakłada się występowanie lepkości. Siła pozioma opisana jest wyrażeniem:

$$F_x = D_x \cdot \dot{x} \tag{5}$$

4. Algorytmy sterowania

Do testów zostały wybrane dwa algorytmy sterowania ruchem. Każdy z nich został pierwotnie stworzony dla zastosowań sterowania robotami kroczącymi. Celem testów jest określenie przydatności tych algorytmów w zadaniu sterowania ruchem układu biomechanicznego, który charakteryzuje się ograniczoną możliwością pozyskiwania informacji sensorycznej oraz wysoko zlokalizowanym środkiem masy układu. Poniżej przedstawiono zwięzłą charakterystykę algorytmów, z których pierwszy zastosowany zostanie dla zadania stabilizacji postawy, zaś drugi do sterowania chodem.

4.1. Virtual Model Control

Celem sterowania jest stabilizacja położenia środka masy układu. Prawo sterowania polega na określeniu siły wirtualnej [8,9], która powstałaby, gdyby między położeniem zadanym a położeniem aktualnym umieścić równoległy element lepko-sprężysty. Siła ta opisana jest wyrażeniem:

$$F = \begin{bmatrix} F_x \\ F_y \end{bmatrix} = \begin{bmatrix} k_x \\ k_y \end{bmatrix} \cdot e + \begin{bmatrix} c_x \\ c_y \end{bmatrix} \cdot \dot{e} + \begin{bmatrix} 0 \\ Q \end{bmatrix} \tag{6}$$

gdzie: e jest uchybem regulacji $e = \begin{bmatrix} x_0 - x \\ y_0 - y \end{bmatrix}$ zaś Q jest ciężarem całego obiektu, parametry wirtualnych elementów sprężystego i lepkiego (k i c) wyznaczane są empirycznie.

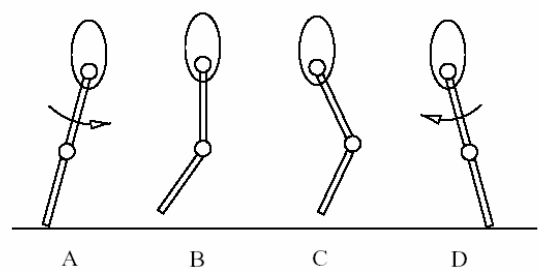
Moment w każdym ze stawów wyznaczany jest na podstawie transformacji:

$$\tau = \begin{bmatrix} \tau_s \\ \tau_k \\ \tau_b \end{bmatrix} = J^T F \tag{7}$$

gdzie: J jest jakobianem układu transformującym układ globalny XY do układu lokalnego X_mY_m związanego ze środkiem masy tułowia (rys. 1).

Ponieważ algorytm jest stosowany dla zadania stabilizacji postawy, zakłada się, że nogi są złączone i na każdą z nich przypada jednakowe obciążenie, stąd momenty w stawach nogi lewej i prawej są równe i wynoszą:

$$\tau_{l/p} = \tfrac{1}{2}\tau \tag{8}$$



Rys. 4. Fazy chodu z wykorzystaniem dynamiki własnej. Źródło: [8]

4.2. Sterowanie z wykorzystaniem dynamiki własnej

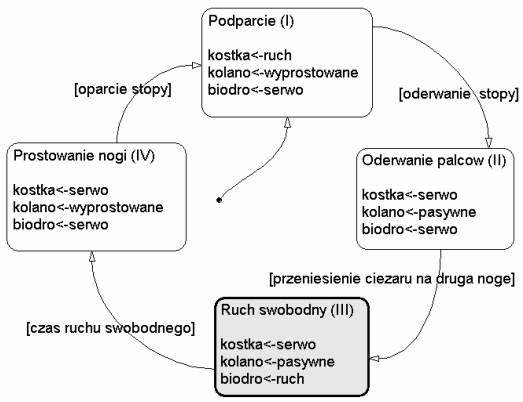
W algorytmie tym wykorzystuje się właściwości dynamiczne oraz istnienie naturalnych ograniczeń wynikających z budowy anatomicznej układu [8].

Ruch każdej nogi w czasie pojedynczego kroku podzielony jest na 4 fazy (rys. 4). Cykl sterowania jest reprezentowany przez maszynę stanów (rys. 5), odrębną dla każdej z nóg. Każdemu ze stanów przypisany jest określony tryb sterowania danym stawem. Pętle regulacyjne dla każdego ze stawów są niezależne i mogą pracować w następujących trybach:

- pasywny – w stawie nie działa żaden moment,
- ruch – w stawie działa stały moment o niezerowej wartości,
- serwo – pętla realizuje regulację PD, gdzie wartością zadaną jest kąt.

W fazie podparcia w stawie skokowym generowany jest stały, dodatni moment, który wymusza przemieszczanie środka masy w przód. Istnienie ograniczenia ruchomości w stawie skokowym prowadzi do automatycznego uniesienia pięty, gdy $\varphi_s \geq \varphi_{\max}$.

W fazie ruchu swobodnego dla stawu biodrowego realizowany jest początkowo tryb ruch, a po upływie ustalonego czasu tryb serwo (przez zmianę wartości kąta zadanego możliwe jest modyfikowanie długości kroku), jednocześnie staw kolanowy pozostaje pasywny. Przy gwałtownym ruchu w stawie biodrowym – zgodnie z zasadą zachowania momentu pędu – następuje ugięcie nogi w stawie kolanowym, a po zatrzymaniu ruchu w stawie biodrowym, dochodzi do całkowitego wyprostu nogi w stawie kolanowym. Dzięki istnieniu ograniczenia zakresu ruchu w tym stawie kończyna powinna pozostać wyprostowana bez konieczności generowania zewnętrznego momentu siły.

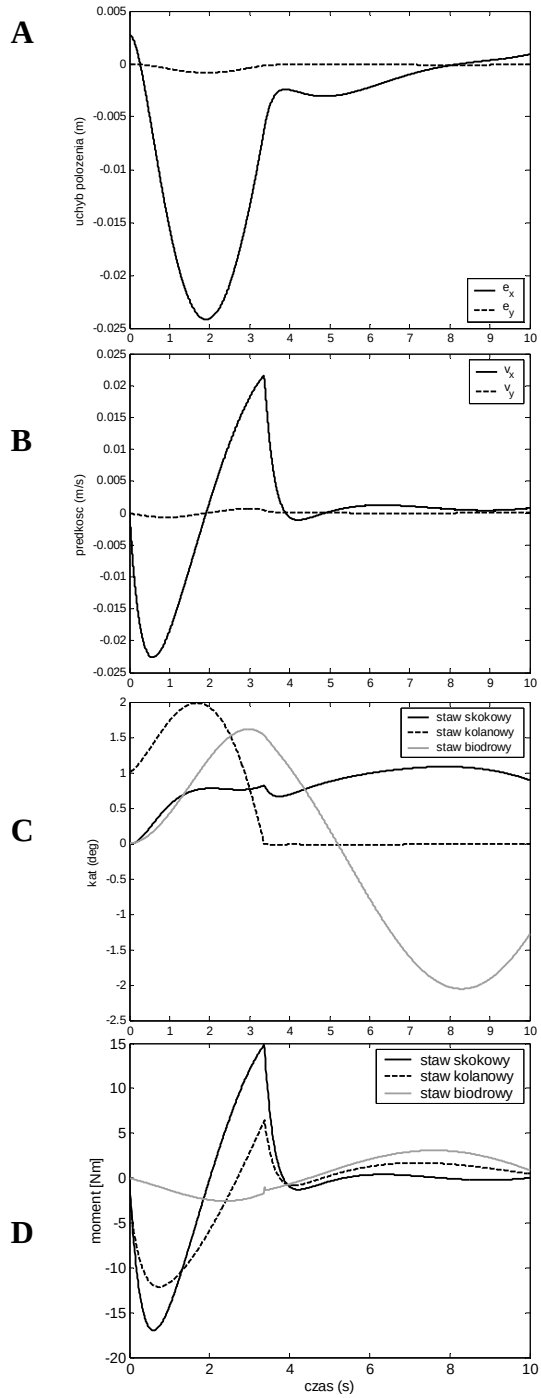


Rys. 5. Maszyna stanów zaimplementowana w MATLAB StateFlow Toolbox dla algorytmu sterowania wykorzystującego dynamikę własną układu

5. Wyniki

5.1. Stabilizacja postawy

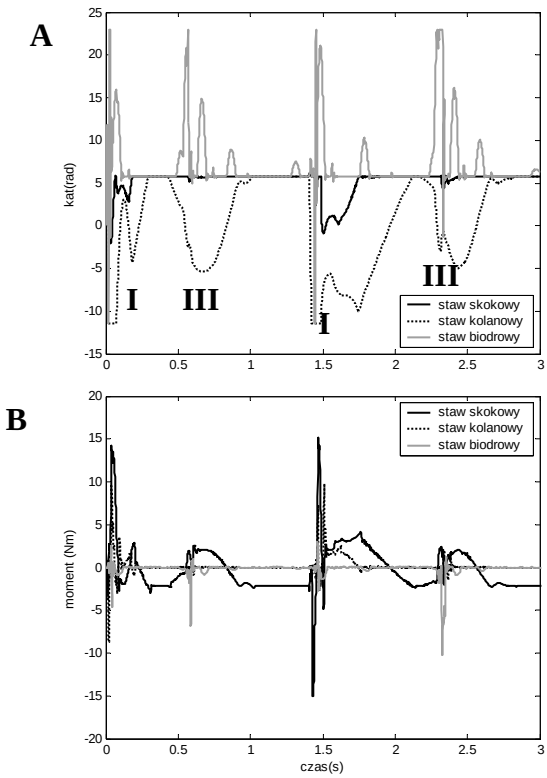
Symulację przeprowadzono w środowisku MATLAB. Ruch układu odbywał się w płaszczyźnie XY (2D). Na rys. 6 przedstawiono efekt działania algorytmu wykorzystującego siły wirtualne w celu stabilizacji pozycji ciała. Układ w chwili początkowej ma niezerowe warunki początkowe (kąt w stawie kolanowym jest niezerowy). Po około 3 s symulacji obserwuje się zmianę charakterystyki uchybu oraz jego pochodnej polegającą na gwałtownym stłumieniu oscylacji położenia środka masy. Ma to związek z istnieniem ograniczenia ruchomości w stawie kolanowym. Po tym czasie kąt w stawie kolanowym osiąga wartość graniczną, co oznacza, że noga znajduje się w maksymalnym wyproście, a równowaga utrzymywana jest poprzez ruch w stawie skokowym i biodrowym. Pomimo zablokowania stawu kolanowego działa tam pewien niezerowy moment siły. Dla symulacji układu z innymi warunkami początkowymi układ był stabilny pod warunkiem, że w chwili początkowej rzut środka masy znajdował się w obrębie stóp.



Rys. 6. Wyniki symulacji stabilizacji postawy. A – Przebieg uchybu położenia środka masy, B – pochodna uchybu, C – wartości kątów w stawach, D – wartości momentów w stawach

5.2. Sterowanie chodem

Symulację przeprowadzono przy użyciu biblioteki ODE. Ruch układu odbywał się w płaszczyźnie XYZ (3D). Na rys. 7 przedstawiono zmiany wartości kątów i momentów w stawach lewej nogi podczas dwóch kroków. W fazie ruchu swobodnego nogi (faza III) obserwuje się znaczną, długotrwałą zmianę kąta w stawie kolanowym, która jest efektem gwałtownego ruchu w stawie biodrowym, nie zaś wynikiem działania zewnętrznego momentu siły w tym stawie. W stawie kolanowym niezerowy moment jest generowany wyłącznie w fazie podparcia (I), a największą wartość osiąga na początku tej fazy, kiedy to wskutek kontaktu stopy z podłożem powstaje gwałtowne obciążenie dynamiczne i dochodzi do zgięcia w stawie kolanowym.

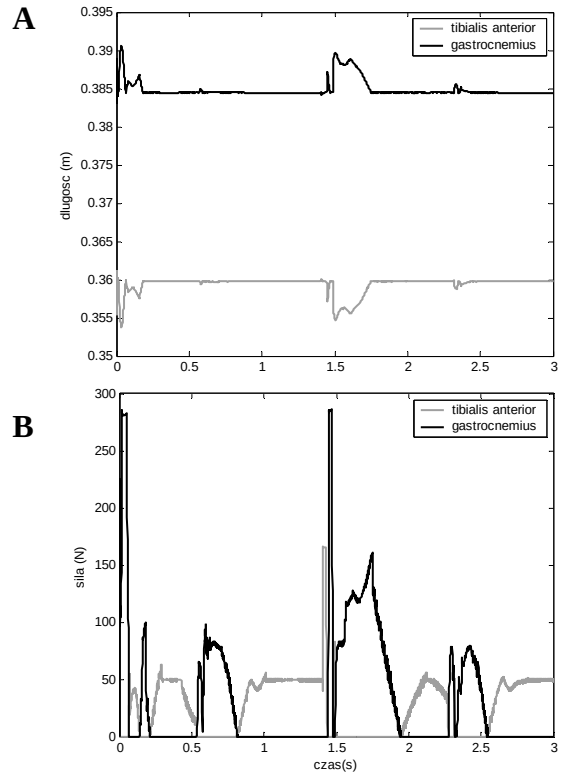


Rys. 7. Przebieg wartości kątów (A) i momentów (B) w lewej nodze podczas dwóch kroków chodu. I – faza podparcia, III – faza ruchu swobodnego

Moment siły generowany w stawie biodrowym nogi podtrzymującej ciężar ciała podczas swobodnego ruchu drugiej nogi przeciwdziała nadmiernemu odchyleniu tułowia. Na szczególną uwagę zasługuje przebieg wartości momentów w stawie skokowym. W fazie podparcia działanie dodatniego momentu powinno wymuszać przemieszczanie środka masy całego układu w przód. W tej fazie w stawie skokowym realizowany jest algorytm PD sterowania. Gdy wartość kąta w stawie skokowym zbliża się do wartości zadanej, następuje zmiana znaku momentu, co prowadzi do zmniejszenia prędkości przesuwania się środka masy, a tym samym chroni go przed wyjściem poza obręb stopy, na której jest wsparty. Jest to pewna modyfikacja względem oryginalnego algorytmu, która musiała zostać wprowadzona dla zachowania stabilnego chodu.

Działanie dodatniego momentu w fazie ruchu swobodnego nogi umożliwia zorientowanie stopy w taki sposób, by nie doszło do zahaczenia palców o podłoże.

Dla stawu skokowego wyznaczono zmiany długości mięśni (rys. 8): mięśnia odpowiedzialnego za zgięcie podeszwowe (*gastrocnemius*) oraz mięśnia antagonisty odpowiedzialnego za zgięcie grzbietowe (*tibialis anterior*). Dodatkowo na podstawie znanych momentów działających w stawie oraz znanego ramienia siły obliczono przebieg siły skurczu każdego z tych mięśni w trakcie ruchu.



Rys. 8. Zmiany długości (A) oraz siły skurczu (B) mięśni *tibialis anterior* i *gastrocnemius* podczas dwóch kroków chodu

6. Dyskusja

6.1. Ograniczenia modelu

Prezentowany model zawiera uproszczenia wprowadzone w fazie modelowania. Do najważniejszych zalicza się:

- uwzględnienie wyłącznie pasywnego modelu mięśnia,
- pominięcie wpływu innych tkanek miękkich na dynamikę układu,
- pominięcie wpływu kończyn górnych, które w istotny sposób oddziałują na zachowanie równowagi podczas chodu.

6.2. Ocena algorytmów sterowania

Model biomechaniczny może być wykorzystany w celu oceny algorytmów sterowania ruchem z punktu widzenia zastosowania do kontroli ruchu w układach biologicznych.

Algorytm sterowania oparty o *Virtual Model Control* jest odporny na pojawiające się w układzie zakłócenia. Niestety, nie uwzględnia on naturalnych ograniczeń ruchomości

w stawie kolanowym oraz właściwości dynamicznych układu, co prowadzi do generowania momentów o dużych wartościach. I tak pomimo maksymalnego wyprostu w stawie kolanowym (co powoduje jego zablokowanie) niepotrzebnie generowany jest moment. Algorytm dla poprawnego działania wymaga wyznaczenie jakobianu układu, co wiąże się z dokładnym wyznaczeniem parametrów geometrycznych sterowanego układu oraz umieszczeniem sensorów dostarczających informacji o kącie w każdym stawie. Z uwagi na konieczność dostarczenia pełnej informacji sensorycznej, algorytm ten może się okazać trudny do implementacji w systemie FES. Dodatkowo konieczność ciągłego utrzymywania momentów siły o dużych wartościach może powodować szybkie męczenie się stymulowanych mięśni, a tym samym niemożliwość utrzymania postawy ciała. Wydaje się, że algorytm ten może być przydatny do sterowania ruchem przy użyciu egzoszkieletu. Egzoszkielet nie posiada cech męczliwości, które charakteryzowały układ biologiczny, umożliwia również pozyskanie informacji sensorycznej niezbędnej do określenia jakobianu układu. Niewątpliwą wadą takiego rozwiązania są gabaryty układu oraz zapotrzebowanie na moc znacznie większe niż w przypadku stymulacji mięśni lub nerwów.

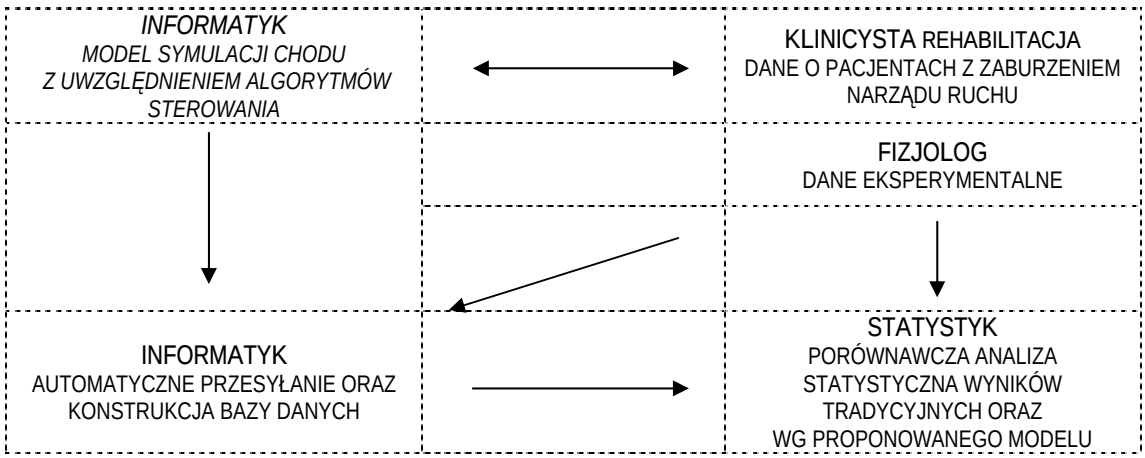
Algorytm uwzględniający dynamikę własną układu wymaga znacznie uboższej informacji sensorycznej. Z punktu widzenia algorytmu istotna jest znajomość kąta w stawie biodrowym (w fazie ruchu swobodnego), wartość kąta w stawie skokowym oraz określenie, czy kolano jest wyprostowane, czy też nie. Istnienie faz ruchu, podczas których w stawie nie działa żaden moment, sprzyja zastosowaniu tego algorytmu jako algorytmu sterowania systemem FES. Wydaje się również, że układ powinien zachowywać się stabilnie, gdy informacja sensoryczna będzie zbiorem wartości rozmytych, hipoteza ta jednak wymaga weryfikacji. Pewnym problemem w zastosowaniu tego algorytmu w systemie FES jest występowanie gwałtownej zmienności momentów, szczególnie w obrębie stawu biodrowego. Siła skurczu mięśnia w trakcie stymulacji elektrycznej wykazuje skończony czas narastania oraz relaksacji [1], [2], stąd niemożliwe jest uzyskanie w układzie biologicznym tak gwałtownego przyrostu siły skurczu. Celowe byłoby przetestowanie działania tego algorytmu przy nałożonym ograniczeniu szybkości narastania siły.

Reasumując, został stworzony model biomechaniczny pozwalający testować różne algorytmy sterowania oraz wstępnie oceniać ich przydatność dla zastosowań sterowania ruchem w układzie biologicznym.

Praca częściowo finansowana z środków KBN; projekt nr 1445/T11/2004/27.

Literatura cytowana

1. Bichler E., Celichowski J.: Changes in the properties of mechanomyographic signals and in the tension during the fatigue test of rat medial gastrocnemius muscle motor units, J. Electromyogr. Kinesiol., 11: 387-394, 2001.
2. Celichowski J., Bichler E.: The time course of the last contractions during incompletely fused tetani of motor units in rat skeletal muscle, Acta Neurobiol. Exp., 62: 7-17, 2002.
3. Enoka R. M.: Neuromechanics of human movement, Third edition, Human Kinetics, United States, 2002.
4. Gerritsen K., van der Bogert A., Nigg B.: Direct Dynamic simulation of the impact phase in heel-toe running, J. Biomech. 28(6): 661-668, 1995.
5. Hog M., Zajac F., Gordon M.: A muscoskeletal model of the human lower extremity: the effect of muscle tendon, and moment arm on the moment-angle relationship of musculotendon actuators of the hip, knee and ankle, J. Biomech., 23(2): 157-169, 1990.
6. Micheau P., Kron A., Bourassa P.: Evaluation of the lambda model for human postural control during ankle strategy, Biol. Cyber., 89: 227-236, 2003.
7. Morecki A., Ekiel J., Fidelus K.: Bionika ruchu, Podstawy zewnętrznego sterowania biomechanizmów i kończyn ludzkich. PWN, Warszawa 1971.
8. Pratt E. J.: Exploating Inherent Roboustness and Natural Dynamics in the Control of Bipedal Walking Robot, PhD Thesis, Massachusetts Institute of Technology, 2000.
9. Pratt J. Dilworth P. Pratt G.: Virtual Model Control of a Bipedal Walking Robot, IEEE International Conferenceon Robotics and Automation, pp.193-198, 1997.
10. Wexler A. S., Ding J., Binder-Macleod S. A.: A mathematical model that predicts skeletal muscle force, IEEE Trans. Biomed. Eng., 44: 337-348, 1997.
11. Yamaguci G. T., Zajac F. E.: Restoring unassisted natural gait to paraplegics via functional neuromuscular stimulation. A computer simulation, IEEE Trans. Biomed. Eng., 37, 1990.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 165-170.

WSPOMAGANY KOMPUTEROWO WIELOŚCIEŻKOWY SYSTEM EGZAMINOWANIA STUDENTÓW MEDYCYNY

A MULTI-PATH COMPUTER-BASED EXAMINATION SYSTEM FOR STUDENTS OF MEDICINE

ANDRZEJ KONONOWICZ[*], MARTA ŻABIŃSKA-POPIELA[**],
IRENA ROTERMAN-KONIECZNA[*], JÓZEF KRZYSIEK[**]

[*] *Zakład Bioinformatyki i Telemedycyny, Collegium Medicum, Uniwersytet Jagielloński, Kraków*
[**] *Klinika Endokrynologii Ginekologicznej, Katedra Ginekologii i Położnictwa, Collegium Medicum, Uniwersytet Jagielloński, Kraków*

Streszczenie. Wadą wielu współczesnych metod egzaminacyjnych w medycynie jest wyrywkowe sprawdzanie wyuczonej wiedzy książkowej. Sposób ten odbiega od praktyki lekarskiej, w której lekarz musi przede wszystkim umieć właściwie postępować z pacjentem, a nie odpowiadać na szczegółowe pytania teoretyczne. W opracowanym przez nas komputerowym systemie egzaminowania student konfrontowany jest z pewnym przypadkiem medycznym (pacjentem indywidualnym), a jego zadaniem jest samodzielny wybór właściwej metody leczenia pacjenta. Podejmowane decyzje wpływają na to, jakie pytania będą stawiane podczas dalszego przebiegu testu. Niewłaściwe odpowiedzi mogą spowodować dalekie odejście osoby egzaminowanej od właściwej ścieżki postępowania w leczeniu pacjenta.
Opis egzaminów w realizowanym przez nas systemie jest niezależny od aplikacji służącej do ich przeprowadzania. Opracowane modele testów mają strukturę grafów skierowanych. Opisy przechowywane są w formacie XML, co sprzyja prostemu tworzeniu nowych egzaminów i łatwej ich wymianie. Utworzony został przykładowy wielościeżkowy test sprawdzający wiedzę z dziedziny położnictwa.
Omawiany przez nas wspomagany komputerowo system egzaminowania może stać się cennym uzupełnieniem klasycznych metod egzaminacyjnych przeznaczonych dla studentów medycyny.

Słowa kluczowe: e-Learning, położnictwo, komputerowo wspomagane systemy egzaminowania

Abstract. Randomized check of detailed medical facts is a disadvantage of many present examination methods in medicine. In fact this approach diverges from the practical medicine, where a physician should know how to deal with patients and not answer detailed theoretical questions. Hence in our computer-based assessment system we present a medical case and the student's task is to choose the correct treatment steps. Student's decisions influence what questions will be asked during further examination process. Wrong answers may lead the student far away from the correct treatment path.
In our approach, examination descriptions are independent of the testing application. The examination models have the structure of a directed graph. Models are stored in XML format which enables us to create, extend, and exchange examinations easily. We have created an exemplary examination model for testing knowledge in midwifery.
Our computer-based examination system may become a useful extension of classical examinations for students of medicine.

Key words: e-Learning, midwifery, computer based assessment

1. Wprowadzenie

Istotnym elementem nauki jest możliwość weryfikacji zdobytej wiedzy. Pozwala to na obiektywną ocenę stopnia opanowania poznawanego materiału. Uczeń spostrzega w ten sposób braki informacji, które powinien uzupełnić. Nie bez powodu mówi się, że człowiek najlepiej uczy się na własnych błędach. Egzaminy pozwalają popełniać błędy bez ponoszenia konsekwencji, które związane byłyby z nimi w warunkach rzeczywistych. Szczególnie w praktyce medycznej, gdzie decyzja lekarska wpływa bezpośrednio na zdrowie lub życie pacjenta, bardzo istotne jest gruntowne sprawdzenie wiedzy studenta przed dopuszczeniem go do wykonywania zawodu.

Przeprowadzanie egzaminów związane jest zwykle z niemałym nakładem pracy. Przygotowanie pytań i sprawdzenie odpowiedzi kosztuje dużo czasu. Technologie e-Learning pozwalają odciążyć nauczyciela, szczególnie jeżeli chodzi o czynność sprawdzania odpowiedzi studenta. Osoba egzaminowana również zyskuje, gdyż poznaje wyniki sprawdzianu od razu po jego zakończeniu. W praktyce

komputerowe systemy egzaminacyjne z powodzeniem stosowane są podczas egzaminów językowych lub przy sprawdzaniu wiedzy teoretycznej kandydatów na kierowców.

Niezwykle cenną cechą przeprowadzanych komputerowo egzaminów, w porównaniu z klasycznymi drukowanymi testami (często jednak niewykorzystywaną), jest możliwość dynamicznego dostosowywania pytań w zależności od odpowiedzi udzielanych przez studenta. Wadą wielu współczesnych metod egzaminacyjnych w medycynie jest wyrywkowe sprawdzanie wyuczonej wiedzy książkowej. Sposób ten nie odpowiada praktyce lekarskiej, w której lekarz musi przede wszystkim umieć właściwie postępować z pacjentem, a nie odpowiadać na szczegółowe pytania teoretyczne. W opracowanym przez nas komputerowym systemie egzaminowania student konfrontowany jest z pewnym przypadkiem medycznym, a jego zadaniem jest samodzielny wybór stosownej metody leczenia pacjenta. Podejmowane decyzje wpływają na to, jakie pytania będą stawiane podczas dalszego przebiegu testu. Niewłaściwe odpowiedzi mogą spowodować dalekie odejście osoby egzaminowanej od prawidłowej ścieżki postępowania w leczeniu pacjenta.

W artykule przedstawiamy opracowany przez nas model egzaminu. Pokazujemy sposób reprezentacji opisanego modelu, stosując bardzo popularny obecnie standard XML. Przedstawimy również prototyp systemu egzaminacyjnego o nazwie BIT, służący do przeprowadzania egzaminów zgodnych z naszym modelem. Jako przykład egzaminu przedstawimy opracowany przez nas egzamin sprawdzający wiedzę z dziedziny położnictwa.

2. Model egzaminu

Opracowany model egzaminu, który tu przedstawiono, jest grafem skierowanym <V,E>, gdzie V to zbiór wierzchołków stanu, a E to zbiór krawędzi pytań. Każdy wierzchołek $v \in V$ stanowi uporządkowana trójka v = [iv,d,q], gdzie iv oznacza identyfikator stanu, d oznacza opis stanu egzaminu (kontekst zadawanego pytania), a q oznacza treść pytania. Każda krawędź $e \in E$, $e=[v_1,v_2]$, $v_1,v_2 \in V$ oznacza możliwą w stanie v_1 odpowiedź, której udzielenie spowoduje przejście do stanu v_2. Z każdą krawędzią $e \in E$ powiązana jest piątka ε=[ie,a,c,p,r], gdzie ie oznacza identyfikator odpowiedzi, a – tekst odpowiedzi, c – ocenę udzielonej odpowiedzi (w skali należącej do przedziału $c \in [0;1]$, $c \in R$. Wartość 1 jest odpowiedzią absolutnie poprawną, a 0 absolutnie błędną). Element p stanowi zbiór identyfikatorów optymalnych ścieżek, do których dana krawędź należy, r – tekst komentarza do udzielonej odpowiedzi. Każdy model wyróżniony ma jeden wierzchołek startowy (zawsze z iv=0) zawierający początkowe pytanie oraz jeden lub więcej wierzchołków końcowych (iv≥1000), z których nie wychodzi już żadna krawędź. Pomiędzy węzłem początkowym oraz przynajmniej jednym węzłem końcowym musi istnieć co najmniej jedna ścieżka optymalna zawierająca same poprawne odpowiedzi. Komentarz do odpowiedzi może – ale nie musi – zawierać informacje dotyczące poprawności wybranej przez użytkownika krawędzi.

3. Tworzenie modeli egzaminów

Podczas tworzenia testów należy pamiętać o konieczności ograniczania złożoności modelu egzaminu. Jeżeli każda możliwa odpowiedź byłaby źródłem nowej, niezależnej od innych ścieżki, liczba pytań w modelu wyrażona byłaby wzorem:

$$\overline{V} = \frac{1-q^{n+1}}{1-q}\ , \qquad (1)$$

gdzie: q to liczba możliwych odpowiedzi w jednym pytaniu, n to długość optymalnej ścieżki (wysokość drzewa). Liczba pytań w takim drzewie dla q=4 i n=10 przekracza milion.

Zmniejszenie złożoność modelu osiąga się poprzez łączenie ścieżek. Nie zawsze możliwe jest płynne połączenie dwóch ścieżek. Czasem trzeba w komentarzu do udzielonej odpowiedzi poinformować, iż wybrana przez użytkownika ścieżka na tyle odbiega od prawidłowego postępowania, iż będzie ona zaniechana, a użytkownik skierowany będzie przez system na właściwą drogę.

Jako szablony do tworzenia wielościeżkowych egzaminów mogą być wykorzystane standardy medyczne i ścieżki kliniczne. Na przykład w opracowanym przez nas systemie wykorzystano ścieżkę dotyczącą terapii hiperprolaktynemii, zaczerpniętą ze Standardów Endokrynologii [6].

4. Reprezentacja modelu egzaminu za pomocą XML

By móc w praktyce wykorzystać opisany powyżej sposób modelowania egzaminów, należy dobrać dla modelu odpowiednią reprezentację danych. Zdecydowaliśmy się zapisać model w formacie XML.

XML jest opracowanym przez konsorcjum W3C uniwersalnym sposobem wymiany danych. XML nie jest zwykłym językiem programistycznym, lecz sposobem definiowania własnych języków znacznikowych. W odróżnieniu od HTML, znaczniki XML mogą być dowolnie definiowane przez użytkowników; nie mają również *a priori* zdefiniowanego sposobu interpretacji. Niewątpliwą zaletą XML jest to, iż jest on czytelny, zarówno dla ludzi, jak i komputerów, pozwala oddzielić dane od sposobu ich prezentacji, jest przenośny platformowo i można go w prosty sposób rozszerzać. Do definicji struktury dokumentu XML służą mechanizmy DTD lub XML Schema. XML Schema jest sposobem nowocześniejszym i dającym większe możliwości, DTD jest jednak bardziej zwięzły i czytelny a zarazem wciąż bardzo popularny. Z tego względu w artykule zdecydowaliśmy się przedstawić zaproponowaną przez nas strukturę modelu egzaminu w standardzie DTD. Tab. 1 przedstawia opis DTD struktury modelu egzaminu.

Wszystkie dane egzaminu zawarte są w znaczniku examination. Bezpośrednimi potomkami węzła examination są: examName – określający nazwę egzaminu, examDscr – zawierający skrótowy opis egzaminu, author – zawierający dane o autorze egzaminu oraz znacznik nodes, w którym zawarte są wszystkie wierzchołki $v \in V$ z modelu egzaminu reprezentowane przez znaczniki node. Każdy znacznik node zawiera znaczniki dscr i question, odpowiadające opisowi stanu d i tekstowi pytania q z modelu egzaminu. Identyfikator stanu zrealizowany jest jako atrybut id znacznika node. Znacznik node zawiera również znacznik edges zawierający krawędzie $e \in E$ wywodzące się z aktualnego wierzchołka v. Znacznik edge ma atrybut id oznaczający identyfikator danej odpowiedzi. Każdy znacznik edge musi zawierać znaczniki: answer, result, target, credit, oznaczające kolejno – tekst

odpowiedzi, tekst komentarza po udzieleniu danej odpowiedzi, identyfikator stanu, do którego przejdzie system po odpowiedzi na pytanie (atrybut node) oraz liczbę punktów przydzielanych za wybranie danej odpowiedzi (atrybut value). Opcjonalnym znacznikiem jest znacznik optimal, oznaczający przynależność do optymalnej ścieżki. Identyfikator ścieżki reprezentowany jest przez atrybut id. Jeżeli dana odpowiedź należy do wielu optymalnych ścieżek, znacznik edge może zawierać więcej niż jeden znacznik optimal.

Tab. 1. Opis DTD modelu egzaminu

```
<!ELEMENT examination examName,examDscr,author,nodes)>
<!ELEMENT examName (#PCDATA)>
<!ELEMENT examDscr (#PCDATA)>
<!ELEMENT author (#PCDATA)>
<!ELEMENT nodes (node*)>
<!ELEMENT node (dscr,question,edges)>
<!ATTLIST node id ID #REQUIRED >
<!ELEMENT dscr (#PCDATA)>
<!ELEMENT question (#PCDATA)>
<!ELEMENT edges (edge*)>
<!ELEMENT edge (answer,result,target,optimal*,credit)>
<!ATTLIST edge id ID #REQUIRED>
<!ELEMENT answer (#PCDATA)>
<!ELEMENT result (#PCDATA)>
<!ELEMENT target EMPTY>
<!ATTLIST target node CDATA #REQUIRED>
<!ELEMENT optimal EMPTY>
<!ATTLIST optimal id CDATA #REQUIRED>
<!ELEMENT credit EMPTY>
<!ATTLIST credit value CDATA #REQUIRED>
```

5. Implementacja systemu

Opracowane przez nas modele egzaminów dostępne są przez Internet jako aplikacja sieciowa. Do rozwiązania testu wymagana jest wyłącznie zainstalowana przeglądarka internetowa.

Aplikacja do przeprowadzania testów zrealizowana została przy użyciu technologii JSP oraz programowalnych arkuszy styli XSLT. JSP (Java Server Pages) jest dobrze znaną, działającą po stronie serwera technologią ułatwiającą dynamiczną generację stron internetowych.

XSLT (XSL Transformation) jest opracowaną przez W3C technologią transformacji dokumentów XML do postaci języka XHTML, innego dialektu XML bądź też innego formatu, takiego jak PDF, RTF lub zwykłego tekstu. Przekształcenia, jakim ma być poddane drzewo transformowanego pliku XML, opisywane są w arkuszach transformacji XSLT. Transformacja wykonywana jest przez procesor XSLT (np. Xalan, Saxon; procesory XSLT zawierają również popularne przeglądarki internetowe, takie jak Internet Explorer czy Firefox). Transformacja XSLT polega na dopasowaniu fragmentu przeglądanego drzewa XML do zdefiniowanych w pliku XSLT masek. Jeżeli takie dopasowanie powiedzie się, w miejsce dopasowanego poddrzewa XML wklejana jest zawartość zdefiniowana przez arkusz transformacji. Maski pozwalające wybrać fragment drzewa XML reprezentuje się jako wyrażenia XPath.

W opracowanym przez nas systemie wykorzystujemy technologię JSP/Java Beans do przechowywania stanu sesji użytkownika, wyświetlania graficznego szkieletu strony, jak również obsługi transformacji XSLT. Transformację XSTL stosujemy do wydobywania z modelu egzaminu wymaganych w danym momencie informacji – m.in. do uzyskania informacji na temat dostępnych egzaminów, wyświetlenia aktualnie wybranego pytania (rys. 1), komentarza do udzielonej odpowiedzi, optymalnej i wybranej przez użytkownika ścieżki (rys. 2). W implementacji systemu użyliśmy procesora XSLT Xalan 2.6 oraz serwera aplikacyjnego Tomcat 5.1.

6. Dalsze plany

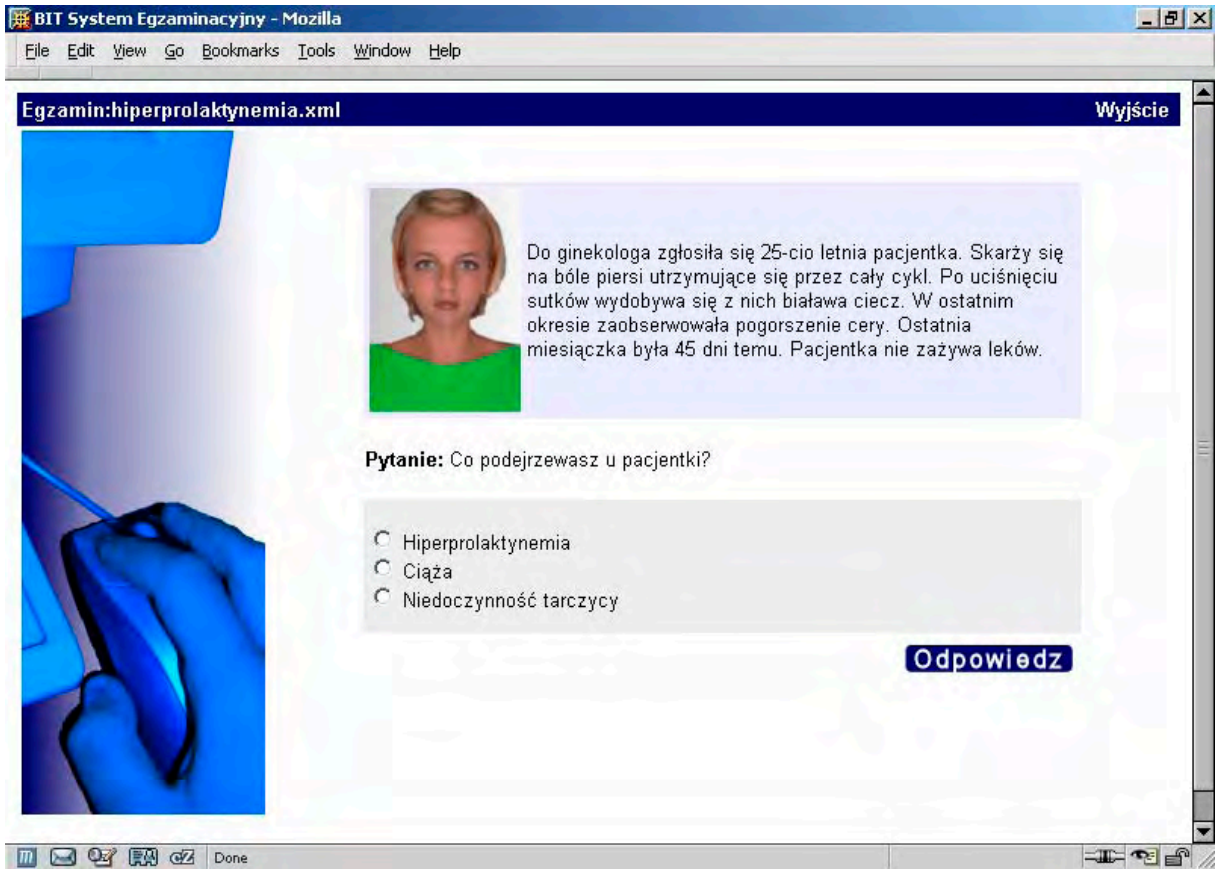
Do naszych dalszych planów należy stworzenie edytora wielościeżkowych modeli egzaminów. Aplikacja ta ma być łatwa w obsłudze i dostępna przez sieć Internet bez konieczności instalowania dodatkowego oprogramowania po stronie klienta. Celem edytora jest umożliwienie samodzielnego tworzenia wielościeżkowych egzaminów. Rozważana jest również możliwość wprowadzenia czynnika przypadkowości do tekstu pytań – wyniki badań mogłyby być losowane według określonego w tekście rozkładu prawdopodobieństwa. Planuje się również stworzenie dużej bazy egzaminów, która mogłaby być udostępniona studentom jako dodatkowa pomoc dydaktyczna.

7. Podsumowanie

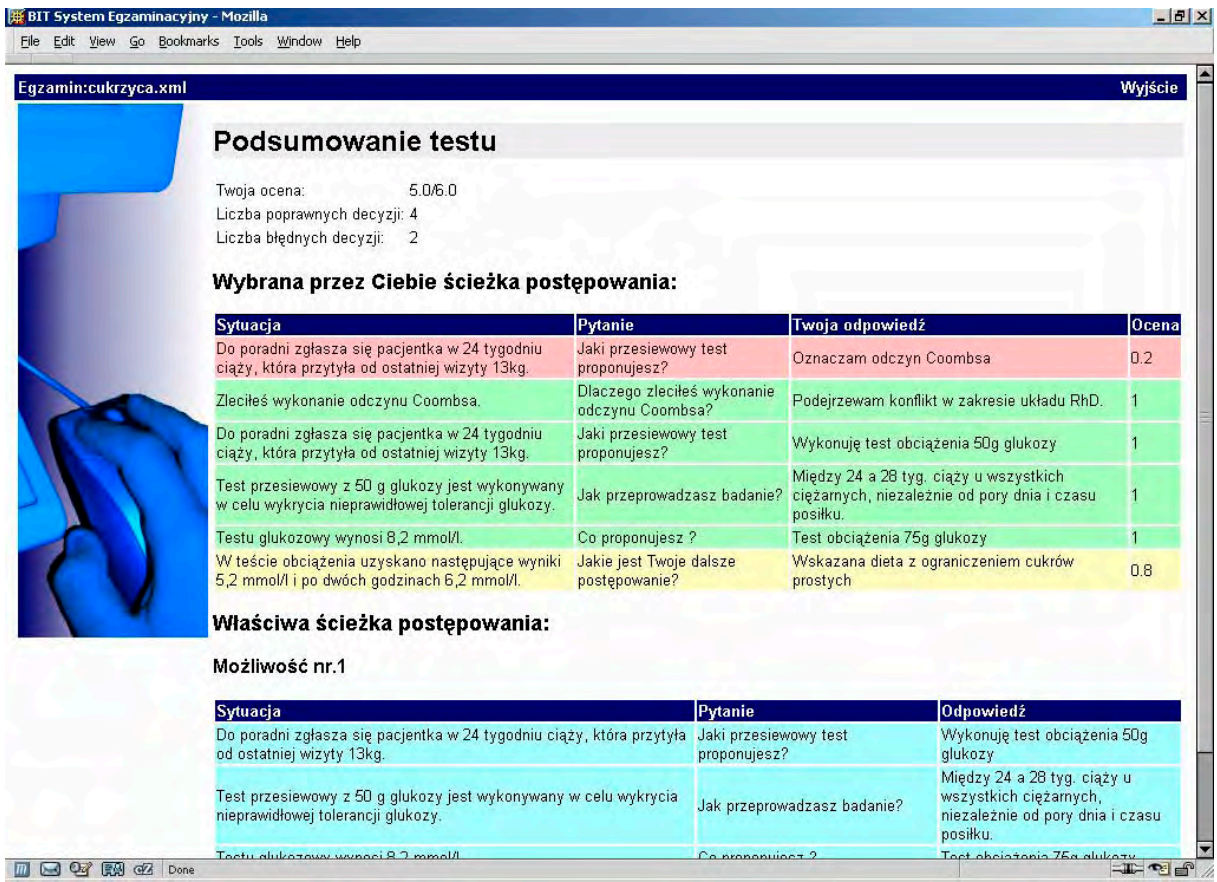
Przedstawiona pilotażowa wersja wielościeżkowego systemu egzaminacyjnego BIT jest w naszym odczuciu interesującą formą uzupełniającą tradycyjne sposoby sprawdzania wiedzy medycznej. O wartości systemu świadczyć będzie mnogość i złożoność dostępnych w nim testów. Podstawową trudnością na obecnym etapie projektu jest duży nakład pracy związany z przygotowaniem testu. Pewną pomocą w rozwiązaniu tego problemu może stać się planowany edytor testów, który pozwoli samodzielnie tworzyć egzaminy. Wykorzystanie standardów medycznych i ścieżek klinicznych jako szablonów egzaminów ułatwia proces tworzenia wielościeżkowych testów.

Literatura cytowana

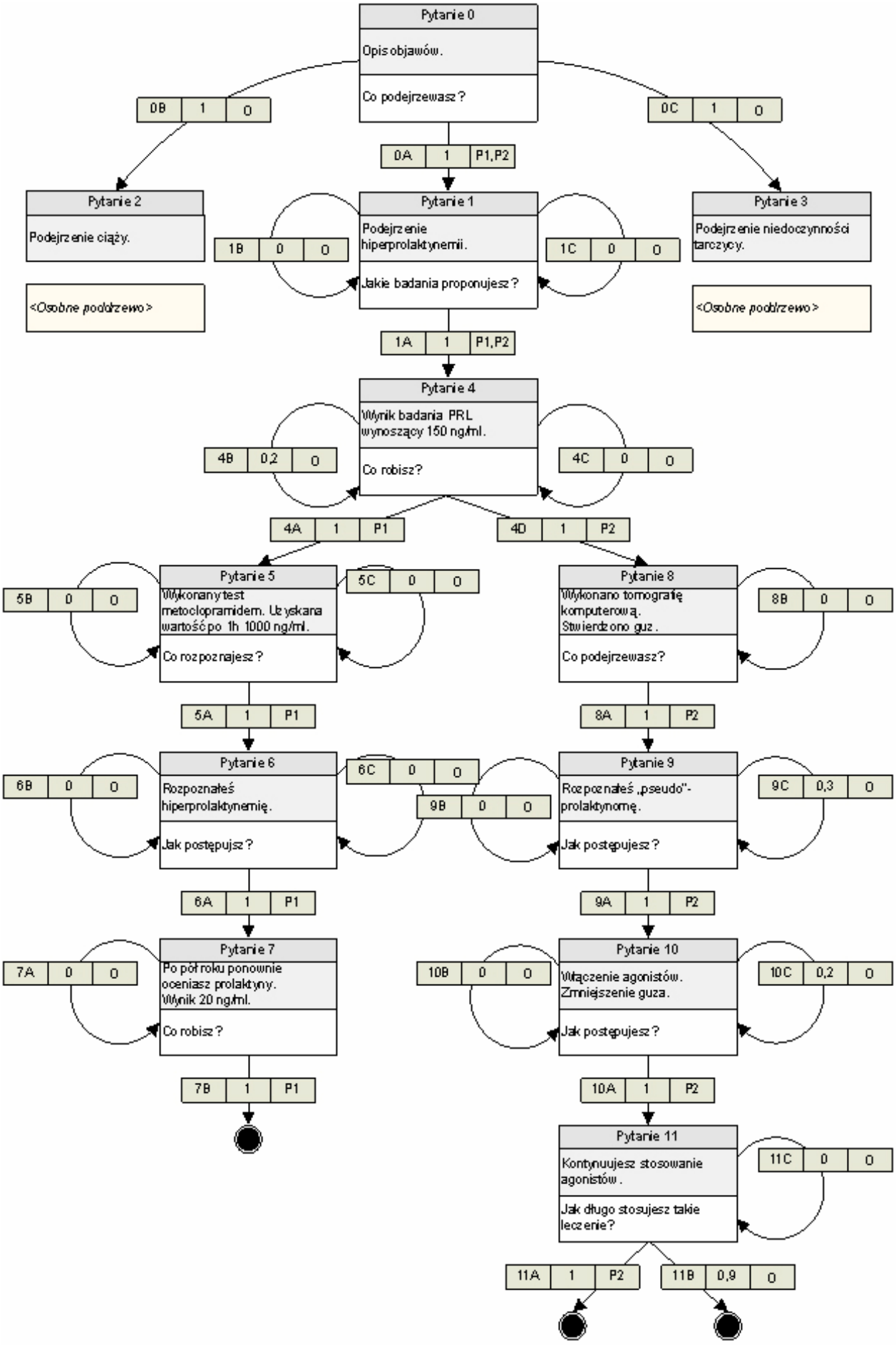
1. Hellmann W.: Klinische Pfade – Konzepte, Umsetzung, Erfahrungen, ecoMed, Landsberg 2002.
2. Holzner S.: XSLT – Anwendung und Referenz, Markt+Technik Verlag, München 2002.
3. Nakhimovsky A., Myers T.: XML Programming: Web Applications and Web Services with JSP and ASP, apress Berkeley 2002.
4. Patzer A.: JSP Examples and Best Practices, apress Berkeley 2002.
5. Romowicz W.: HTML i JavaScript, Helion, Warszawa 1998
5. Zgliczyński S., Zgliczyński W.: Standardy endokrynologii, Studio PIN : Narodowa Fundacja Endokrynologii im. Walentego Hartwiga, Warszawa 2002.



Rys. 1. Przykładowe pytanie w systemie egzaminacyjnym BIT



Rys. 2. Podsumowanie testu w systemie egzaminacyjnym BIT


Pytanie 0
Opis objawów.
Co podejrzewasz?
0B 1 0
0C 1 0
0A 1 P1,P2
Pytanie 2
Podejrzenie ciąży.
<Osobne poddrzewo>
Pytanie 1
Podejrzenie hiperprolaktynemii.
Jakie badania proponujesz?
1B 0 0
1C 0 0
Pytanie 3
Podejrzenie niedoczynności tarczycy.
<Osobne poddrzewo>
1A 1 P1,P2
Pytanie 4
Wynik badania PRL wynoszący 150 ng/ml.
Co robisz?
4B 0,2 0
4C 0 0
4A 1 P1
4D 1 P2
Pytanie 5
Wykonany test metoclopramidem. Uzyskana wartość po 1h 1000 ng/ml.
Co rozpoznajesz?
5B 0 0
5C 0 0
Pytanie 8
Wykonano tomografię komputerową. Stwierdzono guz.
Co podejrzewasz?
8B 0 0
5A 1 P1
8A 1 P2
Pytanie 6
Rozpoznałeś hiperprolaktynemię.
Jak postępujesz?
6B 0 0
6C 0 0
9B 0 0
Pytanie 9
Rozpoznałeś „pseudo"-prolaktynomę.
Jak postępujesz?
9C 0,3 0
6A 1 P1
9A 1 P2
Pytanie 7
Po pół roku ponownie oceniasz prolaktyny. Wynik 20 ng/ml.
Co robisz?
7A 0 0
Pytanie 10
Włączenie agonistów. Zmniejszenie guza.
Jak postępujesz?
10B 0 0
10C 0,2 0
7B 1 P1
10A 1 P2
Pytanie 11
Kontynuujesz stosowanie agonistów.
Jak długo stosujesz takie leczenie?
11C 0 0
11A 1 P2
11B 0,9 0

INFORMATYK *AUTOMATYCZNE PRZESYŁANIE ORAZ KONSTRUKCJA BAZY DANYCH*	⟷	*KLINICYSTA* *GINEKOLOG-POŁOŻNIK* *PROWADZENIE CIĄŻY PRAWIDŁOWEJ I PATOLOGICZNEJ*
SYSTEM OTWARTY NA DOWOLNĄ DYSCYPLINĘ MEDYCZNĄ	⟷	OCZEKIWANA SZEROKA WSPÓŁPRACA Z INNYMI DYSCYPLINAMI KLINICZNYMI

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 171-174.

ZASTOSOWANIE WIELOROZDZIELCZEJ ANALIZY SYGNAŁU DO OPISU AKTYWNOŚCI MIĘŚNIA PODCZAS CHODU

APPLICATION OF THE MULTIRESOLUTION SIGNAL ANALYSIS TO REPRESENT MUSCLE ACTIVITY DURING WALK

PRZEMYSŁAW MAZURKIEWICZ

Instytut Automatyki i Inżynierii Informatycznej Politechniki Poznańskiej, Poznań, mazurek@ar-kari.put.poznan.pl

Streszczenie. W artykule zaproponowano algorytm przetwarzania sygnału EMG w oparciu o wielorozdzielczą analizę sygnału. Prezentowany algorytm stanowi nowe podejście do uśredniania sygnału EMG, rejestrowanego podczas chodu. Jego działanie zaprezentowano, wykorzystując sygnały rzeczywiste. Pokazano podobieństwo między wynikami uśrednień za pomocą przedstawianego algorytmu a wartością RMS liczoną w danym oknie. Omówiono także problem doboru optymalnej falki do postawionego zadania z punktu widzenia filtracji sygnału. Uzyskane wyniki pozwalają wnioskować o przydatności proponowanej metody w analizie chodu.

Słowa kluczowe: EMG, analiza wielorozdzielcza, falki, RMS

Abstract. In this article a new algorythm based on the multiresolution analysis of the EMG signal has been proposed. The algorythm presented is a new approach to averaging the EMG signal recorded during walk. Real signals have been used to test the presented method. Similarity between results obtained from the implemented algorythm and the results of the RMS method has been shown. Moreover, optimal selection of the wavelet function applicable to described problem has been presented. Obtained results allow the conclusion that the featured method may be useful in gait cycle analysis.

Key words: EMG, multiresolution analysis, wavelets, RMS

1. Wstęp

Ważnym elementem każdego systemu biocybernetycznego jest sprzężenie zwrotne. W systemie FES sprzężenie zwrotne można zamknąć, wykorzystując wyniki prowadzonej na bieżąco analizy sygnału aktywności elektrycznej mięśnia (EMG). Powstawanie tego sygnału jest ściśle związane z pracą układu nerwowo-mięśniowego, którego część jest obiektem sterowania we wspomnianym systemie. Charakter sygnałów EMG wynika z morfologii mięśni, które unerwiane przez motoneurony, kurczą się w takt ich potencjałów czynnościowych (AP) i tworzą wspólnie podstawowe jednostki układu nerwowo-mięśniowego – jednostki motoryczne (MU). Znajomość sygnału EMG, będącego sumą potencjałów czynnościowych jednostek motorycznych (MUAP) mięśni aktywnych generowanych w otoczeniu elektrod rejestrujących, jest źródłem cennych informacji dla procesu sterowania układem nerwowo-mięśniowym.

2. Metody przetwarzania sygnału EMG

Ważną informacją we wspomnianym systemie jest czas aktywności mięśnia. tj. czas, w którym następuje jego skurcz. Porównanie czasów aktywności mięśni uczestniczących w ruchu danej kończyny pozwala badać relacje czasowe zachodzące między tymi chwilami, tym samym pozwala określić fazy ruchu, w jakiej znajduje się kończyna.

Istotnym sprzężeniem z punktu widzenia sterowania jest znajomość siły danego mięśnia. Wiadomo, że istnieje nieliniowa relacja między siłą skurczu a amplitudą sygnału EMG [4]. Z pewnym przybliżeniem można powiedzieć, że im większą siłę wytwarza mięsień, tym większe amplitudy mają potencjały w sygnale EMG. Ponieważ poziomy amplitud potencjałów sygnału EMG zależą od tego, jak daleko dane jednostki motoryczne znajdują się od elektrod rejestrujących, dlatego też stosuje się metody uśredniania zapisu EMG. Uśredniony zapis lepiej obrazuje zmiany amplitudy EMG, tym samym stanowi podstawowy zapis służący estymacji siły skurczu. Dodatkowo, poprzez prosty system decyzyjny oparty o progowanie z uśrednionych przebiegów EMG uzyskuje się informację o czasie aktywności danego mięśnia.

Jedną z metod uśredniania EMG jest filtracja wartości bezwzględnej sygnału filtrem dolnoprzepustowym o częstotliwości odcięcia 5-10 Hz. Uzyskane w ten sposób wyniki stanowią obwiednie analizowanego sygnału.

Podobne wyniki uzyskuje się poprzez wyznaczanie wartości RMS (1) analizowanego sygnału w zadanym oknie. Standardowe okno w badaniach EMG wynosi 25 ms.

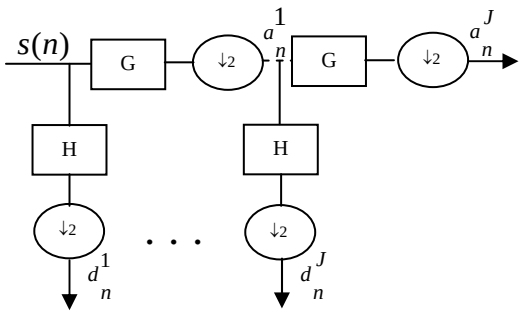
$$x_{RMS} = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_N^2}{N}} \quad (1)$$

Wyznaczona wartość jest funkcją chwili czasowej związanej z środkiem okna.

Zaprezentowane metody dają podobne rezultaty. Różnice pojawiają się w przypadku filtracji dużych częstotliwości. Metoda RMS słabiej tłumi duże częstotliwości niż filtr dolnoprzepustowy.

Przedstawione metody przetwarzania stanowią standardowe narzędzie w badaniach chodu. W dalszej części pracy przedstawiono inną metodę przetwarzania sygnału EMG w celu opisu aktywności mięśnia, opartą o wielorozdzielczą analizę sygnału.

3. Opis aktywności mięśnia za pomocą wielorozdzielczej analizy sygnału EMG



Rys. 1. Algorytm wielorozdzielczej analizy sygnału (algorytm Mallata)

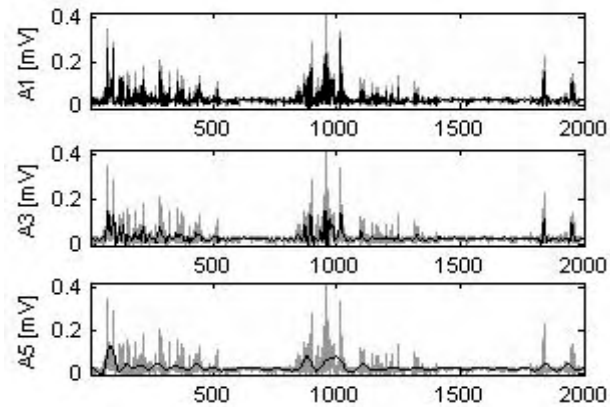
Analiza wielorozdzielcza stanowi jedno z podejść przetwarzania sygnału [2], [3]. Jej wynikiem jest reprezentacja sygnału w postaci współczynników reprezentujących detale d_n^j analizowanego sygnału na *j* (*j*=1,2,...J) poziomach dekompozycji oraz aproksymację a_n^J sygnału na J poziomie (rys. 1).

Detale uzyskuje się w wyniku filtracji górnoprzepustowej (H) i decymacji (↓2), natomiast współczynniki aproksymacji w wyniku filtracji dolnoprzepustowej (G) i decymacji. Tak realizowany algorytm umożliwia zdekomponowanie analizowanego sygnału na J pasm odpowiadających detalom oraz pasmo niskoczęstotliwościowe odpowiadające aproksymacji. Należy dodać, że wykorzystywane filtry muszą posiadać określone właściwości [1] i dodatkowo są one związane z funkcją skalującą (filtr G) oraz falką (filtr H). Powyższy algorytm nazywa się algorytmem Mallata i służy do wyznaczenia współczynników dyskretnej transformaty falkowej analizowanego sygnału (DWT).

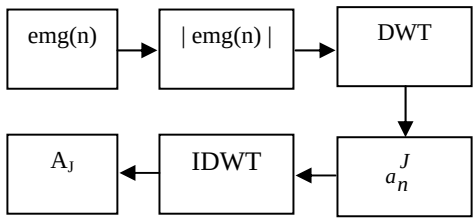
Szczególnie interesująca z punktu widzenia opisu aktywności mięśnia wydaje się własność uśredniania sygnału w torze aproksymacji w algorytmie Mallata. Im wyższy poziom dekompozycji sygnału, tym mniej szczegółów zawiera zrekonstruowany na danym poziomie sygnał aproksymacji (rys. 2). Z tego też powodu zaproponowano metodę przetwarzania elektromiogramu opartą o algorytm Mallata. Schemat proponowanej metody przedstawia rys. 3.

Dyskretnej transformacji falkowej (DWT) poddano wartość bezwzględną analizowanego sygnału emg(n) (|emg(n)|) celem wyznaczenia współczynników a_n^J określających aproksymację analizowanego sygnału na J (ostatnim) poziomie dekompozycji. Po dokonaniu odwrotnej dyskretnej transformaty falkowej (IDWT) uzyskanych sekwencji współczynników otrzymano uśredniony przebieg analizowanego sygnału A_J.

W dalszej części pracy zostaną zaprezentowane wyniki przetwarzania proponowaną metodą rzeczywistych sygnałów EMG rejestrowanych podczas chodu zwierzęcia doświadczalnego (szczura).



Rys. 2. Wartość bezwzględna sygnału oraz aproksymacje A_j (czarna linia) dla j = 1,3,5 o pasmach odpowiednio: 0-500 Hz, 0 – 125 Hz, 0 – 31.25 Hz (fp = 2000 Hz)

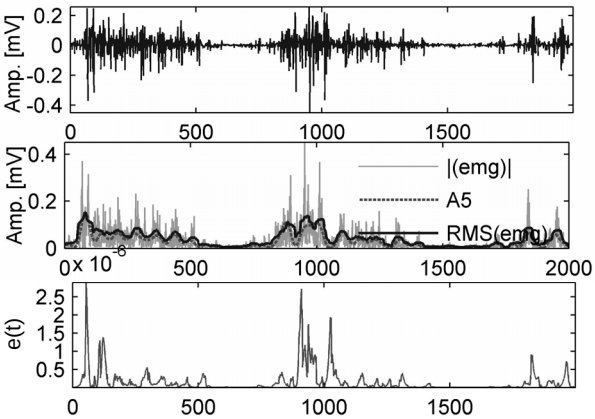


Rys. 3. Proponowany algorytm przetwarzania sygnału EMG celem analizy aktywności mięśnia

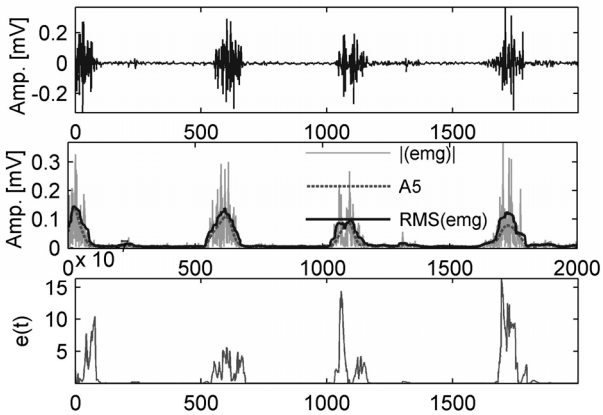
4. Wyniki

Algorytm zaimplementowano w środowisku MATLAB, wykorzystując wavelet toolbox. Przedmiotem analizy były sygnały z dwóch mięśni – *tibialis anterior* oraz *soleus*, zarejestrowane u szczura podczas chodu. Przed przystąpieniem do przetwarzania i analiz sygnału rozważono dwa zagadnienia związane z zastosowaną analizą wielorozdzielczą.

Ważnym zagadnieniem w proponowanym podejściu do uśredniania sygnału jest dobór właściwej falki, ponieważ jej wybór narzuca rodzaj funkcji skalującej, a tym samym wpływa na własności filtru dolnoprzepustowego w torze aproksymacji. Istotnym kryterium doboru falki jest płaska charakterystyka w paśmie przepustowym oraz liniowa faza. Okazuje się, że kryterium fazy jest niewystarczające, ponieważ filtry stosowane w wielorozdzielczej analizie sygnału powinny być filtrami liniowo-fazowymi. Filtry związane z danymi falkami, dostępnymi w przyborniku MATLAB posiadają mniej (rodzina falek dbN) lub bardziej liniową fazę (falki symN, biorN_d.N_r).



Rys. 4. a) Sygnał EMG mięśnia *soleus*. b) Przebiegi aproksymacji A5 uzyskane za pomocą filtru związanego z funkcją skalującą bior3.3 oraz przebieg wartości RMS liczonej w oknie 16 ms. c) Błąd średniokwadratowy między A5 a wartością RMS



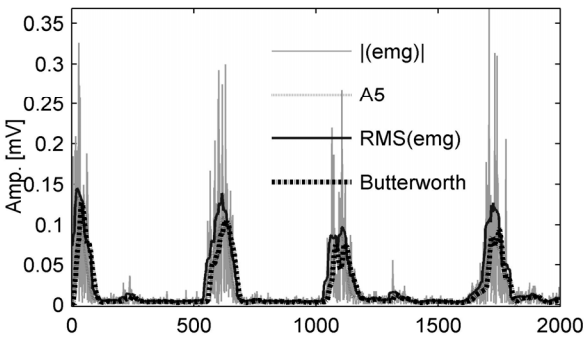
Rys. 5. a) Sygnał EMG mięśnia *tibialis*. b) Przebiegi aproksymacji A5 uzyskane za pomocą filtra związanego z funkcją skalującą bior3.3 oraz przebieg wartości RMS liczonej w oknie 16 ms. c) Błąd średniokwadratowy między A5 a wartością RMS

W celu wyboru właściwej falki jako dodatkowe kryterium przyjęto stałość opóźnienia grupowego w całym paśmie przepustowym danego filtru. Właściwość taką mają falki z rodziny symlet oraz bioortogonalne. W wyniku przeprowadzonych doświadczeń wybrano falkę z rodziny biortogonalnych. Dodatkowym problemem do rozwiązania był dobór długości nośnika stosowanego filtru, rozumianego jako liczba współczynników filtru. Wiąże się to z wyborem określonej falki z rodziny falek biortogonalnych. Wyselekcjonowana rodzina falek wykorzystuje dwa filtry dolnoprzepustowe – jeden do analizy, drugi do syntezy, odpowiada to falce biorN_d.N_r, gdzie N_d jest długością nośnika filtru dekompozycji a N_r długością nośnika filtru rekonstrukcji. Podobnie jak w przypadku doboru rodziny falek, wyboru długości nośników dokonano według tych samych kryteriów. Do dolnoprzepustowej filtracji sygnału EMG zaproponowano falkę bior3.3 ze względu na jej płaską charakterystykę amplitudową w paśmie przepustowym (rys. 4a), liniową charakterystykę fazową (rys. 4b) oraz stałe opóźnienie grupowe. Z wybraną falką związany jest filtr o długości nośnika równej $N_d = N_r = 8$. Dla takiego filtru oraz $f_s = 2000$ Hz opóźnienie grupowe liczone według (2) [1.5] wynosi G = 2 ms.

$$G = \frac{Nt_s}{2} \tag{2}$$

W równaniu (2) $N=2*N_{d,r}+2$ jest rzędem filtru, natomiast t_s jest okresem próbkowania sygnału.

Na rys. 4 przedstawiono wyniki uśredniania sygnału zarejestrowanego z mięśnia *soleus* (rys. 5a). Uśrednienia dokonano za pomocą filtru stowarzyszonego z funkcją skalującą bior3.3., na 5 poziomie dekompozycji wielorozdzielczej analizy (f_{cutoff} = 31.25 Hz). Rys. 4b przedstawia porównanie uzyskanego uśrednienia z wartością RMS liczoną w oknie 16 ms. Podobne zestawienie wyników uśrednienia zawiera rys. 5 dla mięśnia *tibialis interior*. W celu porównania uzyskanych wyników wyznaczono błąd kwadratowy między uzyskaną aproksymacją (A5) a obliczoną wartością RMS w zadanym oknie.



Rys. 6. Porównanie przebiegów aproksymacji A5 , wartości RMS liczonej w oknie 16 ms z sygnałem EMG odfiltrowanym filtrem dolnoprzepustowym typu Butterwortha 10-go rzędu (f_{cutof} = 32Hz)

5. Wnioski i spostrzeżenia

Analiza uzyskanych wyników prowadzi do wniosku, że wraz ze wzrostem poziomu dekompozycji aproksymacje zawierają mniej szczegółów (rys. 4), uśredniając analizowany sygnał w coraz wyższym stopniu. Uzyskane aproksymacje reprezentują zmianę średniej amplitudy sygnału EMG mięśni w trakcie ich skurczów. W trakcie licznych doświadczeń zaobserwowano, że dobre uśrednienie (pozbawione szczegółów związanych z szybkimi oscylacjami) ma miejsce już na poziomie 3. Wybór aproksymacji, która miałaby służyć do analizy EMG, zależy od ilości szczegółów, które chcemy zachować w sygnale. Pewnym odniesieniem mogą być wyniki uzyskane za pomocą okienkowej metody RMS, również stosowanej w badaniach EMG, w której standardowa szerokość okna wynosi 25 ms. W niniejszej pracy dla metody RMS przyjęto szerokości okna 16 ms. Jest to związane z zależnością między uśrednieniem RMS w oknie o szerokości T a filtracją dolnoprzepustową, której wyniki są podobne. Związek ten ma postać:

$$f_{cutoff} = \frac{1}{2T} \tag{3}$$

W przypadku analizowanych sygnałów f_p = 2000Hz, a ponieważ w analizie wielorozdzielczej pasma sygnałów

dzielone są przez 2^j (j=1,2..J), dlatego też najbliższą częstotliwością f_{cutoff} filtru dolnoprzepustowego w torze aproksymacji odpowiadającej RMS liczonemu w oknie 25 ms jest f_{cutoff} = 31.25 Hz (A5). Na podstawie (3) częstotliwości tej odpowiada okno RMS o szerokości 16 ms.

Porównując przebiegi RMS zarejestrowanych sygnałów oraz przebiegi A5 można zauważyć, że oba wskaźniki podobnie opisują pracę mięśni. Nieznaczne różnice pojawiają się przy krótkich skurczach. Obie metody wprowadzają małe opóźnienie. Dla porównania, wykonano dodatkową filtrację dolnoprzepustową filtrem Butterwortha 10-go rzędu (rys. 6). Uzyskane uśrednienie jest wyraźnie opóźnione w stosunku do A5. Zatem zaproponowane narzędzie analizy sygnału EMG dostarcza podobnych wyników, jak standardowa metoda (RMS). Jej przewagą jest natomiast łatwa implementacja, co może być szczególnie istotne przy przetwarzaniu dużej ilości danych. Dodatkowo, poprzez związek analizy wielorozdzielczej z transformatą falkową przedstawiona metoda stanowi uniwersalne narzędzie przetwarzania sygnału EMG, dostarczając z jednej strony informacji o średnich zmianach amplitudy sygnału, a z drugiej informacji o rozkładzie energii sygnału na płaszczyźnie skala-czas.

Analiza konkretnego sygnału EMG rejestrowanego podczas chodu zwierzęcia, przedstawionego na rys. 5b i na rys. 6b, wskazuje na zróżnicowanie przebiegów skurczów poszczególnych mięśni. Skurcze mięśnia *soleus* są dłuższe, w przeciwieństwie do skurczów mięśnia *tibialis*. Wnioski te można wyciągnąć zarówno na podstawie wyliczonej wartości RMS sygnału EMG, jak i na podstawie aproksymacji A5.

Praca częściowo finansowana ze środków KBN: projekt 1445/T11/2004/27.

Literatura cytowana

1. Augustyniak P.: Transformacje falkowe w zastosowaniach elektrodiagnostycznych, Wydawnictwo AGH, Kraków 2003.
2. Białasiewicz J.: Falki i aproksymacje, WNT, Warszawa 2000
3. Chan Y. T.: Wavelet basics, Kluwer academic publisher, Norwell 1995.
4. De Luca C. J.: The use of surface electromyography in biomechanics, Journal of Applied Biomechanics, 13(2): 135-163, 1997.
5. Lyons R. G.: Wprowadzenie do cyfrowego przetwarzania sygnałów, WKŁ, Warszawa 2000.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 175-182.

METODY BADAŃ EKSPRESJI MOWY U OSÓB PO USUNIĘCIU KRTANI

METHODS FOR SPEECH EXPRESSION ANALYSIS IN PATIENTS AFTER TOTAL LARYNGECTOMY

RAFAŁ PIETRUCH*, MARTA MICHALSKA**, WIESŁAW KONOPKA**, ANTONI GRZANKA*

**Zespół Aparatury Biocybernetycznej Instytutu Systemów Elektronicznych Politechniki Warszawskiej, Warszawa, rpietruc@elka.pw.edu.pl, antekg@ise.pw.edu.pl*
***Zakład Diagnostyki Audiologicznej i Elektrofizjologicznej Katedry Chirurgii Głowy i Szyi Uniwersytetu Medycznego w Łodzi, martamichalska@o2.pl, konopka@achilles.wam.lodz.pl*

Streszczenie. Artykuł ma na celu zaprezentowanie sposobów komputerowego przetwarzania obrazów i dźwięków wykorzystywanych w badaniach pacjentów po usunięciu krtani. Metody komputerowe mają na celu wizualizację widma głosu pacjentów, umożliwienie komputerowej analizy otrzymanych obrazów, wyselekcjonowanie najważniejszych cech sygnałów, zarchiwizowanie danych oraz analizę systematyczną otrzymanych wyników. Celem badań jest porównanie obiektywnych właściwości sygnałów z subiektywną oceną postępu pacjenta w rehabilitacji.

Słowa kluczowe: analiza widmowa sygnałów mowy, filtry adaptacyjne, przetwarzanie obrazów, programowanie DirectShow, usunięcie krtani

Abstract. The paper shows the methods for processing speech signals and imaging power spectral density of speech. The methods were applied in medicine as a tool for improving the patients' rehabilitation after tracheotomy. The adaptive algorithms and transversal filters were implemented to estimate the transfer function of human speech modeling tract. The estimation methods were based on autoregressive model. The pilot study on speakers' images, especially movements of lips and jaw, and their correlation with speech frequency has been presented.

Key words: speech signal spectrum analysis, adaptive filters, image processing, DirectShow programming, total laryngectomy

1. Wstęp

Konsekwencją usunięcia krtani jest utrata przez chorego narządu głosu. Kalectwo to znacznie utrudnia komunikowanie się z otoczeniem. Głównym celem rehabilitacji foniatrycznej jest przywrócenie osobom po operacji możliwości swobodnego kontaktu słownego z otoczeniem. Sposobem porozumiewania się osób laryngektomowanych, czyli ludzi z usuniętą krtanią, jest pseudoszept lub wykształcenie mowy zastępczej przełykowej lub gardłowej.

W zastępstwie krtani stosowane mogą być protezy głosowe auto- lub alloplastyczne. W takim przypadku następuje połączenie drogi oddechowej i pokarmowej dla celów fonacyjnych. Protezy stosowane są dość rzadko. Około 5% pacjentów korzysta z protez elektronicznych.

Po operacji usunięcia krtani często wykształca się u pacjentów pseudoszept ustno-gardłowy, pozbawiony dźwięczności, mało zrozumiały przez otoczenie. Od 60 do 70% chorych dzięki prawidłowej rehabilitacji posługuje się mową zastępczą przełykową.

Cechami głosu przełykowego są niska częstotliwość tonu podstawowego – około 60-80 Hz, duży komponent szmerowy, uboga fonacja, długi czas trwania samogłosek, podwyższone częstotliwości formantów F1 i F2, skrócenie czasu fonacji oraz zwolnienie tempa mowy. Zrozumiałość mowy przełykowej waha się od 30 do 80%. Zależy ona między innymi od rozległości zabiegu chirurgicznego, uzupełniającej radioterapii oraz rehabilitacji foniatrycznej.

Istnieją różne metody oceny głosu. Subiektywna oparta jest na ocenie słuchowej, a obiektywne oparte są na obrazach analiz akustycznych, spektrograficznych, sonograficznych oraz zapisach czasowo-częstotliwościowych sygnału mowy.
Celem pracy była:

- analiza czasowo-częstotliwościowa widma głosu u osób po laryngektomii w zależności od okresu po zabiegu operacyjnym z oceną między innymi częstotliwości formantów F1 i F2 i czasu trwania fonacji,
- ocena możliwości zastosowania techniki video w procesie rehabilitacji mowy zastępczej,

– ocena wpływu ruchów ust i twarzy na wybrane parametry widma głosu,
– ocena wpływu leczenia radioterapią na parametry widma głosu laryngektomowanych.

Materiał badań stanowią nagrania 50-ciu osób po całkowitym usunięciu krtani. Pacjentów podzielono na cztery grupy w zależności od stosowanego leczenia pooperacyjnego (radioterapii) czy rehabilitacji foniatrycznej.

Grupę I stanowią osoby nie poddawane rehabilitacji foniatrycznej, bez uzupełniającej radioterapii. Do grupy II należą osoby nie poddane rehabilitacji foniatrycznej, ale po radioterapii. Grupa III to osoby rehabilitowane foniatrycznie, bez radioterapii. Grupa IV to osoby rehabilitowane foniatrycznie po radioterapii.

Dodatkowo wszystkie grupy podzielono na podgrupy w zależności od okresu po zabiegu chirurgicznym. Podgrupę A stanowią pacjenci od 6 miesięcy do 5 lat po zabiegu, a podgrupę B pacjenci powyżej 5 lat po leczeniu operacyjnym.

Grupę porównawczą stanowiło 30 zdrowych osób w podobnym przedziale wieku.

2. Metodyka badań

Rejestracje głosu, ruchu ust i mimiki twarzy wykonano przy pomocy cyfrowej kamery z podłączonym zewnętrznym mikrofonem w odpowiednio wyciszonym pomieszczeniu. Sygnał wraz z obrazem został zapisywany na cyfrowym nośniku DV SD. Materiał lingwistyczny stanowiły wybrane słowa jednosylabowe i dwusylabowe z list wg Zakrzewskiego, Pruszewicza i Kubzdeli [14].

Wykonano analizę czasowo-częstotliwościową widma nagranego głosu z oceną między innymi czasu fonacji i częstotliwości formantów F1 i F2 z zastosowaniem techniki predykcji liniowej. Analizy zostały wykonane w Instytucie Systemów Elektronicznych Politechniki Warszawskiej.

W następnym etapie badań planowane są ćwiczenia mówienia laryngektomowanych z zastosowaniem zarejestrowanych wcześniej obrazów wideo i nagranej mowy przełykowej. Uzyskane dane poddane zostaną analizie statystycznej celem usystematyzowania wyników badań.

3. Technika predykcji liniowej w opisie sygnału mowy

Sygnał mowy x(n) powstaje w wyniku przekształcenia sygnału tonu krtaniowego g(n) w trakcie głosowym o transmitancji, funkcji przejścia, będącej transformatą *Z* odpowiedzi impulsowej h(n), przybliżonej wzorem:

$$H(z) = \frac{Q}{1 - \sum_{k=1}^{p} a_k z^{-k}}, \qquad (1)$$

Źródło: [13].

gdzie: Q – minimalny błąd średniokwadratowy predykcji [11], czemu odpowiada autoregresywna zależność w dziedzinie dyskretnego czasu:

$$x(n) = \sum_{k=1}^{p} a_k x(n-k) + Gg(n) \qquad (2)$$

Źródło: [13].

Parametry traktu głosowego a_k estymuje za pomocą modelu odwrotnego transmisji.

$$g(n) = x(n) - \sum_{k=1}^{p} a_k x(n-k) \qquad (3)$$

$$G(z) = 1 - \sum_{k=1}^{p} a_k z^{-k} \qquad (4)$$

Źródło: [13].

W badaniach zaimplementowano strukturę transwersalną filtru predykcyjnego. Zastosowano dwa adaptacyjne algorytmy wyznaczania estymat współczynników a_k filtru LMS i RLS.

Algorytm LMS:

$\hat{\mathbf{h}}$ *(0)* = 0;

x(-1) = x(-2) = ... = x(-N) = 0

Dla n = 1, 2, …

$$\mathbf{u}(n) = [x(n-1) \quad x(n-2) \quad \ldots \quad x(n-N)]^T$$
$$e(n) = x(n) - \hat{\mathbf{h}}^T(n-1)\mathbf{u}(n)$$
$$\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + 2\alpha \cdot e(n) \cdot \mathbf{u}(n)$$

Algorytm RLS:

$\hat{\mathbf{h}}$ *(0)* = 0;

x(-1) = x(-2) = ... = x(-N) = 0

R*(0) = δ*$^{-1}$I, *δ- mała liczba*

Dla n = 1, 2, …

$$\begin{aligned}
&\mathbf{u}(n) = [x(n-1) \quad x(n-2) \quad \ldots \quad x(n-N)]^T \\
&e(n) = x(n) - \mathbf{h}^T(n-1)\mathbf{u}(n) \\
&\boldsymbol{\pi}(n) = \mathbf{u}^T(n)\mathbf{R}(n-1) \\
&\kappa(n) = \lambda + \boldsymbol{\pi}(n)\mathbf{u}(n) \\
&\mathbf{k}(n) = \frac{\mathbf{R}(n-1)\mathbf{u}(n)}{\kappa(n)} \\
&\hat{\mathbf{h}}(n) = \hat{\mathbf{h}}(n-1) + \mathbf{k}(n)e^*(n) \\
&\mathbf{R}'(n-1) = \mathbf{k}(n)\boldsymbol{\pi}(n) \\
&\mathbf{R}(n) = \frac{1}{\lambda}[\mathbf{R}(n-1) - \mathbf{R}'(n-1)]
\end{aligned} \qquad (5)$$

Źródło: [3].

Wektor $\hat{\mathbf{h}}$ tworzą kolejne estymaty wartości współczynników a_k [$\hat{a}_1 \; \hat{a}_2 \ldots \hat{a}_p$]

Z badań wynika, że do śledzenia składowych częstotliwościowych sygnałów niestacjonarnych nadają się zarówno algorytmy LMS, jak i RLS. Współczynniki algorytmów: α, λ zostały wyznaczone dla każdego filtru doświadczalnie. Dla

sygnałów mowy w programie służącym do badań wybrano następujące wartości:

$$\alpha = 0.05, \ \lambda = 0.985.$$

W zależności od wartości współczynników algorytmy są zbieżne lub rozbieżne, jeśli są zbieżne, to zależy od nich szybkość zbieżności i dokładność estymacji, czyli błąd estymacji. Według Haykin [3] podany algorytm RLS (5) charakteryzuje się większą stabilnością filtru ze względu na propagację błędu kwantyzacji.

4. Estymacja widma

Dysponując w każdej chwili czasowej n zestawem wag filtra a_k, możemy je wykorzystać do estymacji chwilowego widma częstotliwościowego $X_n(e^{j\Omega})$ sygnału mowy x(n), gdyż stanowią one współczynniki aproksymującego go modelu AR:

$$X_n(e^{j\Omega}) = H(z)\Big|_{z=e^{j\Omega}} \frac{1}{1-\sum_{k=1}^{p} a_k e^{-jk\Omega}} \quad (6)$$

gdzie Ω oznacza pulsację unormowaną względem częstotliwości próbkowania $2\pi f/f_p$.

Z widma wyznacza się charakterystykę amplitudową $A(f) = |X_n(e^{j\Omega})|$.

$$A(\Omega) = \frac{1}{\sqrt{\Re e^2 + \Im m^2}}$$
$$\Re e = 1 - \sum_{n=1}^{N} h_{n-1}\cos(n\Omega) \quad (7)$$
$$\Im m = \sum_{n=1}^{N} h_{n-1}\sin(n\Omega)$$

Dla większej czytelności widma sygnału stosuje się na sygnale wejściowym tzw. preemfazę. Jest to zabieg powodujący uwypuklenie wyższych częstotliwości, a tym samym wyrównanie widma tonu krtaniowego, którego amplituda ma postać zbliżoną do malejącej wykładniczo funkcji wraz ze wzrostem częstotliwości [7]. W celu preemfazy został użyty filtr MA o współczynnikach b = [1 -0.9735], a = 1.

5. Lokalizacja twarzy

W ramach współpracy autorów publikacji zaimplementowano szereg algorytmów przetwarzania obrazu. Sprawdzono działanie algorytmów na wybrane przetwarzane klatki nagranych filmów. Kolejne procesy przetwarzania obrazów obejmowały:

- Tworzenie obrazów różnicowych kolejnych klatek, czułych na ruch głowy i ust osób występujących w filmie (rys. 1).

Rys. 1. Obraz różnicowy dwóch klatek przykładowego filmu

- Przekształcenie obrazu w binarny za pomocą zdefiniowanej granicy przynależności bitowej dla wartości pikseli (rys. 2).

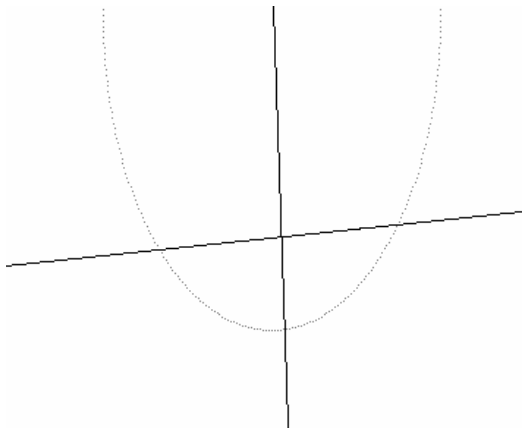
Rys. 2. Binarny obraz różnicowy dwóch klatek przykładowego filmu

- Wyznaczenie szkieletu obrazu binarnego za pomocą algorytmu Hilditch'a opisanego przez Azar (rys. 3). Źródło: [1].

Rys. 3. Szkielet obrazu różnicowego dwóch klatek przykładowego filmu

- Zastosowanie metody wyznaczania symetrii skrajnych punktów w liniach poziomych obrazów. Do punktów symetrii stosowana jest następnie transformata Hougha dla linii [8] w celu wyznaczenia prostej symetrii głowy. Linia ust znajdowana jest za pomocą transformaty Hougha dla linii prostopadłych do symetrii głowy (rys. 4).

- Zastosowanie transformaty Hougha dla elips [8] do wyznaczenia wielkości i umiejscowienia głowy (rys. 4). Występuje redukcja parametrów elipsy ze względu na symetrię głowy oraz założenia odnośnie umiejscowienia głowy w nagranych obrazach.



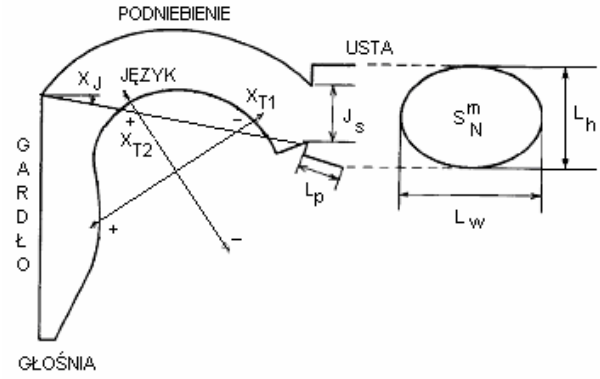
Rys. 4. Elipsa wyznaczająca miejsce głowy w obrazie i proste wyznaczające miejsce ust i symetrię głowy

6. Charakteryzacja ekspresji mowy

W celu sparametryzowania otwarcia ust i ich kształtu opisano je przykładowo za pomocą trzech wielkości: wysokości otwarcia L_h, szerokości otwarcia L_w oraz wysunięcia ust L_P. Kształt ust zmienia się w pewnym stopniu zależnie od kąta rozwarcia szczęki. Następujące wielkości zależne są od pozycji szczęki w przybliżonych zależnościach:

$$\begin{aligned} L_h &= \overline{L}_h + k_1 J_S + L'_h \\ L_w &= \overline{L}_w + k_2 J_S^{k_3} + L'_w \end{aligned} \tag{8}$$

gdzie J_s jest odległością między górnymi a dolnymi siekaczami, L'_h i L'_w wiążą się z dodatkowym ruchem ust. Stałe k_1, k_2 i k_3 opisują zależności między rozwarciem szczęki a kształtem ust. $\overline{L}_h$ i $\overline{L}_w$ są wielkościami opisującymi kształt ust w czasie zamknięcia szczęki w neutralnej pozycji (rys. 5).



Rys. 5. Charakterystyczne parametry traktu głosowego. Źródło: [12]

Dynamika ruchu ust opisana została wielkościami L'_h, L'_w oraz L_P. Te składowe są ze sobą skorelowane. Do analizy zależności między tymi wielkościami można zastosować rozkład na składowe główne (PCA – z ang. *Principal Component Analysis*) otrzymując :

$$\begin{bmatrix} L'_h & L'_w & L_P \end{bmatrix}^T = \begin{bmatrix} l_h & l_w & l_P \end{bmatrix}^T X_L + \overline{L}\,, \tag{9}$$

gdzie X_L jest skalarem opisującym zmianę kształtu ust, (l_h, l_w, l_P) jest wektorem jednostkowym wskazującym kierunek główny zmian tego kształtu. Ostatecznie kształt ust zdefiniowany jest następująco:

$$\begin{bmatrix} L_h \\ L_w \\ L_P \end{bmatrix} = \begin{bmatrix} k_1 J_S + \overline{L}_h \\ k_2 J_S^{k_3} + \overline{L}_w \\ \overline{L}_P \end{bmatrix} + \begin{bmatrix} l_h \\ l_w \\ l_p \end{bmatrix} X_L \tag{10}$$

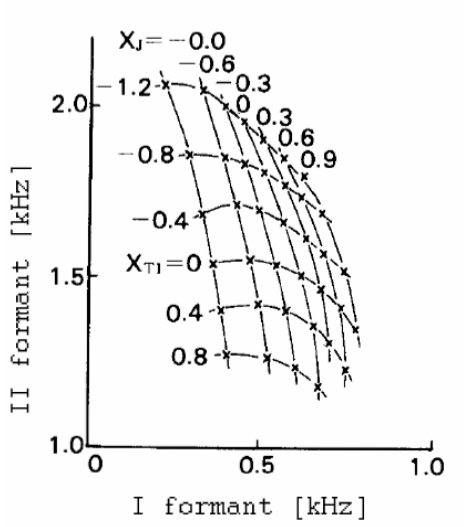
Wartości stałych występujących w równaniach podane przez Saito [12] wynoszą:

(L_h, L_w, L_P) = (0.5, 1.1, 1.0 cm),
k_1=0.6, k_2=1.61, k_3=0.2
(l_h, l_w, l_P)=(-0.38, -0.91, 0.18)

Powierzchnie otwarcia ust liczymy wg następującego wzoru:

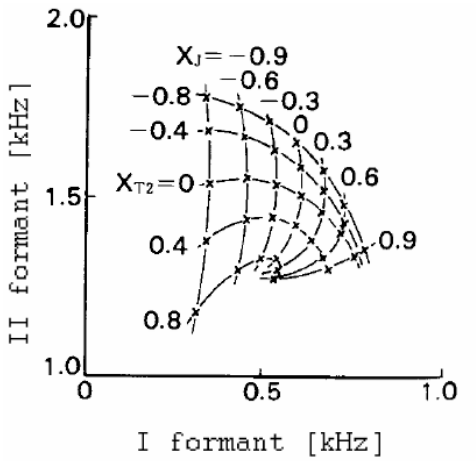
$$S_N^m = \frac{\pi \cdot L_h \cdot L_w}{4} \tag{11}$$

Relacje między parametrami artykulacji a częstotliwościami formantowymi na podstawie literatury [12] ukazują rysunki 6 i 7. Wpływ ruchu szczęki, czyli zmian X_j jest widoczny dla częstotliwości pierwszego formantu, a w mniejszym stopniu dla drugiego. Z formantem drugim silnie skorelowane jest wysunięcie języka, oznaczone jako parametr X_{T1} na rys. 5. X_{T1} w mniejszym stopniu wpływa na formant pierwszy. Zakresy X_{T1} oraz X_J rozciągają przestrzeń samogłosek /a/ /e/ oraz /i/.



Rys. 6. Zależności między formantami I i II a parametrami artykulacji X_{T1} oraz X_J. Źródło: [12]

Według Saito [12], jeśli wysokość języka, oznaczona na Rys. 5 jako X_{T2}, zwiększa się, to częstotliwości formantów wchodzą w region samogłoski /u/ w zależności od pozycji szczęki. Wartość X_{T2} jest charakterystyczna dla głoski /u/.

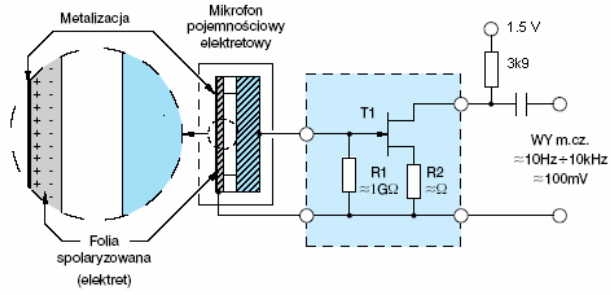


Rys. 7. Zależności między formantami I i II a parametrami artykulacji X_{T2} oraz X_J. Źródło:[12].

7. Stanowisko nagraniowe

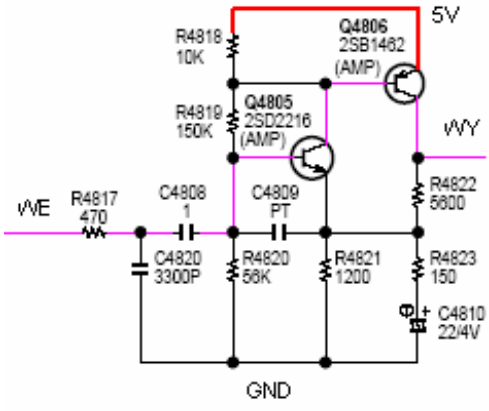
Nagrań dokonano za pomocą kamery Panasonic NV-DS65 w wyciszonej kabinie. Poziom szumów w pomieszczeniu wahał się w granicach -42dB±2dB. Pomiaru dokonano za pomocą oprogramowania CoolEdit Pro 2.0, badając przebieg sygnału nagranego podczas ciszy.

Rejestracje wykonane zostały przy pomocy zewnętrznego mikrofonu elektretowego zasilanego z baterii AA R6 1.5V podłączonej do mikrofonu jak na rys. 8.



Rys. 8. Budowa i zasilanie mikrofonu elektretowego. Źródło: [4]

Dodatkowo zastosowano potencjometr, aby przystosować moc sygnału do wejścia kamery. Wejście mikrofonowe kamery przedstawione na rys. 9 posiadało duże wzmocnienie sygnału, równe 20 dB.



Rys. 9. Mikrofonowy wzmacniacz wejściowy kamery. Źródło: [9]

Kamera ustawiona była na statywie w odległości ok. 1,5 metra od siedzącego pacjenta. Mikrofon umieszczono w odległości ok. 10 cm od ust badanych. Wszystkie dane zapisano na karcie MiniDV. Jakość nagranego dźwięku na karcie MiniDV w formacie PCM to 16 bitów (48 kHz), a jakość obrazu w formacie AVI to 720x576 rozdzielczości i częstość ramek 25 fps.

Podczas rejestracji zoom kamery ustawiono tak, aby w czasie całego nagrania widoczna była część twarzy pacjentów od wysokości łuku brwiowego do podbródka. Nagrania zostały zarchiwizowane także na płytach CD w plikach o formacie MPEG o rozdzielczości 352 na 288 pikseli, i dźwięku 16-bitowym o częstotliwości próbkowania 44,1 kHz.

Do archiwizacji sygnału w plikach użyte zostały karta wideo Pinnacle Studio Deluxe oraz program Pinnacle Studio.

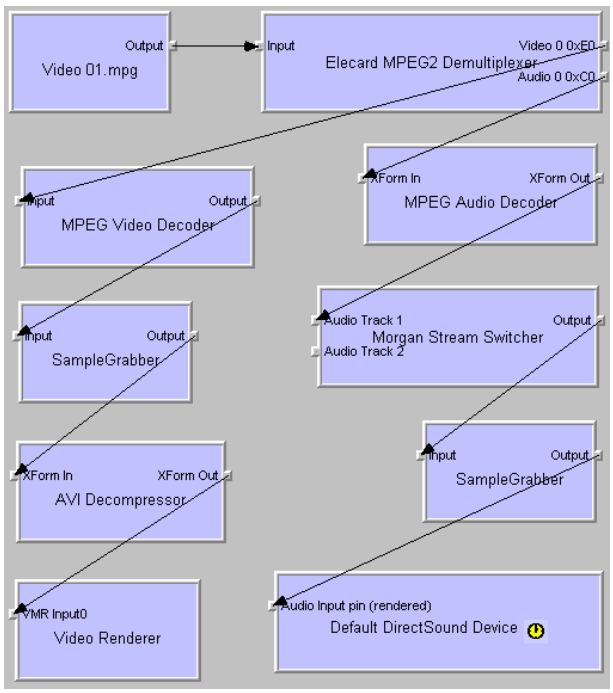
8. Program komputerowy

W ramach współpracy autorów publikacji powstał program komputerowy o nazwie AVW. Jest to aplikacja, która analizuje zapis mowy z plików audio lub wideo. Analizie podlega tylko kanał lewy sygnału audio. Program wyświetla widmo czasowo-częstotliwościowe w oknie w czasie rzeczywistym odtwarzania pliku. Umożliwia zatrzymanie odtwarzania, przewinięcie do początku pliku oraz ustawienie w dowolnym punkcie ścieżki. Dodatkowo program pozwala na podzielenie ścieżki na krótkie części, które wskazując kursorem, można odtworzyć w dowolnej chwili. Podział pliku na fragmenty wraz

z czasami oraz opisami zapisywany jest w plikach dokumentów z rozszerzeniem .avw w tym samym katalogu, co plik dźwiękowy. Występuje możliwość ustawienia barw wyświetlanego widma na kolorowe lub w poziomach szarości oraz śledzenia samych maksimów lokalnych widma. Program działa dla plików o częstotliwości próbkowania 8 kHz, ale istnieje także możliwość odtwarzania plików zapisanych z wyższą częstotliwością oraz zapisanych w formacie MPG. W przypadku wysokich częstotliwości, sygnał mowy możemy poddać operacji decymacji, obniżając częstotliwość próbkowania cztery razy. Użyty w tym celu filtr o skończonej odpowiedzi impulsowej, posiadający 59 współczynników został zaprojektowany metodą okien Kaisera [15].

Aplikacja napisana została w języku C++ za pomocą graficznego środowiska programistycznego Visual Studio .NET. Program jest wielodokumentową aplikacją MFC korzystającą z bibliotek DirectX 9.0 SDK [5]. Do zaprojektowania aplikacji użyto programu GraphEdit z pakietu programów Microsoft DirectX SDK. Program GraphEdit pozwala na projektowanie i testowanie własnych grafów DirectShow oraz na automatyczne tworzenie filtrów do wybranego pliku multimedialnego [10]. Utworzone grafy wykorzystuje się podczas pisania własnej aplikacji (rys. 10).

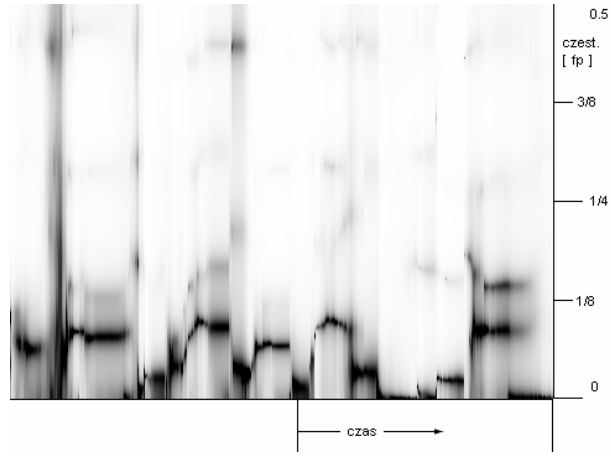
Bufor próbek sygnału audio wypełniany przy wykorzystaniu metod interfejsu ISampleGrabber, udostępnionego w bibliotece DirectX 9.0 SDK. Następnie pobrane dane są przetwarzane i analizowane, a wyniki wyświetlane w okienkach aplikacji AVW. Na rys. 10 ukazany jest graf DirectShow, zbudowany za pomocą programu GraphEdit dla jednego z nagranych filmów.



Rys. 10. Graf filtrów DirectShow

Aplikacja AVW posiada możliwość analizy sygnałów mowy pobieranych z plików o formatach WAV, AVI, MPG. Na komputerze klasy PC z procesorem taktowanym zegarem 1.67 MHz najlepsze wyniki uzyskuje się dla częstotliwości próbkowania sygnału dźwiękowego równej 8 kHz. Jest to najkorzystniejsza częstotliwość próbkowania sygnału mowy w celu estymacji widma w czasie rzeczywistym ze względu na złożoność obliczeniową oraz skuteczność algorytmów.

W okienku aplikacji wizualizowane są charakterystyki częstotliwościowo-czasowe sygnałów (rys. 11).



Rys. 11. Fragment okna programu wizualizującego przebieg spektrum mowy; wymawiane słowa to „boso, łatwo", przy wyłączonej preemfazie

W celu lokalizacji wybranych elementów twarzy napisany został filtr DirectShow [10], analizujący oraz przetwarzający kolejne klatki nagranych filmów. Korzystając z algorytmów opisanych w punkcie 5, zestaw funkcji filtru implementuje poszczególne kroki segmentacji obrazu w celu rozpoznania i umiejscowienia twarzy i ust.

9. Wyniki

Z obrazów zarejestrowanych widm można wyróżnić indywidualne cechy głosu przełykowego. W porównaniu z głosem zdrowego człowieka mowę przełykową charakteryzują m.in. niższe częstotliwości tonów, formantów (Tab. 1) i skrócenie czasu fonacji.

Tab. 1. Częstotliwości pierwszych formantów dla wybranych głosek

głoska	Pacjenci mowa zwykła	Pacjenci mowa ekspresyjna	Grupa porównawcza
/a/	f1 = 0.78, f2 = 1.5, f3 = 2.7	f1 = 0.9, f2 = 1.4, f3 = 2.6.	f1 = 0.9, f2 = 1.5, f3 = 2.8.
/i/	f1 = 0.35, f2 = 1.75, f3 = 2.8.	f1 = 0.32, f2 = 1.8, f3 = 2.8.	f1 = 0.3, f2 = 2.3, f3 = 3.5.
/e/	f1 = 0.6, f2 = 1.8, f3 = 2.8.	f1 = 0.65, f2 = 1.8, f3 = 2.5.	f1 = 0.7, f2 = 1.8, f3 = 2.7.
/y/	f1 = 0.41, f2 = 1.85, f3 = 2.4	f1 = 0.42, f2 = 2.1, f3 = 2.6.	f1 = 0.4, f2 = 1.7, f3 = 2.4.
/o/	f1 = 0.6, f2 = 1.1, f3 = 2.8.	f1 = 0.6, f2 = 1.0, f3 = 2.1.	f1 = 0.7, f2 = 2.0, f3 = 3.0.
/u/	f1 = 0.4, f2 = 1.1, f3 = 1.9.	f1 = 0.4, f2 = 1.2, f3 = 2.1.	f1 = 0.4, f2 = 1.4, f3 = 2.3

Tabela 1 zawiera przykładowe, przybliżone pomiary formantów dla wybranych osób z grupy badanej i porównawczej dla sześciu samogłosek.

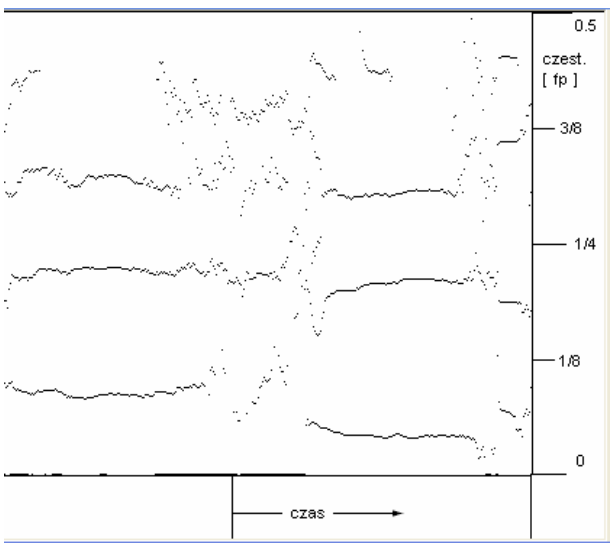
Prezentowane wyniki dotyczą trzech zdrowych osób z grupy porównawczej i trzech badanych posługujących się mową przełykową. Pomiary w grupach uśredniono.

Klasyfikacji formantów dokonano na podstawie orientacyjnego podziału pasm częstotliwości, w których mieszczą się trzy pierwsze formanty głosek mowy polskiej, podanego w literaturze [13], [16].

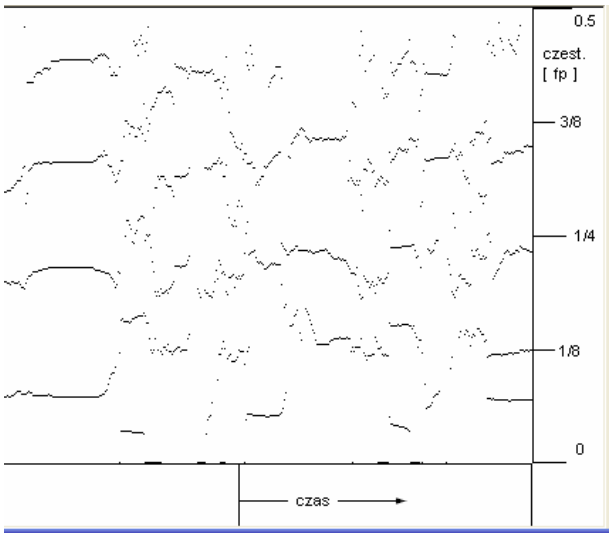
Z przeprowadzonych badań wynika, że pierwsze formanty mają podobne wartości dla grupy laryngektomowanych i grupy porównawczej. Pozostałe formanty II i III posiadają większe odchylenia przy porównywaniu badanych grup. W tym ostatnim przypadku nie zawsze można przyporządkować maksimum lokalne widma chwilowego do określonego numeru formantu.

Zgodnie z teorią zaprezentowaną w punkcie 6 (rys. 6, 7), ekspresja artykulacji ma największy wpływ na formant pierwszy. Potwierdzają to wyniki pomiarów zamieszczone w tabeli 1. Większe jest podobieństwo między wypowiedziami o większej ekspresji mowy pacjentów do widma głosu u osób zdrowych.

Dla osób laryngektomowanych występują trudności w ustaleniu wartości formantów ze względu na krótki czas fonacji i duże wahania częstotliwości formantowych dla pojedynczej głoski. W niektórych przypadkach jednoznaczną wartość formantu oraz jego numer nie udaje się dokładnie określić. Różnice między przebiegami maksimów widma chwilowego mowy u grupy pacjentów oraz grupy porównawczej ukazują rysunki 12 i 13.



Rys. 12. Przebieg formantów głosek /e/, /y/ wymawianych przez zdrowego człowieka z grupy porównawczej



Rys. 13. Przebieg formantów głosek /e/, /y/ wymawianych przez pacjenta posługującego się mową przełykową

10. Podsumowanie

Wytworzenie zastępczego głosu i mowy u osób po całkowitym usunięciu krtani warunkuje ich powrót do pracy i aktywne włączenie się do życia. Czynniki wpływające na rehabilitację głosu związane są z warunkami istniejącymi przed zabiegiem operacyjnym (wiek, płeć, rozległość procesu nowotworowego, wykształcenie), zastosowanym leczeniem (rozległość zabiegu operacyjnego, stosowanie radioterapii), a także ze stosowaniem ćwiczeń foniatrycznych. Niestety, bardzo często jakość głosu i mowy u bezkrtaniowców jest niezadowalająca.

W niniejszej pracy opisano oraz zaimplementowano metody wizualizacji widma głosu ludzkiego. Zaprezentowano metody lokalizacji twarzy w obrazach wizyjnych oraz sposób ekstrakcji cech potrzebnych do usystematyzowania parametrów ekspresji mowy. Podczas prac związanych z badaniami mowy przełykowej powstał system do wspomagania rehabilitacji pacjentów po usunięciu krtani poprzez modelowanie traktu głosowego oraz analizę sygnałów mowy i obrazów twarzy.

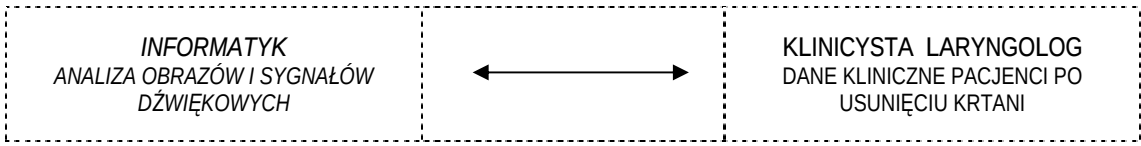
Na podstawie wstępnych wyników, sformułowano następujące wnioski:

- zastosowanie techniki audio-wideo pozwoli na analizę korelacji pomiędzy mimiką a wyrazistością mowy przełykowej, co może mieć zastosowanie praktyczne w procesie rehabilitacji mowy zastępczej,
- im większa będzie ekspresja ruchów ust i twarzy, tym mniejsza spodziewana różnica pomiędzy parametrami widma głosu laryngektomowanych a posługujących się krtanią,
- analiza czasowo–częstotliwościowa z oceną częstotliwości formantów F1 i F2, pomiary czasu trwania fonacji, tempa mowy oraz komponentów szmerowych zobiektywizują postęp rehabilitacji mowy zastępczej.

W chwili obecnej dzięki przedstawieniu osobom bez krtani metod opartych na analizie głosu i obrazu twarzy, istnieje dodatkowa możliwość na sprawne i efektywne opanowanie sztuki porozumiewania się głosem. W toku dalszych prac, poprzez szersze badania kliniczne, oceniony zostanie postęp rehabilitacji osób po laryngektomii z zastosowaniem opisanych tu metod.

Literatura cytowana

1. Azar D.: Hilditch's Algorithm for Skeletonization, University in Montreal, 1997. *http://cgm.cs.mcgill.ca/~godfried/teaching/projects97/azar/skeleton.html#algorithm*
2. Dymarski P.: Predykcyjne i wektorowe metody kompresji sygnału mowy, s.32-49, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2002.
3. Haykin S.: Adaptive filter theory, s.299-333,480-485,692-698, Prentice Hall, Inc, 1991.
4. Kozera R.: Łowcy dźwięków czyli mikrofon kierunkowy, Praktyczny Elektronik 10(99):7. *http://www.elektroda.net/download/dwpl.php?file=inne/pe2/mic_kierunkowy.pdf*
5. Microsoft, MSDN Library *http://msdn.microsoft.com/library/default.asp?url=/library/en-us/directshow/htm/introductiontodirectshow.asp*
6. Montewski M.: Procesor mowy dla bezkrtaniowców, praca dyplomowa, Instytut Systemów Elektronicznych Politechniki Warszawskiej, 1992.
7. Narożnik A.: Realizacja adaptacyjnych filtrów kratowych przy wykorzystaniu procesora sygnałowego, s.1-36, praca dyplomowa, Instytut Systemów Elektronicznych Politechniki Warszawskiej, 1998.
8. Nixon M., Aguado A.: Feature Extraction & Image Processing, s.173-198, Newnes, Oxford, 2002.
9. Panasonic: Digital Video Camera Operating Instructions, Model No. NV-DS65B, *http://www.eserviceinfo.com/download.php?fileid=1619*
10. Pence M.: Programming Microsoft DirectShow for Digital Video and Television, s.198-281, Microsoft Corporation, 2003.
11. Rutkowski L.: Filtry adaptacyjne i adaptacyjne przetwarzanie sygnałów, s.49-154, WNT, Warszawa 1994.
12. Saito S.: Speech Science and Technology, s.51-61, Ohmsha, Ltd., IOS Press, 1992.
13. Tadeusiewicz R.: Sygnał Mowy, s.158-186, WKiŁ, Warszawa 1988.
14. Zakrzewski, Pruszewicz, Kubzdela: New articulation lists matched phonematically and structurally, Otolaryngol Pol., 25(3): 297-306, 1971.
15. Zieliński T.: Od teorii do cyfrowego przetwarzania sygnałów, s.330-344, WEAIiE AGH, Kraków 2002.
16. Żyszkowski Z.: Podstawy elektroakustyki, s.218-226, WNT, Warszawa 1984.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 183-190.

METODA WYMUSZANIA UJEMNYCH IMPULSÓW CIŚNIENIA W POMIARACH PARAMETRÓW DIAGNOSTYCZNYCH UKŁADU ODDECHOWEGO

NEGATIVE PRESSURE IMPULSE METHOD IN MEASUREMENTS OF DIAGNOSTIC PARAMETERS OF HUMAN RESPIRATORY SYSTEM

PIOTR PIWOWAR

Katedra Metrologii Akademii Górniczo-Hutniczej, Kraków, ppiwowar@uci.agh.edu.pl

Streszczenie. W pracy przedstawiono metodę pomiaru parametrów diagnostycznych układu oddechowego polegającą na pomiarze i analizie przepływu powietrza w trakcie sztucznie wywołanych stanów przejściowych. Przedstawiono także system pomiarowy, model identyfikacji oraz wyniki badań dotyczących wnikliwości diagnostycznej prezentowanej metody.

Słowa kluczowe: parametry układu oddechowego, ujemny impuls ciśnienia, identyfikacja

Abstract. The paper presents a new method applied to quantitative evaluation of the diagnostic parameters of human respiratory system. This method is based on measurement and analysis of the breathing flow during artificially caused transient states. These transient states are caused by short-time negative pulses, which are generated at the outlet of the patient's respiratory system. The results of the experiments for the healthy patient and patient with the disfunction of respiratory system is also presented in the paper.

Key words: respiratory system, negative pressure impulse, identification

1. Wstęp

Jedna z metod diagnozowania stanu obiektu polega na identyfikacji parametrów tego obiektu prowadzącej do wyznaczenia opisującego go modelu parametrycznego. Wnioski dotyczące stanu obiektu oraz występujących odchyleń od normy są formułowane na podstawie wyznaczonych w procesie identyfikacji wartości współczynników modelu. Czynnikami warunkującymi dostateczną wnikliwość procesu diagnostycznego są między innymi fizykalność oraz wielowymiarowość modelu. Warunek fizykalności oznacza, że współczynniki modelu posiadają swoją fizyczną interpretację, a mianowicie są one parametrami modelowanego obiektu oraz zjawisk w nim występujących. Wielowymiarowość modelu oznacza natomiast opisywanie przez ten model wielu różnorodnych zjawisk występujących w obiekcie poddawanym diagnostyce [10].

Powszechnie stosowana obecnie metoda identyfikacji polega na pomiarowym wyznaczeniu częstotliwościowych charakterystyk układu oddechowego, a następnie na wyznaczeniu takich wartości współczynników przyjętego modelu, które zapewniają najlepszą według założonego kryterium aproksymację charakterystyk zmierzonych [10]. Przedmiotem pomiaru jest zmienność w funkcji częstotliwości impedancji skrośnej lub impedancji wejściowej. W tym celu generowany jest wokół klatki piersiowej pacjenta lub u wylotu jego dróg oddechowych sinusoidalnie zmienny sygnał ciśnienia o przestrajanej częstotliwości lub sygnał poliharmoniczny o dobranym widmie. Równocześnie jest mierzony przepływ powietrza u wylotu dróg oddechowych [1], [3], [9], [13].

Alternatywą dla metod częstotliwościowych są metody czasowe oparte na pobudzeniu badanego obiektu sygnałem nieokresowym, oraz rejestracji sygnałów generowanych przez obiekt w odpowiedzi na zadane wymuszenie [4], [10].

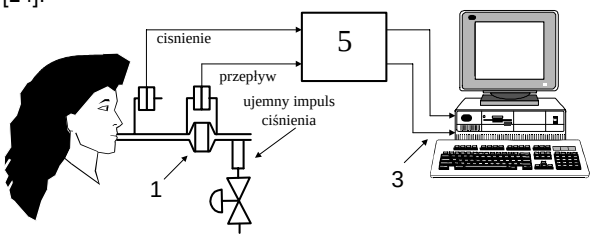
W pracy przedstawiono nową metodę przeznaczoną do diagnozowania stanu dróg oddechowych, model identyfikacji a także wpływ niektórych parametrów algorytmu na jakość identyfikacji. Pokazano także uzyskane, przy wykorzystaniu opisywanej metody, wyniki identyfikacji wybranych parametrów układu oddechowego, dla grupy osób zdrowych i chorych, które mogą być podstawą w rozpoznawaniu istotnych zmian chorobowych układu oddechowego.

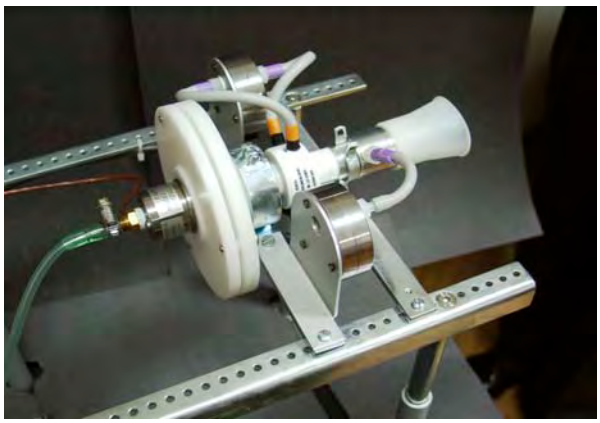
2. Metoda pomiarowa i system pomiarowy

Metoda ujemnych impulsów ciśnienia jest nową metodą pomiarową przeznaczoną do diagnozowania układu oddechowego. Polega ona na wywołaniu u wylotu dróg oddechowych pacjenta ujemnego impulsu ciśnienia o ograniczonej amplitudzie oraz czasie trwania. Zarejestrowana odpowiedź obiektu w postaci zmian przepływu powietrza, mierzonych

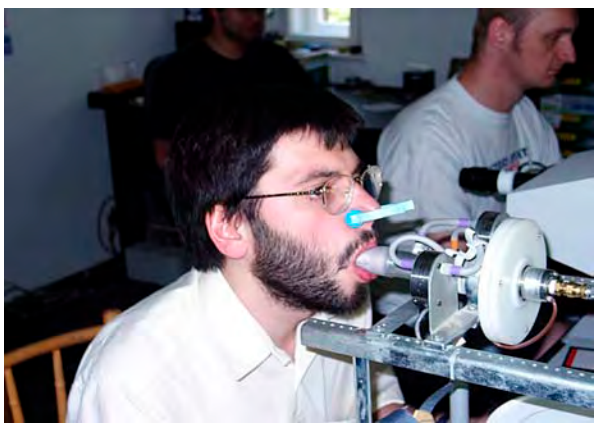
również u wylotu dróg oddechowych, a wywołana wymuszonym ujemnym impulsem, jest podstawą wyznaczenia wartości współczynników przyjętego modelu. Pomiary te są przeprowadzane na tle normalnego oddechu pacjenta, co jednak ze względu na krótki czas pomiaru nie stanowi istotnego zakłócenia [4], [6], [7]. Zaletą omawianej metody jest skrócenie czasu trwania eksperymentu oraz prosty system pomiarowy.

Schemat oraz zdjęcie systemu pomiarowego przedstawiono na rysunku 1. System zawiera tor pomiaru ciśnienia oraz tor pomiaru objętościowego przepływu oddechowego, układ (Venturiego) do generowania ujemnych impulsów ciśnienia oraz system zbierania danych. System oraz jego właściwości w sposób szczegółowy zostały opisane w pracy [14].





Rys. 1. Schemat oraz zdjęcie stanowiska pomiarowego do diagnozowania układu oddechowego metodą ujemnych impulsów ciśnienia. 1 – czujnik ciśnienia wydychanego powietrza, 2 – przepływomierz, 4 – aparat Venturiego umożliwiający generowanie ujemnych impulsów, 5 – układ kondycjonowania sygnałów pomiarowych, 3 – komputer z kartą pomiarową, 6 – czujnik ciśnienia współpracujący z pneumotachometrem, 7 – elektrozawór sterujący przepływem sprężonego powietrza, 8 – ustnik.

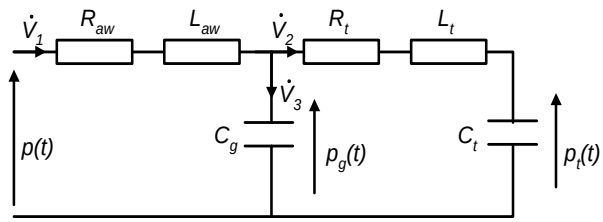


Rys. 2. Pacjent w trakcie badania metodą ujemnych impulsów ciśnienia

3. Modele obiektu identyfikacji

W literaturze [5] można spotkać wiele modeli opisujących mechaniczne właściwości dróg oddechowych człowieka wykorzystywanych do interpretowania wyników identyfikacji mechanicznych właściwości dróg oddechowych częstotliwościowymi metodami pomiarowymi. Modele te są mniej lub bardziej skomplikowane i różnią się od siebie liczbą współczynników (od trzech do dziewięciu).

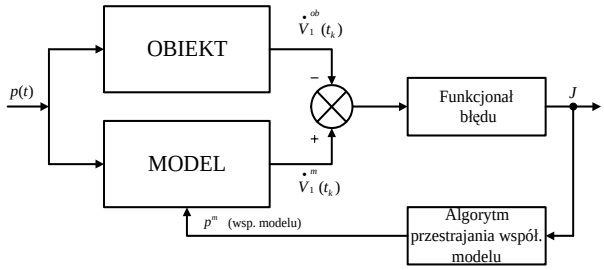
W najprostszych modelach zaproponowanych przez Mead'a (3 elementowy) i Otis'a (4 elementowy) pominięto inertancję poszczególnych elementów układu oddechowego. W pierwszym z nich założono, że składnikiem najbardziej wpływającym na impedancję dróg oddechowych jest podatność tych dróg [2]. Drugi model opiera się na założeniu, że płuca składają się ze zbioru elementów, które różnią się swoimi parametrami mechanicznymi i z powodu tych różnic pracują asynchronicznie [1]. Wpływ inertancji – pominiętej w modelach Otis'a i Mead'a – uwzględniają natomiast modele sześcioelementowy i siedmioelementowy [4], [5]. Najbardziej rozpowszechnionym i najczęściej używanym jest model o sześciu elementach [5], [6], [7]. Uwzględnia on zarówno rezystancję i inertancję dróg oddechowych, parametry części tkankowej, jak i tzw. czynnościową objętość gazu zalegającego, to jest objętość gazu pozostającą w płucach po spokojnym wydechu. Strukturę tego modelu przedstawiono na rys. 3. Częste wykorzystywanie tego modelu spowodowane jest tym, że posiada on strukturę dostatecznie skomplikowaną, by uwzględnić podstawowe zjawiska pneumatyczne i mechaniczne, a zarazem jest dość prosty, by estymacja jego współczynników była wiarygodna. Modele o większej liczbie współczynników są w zasadzie rozbudowanymi modelami sześcioelementowymi, uwzględniającymi dodatkowo wpływ pozatorakalnych dróg oddechowych [3] lub wpływ tkanki brzucha i klatki piersiowej [8]. Fragment modelu opisujący część tkankową ulega w takim przypadku rozbudowie, prowadząc do modelu o dziewięciu elementach.



Rys. 3. Model dróg oddechowych o sześciu elementach. Raw – rezystancja dróg oddechowych, Law – inertancja dróg oddechowych, Rt – rezystancja tkanek, Lt – inertancja tkanek, Ct – podatność tkanek, Cg – ściśliwość gazu pęcherzykowego charakteryzująca tzw. czynnościową objętość gazu zalegającego, to jest objętość gazu pozostającego w płucach po spokojnym wydechu

4. Algorytm identyfikacji

Identyfikacja obiektu została przeprowadzona według algorytmu strojonego modelu AMT (*Adjusted Model Technique*) (rys. 4). Współczynniki identyfikowanego modelu są estymowane w procesie strojenia i za wynik identyfikacji przyjmuje się takie ich wartości, którym odpowiada minimalna wartość funkcjonału *J* (1).



Rys. 4. Schemat blokowy ilustrujący metodę strojonego modelu AMT (*Adjusted Model Technique*)

$$J = \sqrt{\sum_{k=1}^{N} \left(\frac{\dot{V}_1^{ob}(t) - \dot{V}_1^{m}(t)}{\dot{V}_1^{ob}(t)} \right)^2}, \qquad (1)$$

gdzie:

$\dot{V}_1^{ob}(t_k)$ – przepływ zmierzony u wylotu dróg oddechowych w odpowiedzi na wymuszenie,

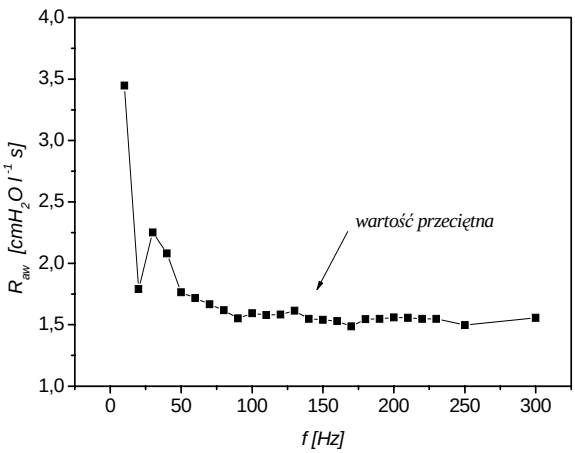
$\dot{V}_1^{m}(t_k)$ – przepływ wyliczony na podstawie przyjętego modelu dla aktualnych wartości jego współczynników i dla tego samego wymuszenia,

N – liczba próbek.

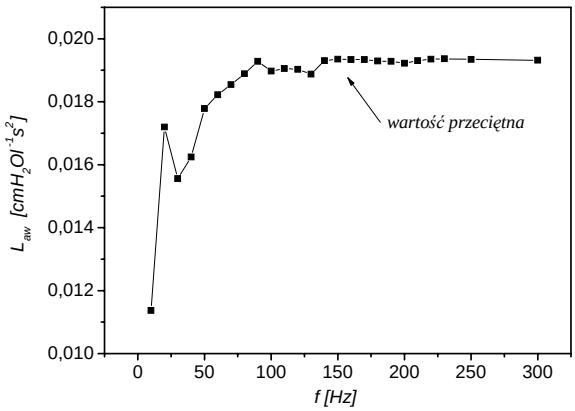
Poszukiwanie minimum kryterium jakości dostrojenia J (1) odbywa się przy wykorzystaniu procedury *fminu*, zawartej w pakiecie *Matlab for Windows™*. W celu poprawienia efektywności poszukiwania, w początkowej fazie procesu strojenia wykorzystano algorytm *Monte Carlo* (1000 iteracji).

Zgodnie z przyjętym algorytmem identyfikacji, w celu ograniczenia wpływu zakłóceń zarejestrowane sygnały pomiarowe są wstępnie filtrowane cyfrowym filtrem dolnopasmowym.

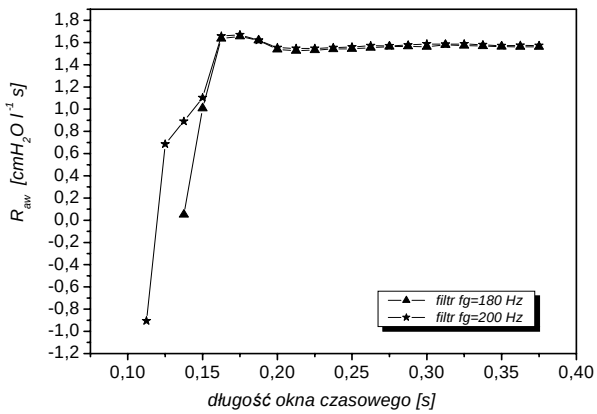
Analiza wpływu częstotliwości granicznej filtra na wynik identyfikacji poszczególnych parametrów została przeprowadzona przy wykorzystaniu modelu sześcioelementowego, a jej wyniki przedstawiono na rys. 5 i 6. Z wykresów tych wynika, że jeżeli częstotliwość graniczna filtra jest większa niż 160-170 Hz, to nie ma on wpływu na jakość identyfikacji. Oznacza to, że w widmie sygnałów pomiarowych powyżej tej częstotliwości nie ma istotnej informacji dotyczącej właściwości badanego obiektu [6], [14].



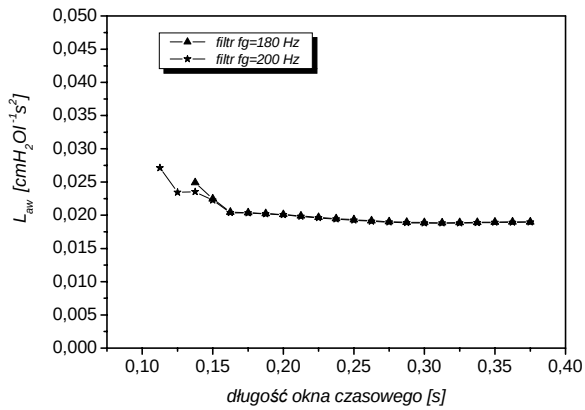
Rys. 5. Wykres zależności parametru R_{aw} modelu sześcioelementowego od częstotliwości granicznej filtra wygładzającego wyniki pomiaru ciśnienia i przepływu



Rys. 6. Wykres zależności parametru L_{aw} modelu sześcioelementowego od częstotliwości granicznej filtra wygładzającego wyniki pomiaru ciśnienia i przepływu



Rys. 7. Wpływ szerokości okna czasowego sygnałów pomiarowych użytych w procesie identyfikacji parametru R_{aw}



Rys. 8. Wpływ szerokości okna czasowego sygnałów pomiarowych użytych w procesie identyfikacji parametru L_{aw}

Innym istotnym parametrem mającym wpływ na jakość identyfikacji jest szerokość okna czasowego, decydująca o czasie obserwacji sygnałów pomiarowych użytych w procesie identyfikacji. Tak jak poprzednio, badanie zależności wyników identyfikacji od parametrów wpływowych zostało przeprowadzone przy użyciu modelu sześcioelementowego. Rysunki 7 i 8 pokazują, że zwiększanie szerokości okna cza-

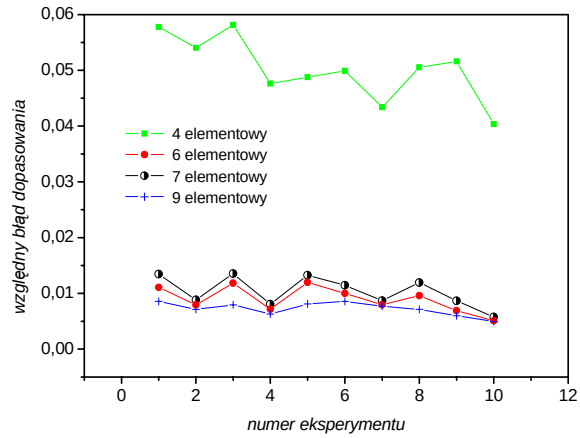
sowego sygnałów pomiarowych powyżej *200 ms* nie wpływa w sposób istotny na wynik identyfikacji parametrów obiektu [6], [14].

Wobec powyższych wniosków przyjęto częstotliwość graniczną filtra równą *180 Hz*, natomiast szerokość okna czasowego *200 ms* .

5. Wybór modelu identyfikacji

W celu wyboru modelu obiektu identyfikowanego, odpowiedniego dla omawianej metody ujemnych impulsów ciśnienia, przeprowadzono identyfikację parametrów układu oddechowego z pięcioma wybranymi modelami, wykorzystując w tym celu dziesięć sygnałów pomiarowych zarejestrowanych podczas badania zdrowego pacjenta. Modelami tymi były: czteroelementowy model Otis'a [1], model sześcioelementowy [3] (rys. 3), model siedmioelementowy [2] oraz model dziewięcioelementowy [9].

Na rysunku 9 przedstawiono wykres błędu dopasowania odpowiedzi poszczególnych modeli do odpowiedzi obiektu mierzony według kryterium (1). Z rysunku tego wynika, że spośród zaproponowanych modeli najgorszą jakość dopasowania uzyskuje się dla czteroelementowego modelu Otis'a. Rezultaty te w sposób znaczny odbiegają od wyników dostrojenia pozostałych modeli, które są do siebie zbliżone.



Rys. 9. Wykres błędu dopasowania odpowiedzi poszczególnych modeli do sygnału pomiarowego

Wyniki identyfikacji parametrów układu oddechowego dla pozostałych modeli przedstawiono w tabeli 1. Wynika z niej że najlepsze wyniki otrzymano dla modelu sześcioelementowego. Wyliczone wartości współczynników tego modelu są najbliższe tym podawanym w literaturze [7] dla metod częstotliwościowych. Jednocześnie dla tego modelu (z wyjątkiem parametru C_g) uzyskuje się najmniejsze odchylenia standardowe poszczególnych parametrów.

Tab. 1. Wyniki identyfikacji dla poszczególnych modeli

Model		9 element.		6 element.		7 elementowy		
Parametry	Wartość Przecięt.	mean	wz.std [%]	mean	wz.std [%]	mean	wz.std [%]	Parametr
Law	0,019	0,01966	6,67	0,01917	5,93	0,01965	5,99	Ld
Raw	1,74	2,48648	28,34	2,0352	23,18	2,08631	25,27	Rd+Rw
Cg	0,0035	0,0035	14,75	0,00287	15,81	0,01994	101,0	Cw
Ct (1)	0,028	-0,4723	365,74	0,02438	23,70	41,6838	33,49	Cp
Lt (1)	0,0026	0,0018	79,67	0,00301	61,12	0,00165	56,02	Rp
Rt (1)	0,87	6,8409	20,64	5,68532	12,43	0,00235	18,41	Cd
Ct (2)		0,0275	135,41					
Lt (2)		0,9809	75,38					
Rt (2)		-8,3409	82,03					

W tabeli miejsca zacieniowane wskazują najlepsze wyniki identyfikacji.

Cechą charakterystyczną modeli wykorzystywanych w praktycznych zastosowaniach jest ich liniowość. O ile w metodzie częstotliwościowej jest to podejście słuszne, gdyż amplituda wymuszającego sygnału ciśnienia nie przekracza *1 [cmH₂O]*, co odpowiada wartości przepływu około *0.5 [l·s⁻¹]*, o tyle w metodach czasowych zakres zmian ciśnień i przepływów jest na tyle duży, że wartości zmiennych obiektu w trakcie badań wykraczają poza zakres zmian fizjologicznych. Ten szeroki zakres zmian ciśnienia powoduje, że zależność miedzy ciśnieniem a przepływem może mieć charakter nieliniowy.

Nieliniowy model rezystancji pneumatycznej wynikający z prawa Rohrer'a [15] zapisany jest w postaci:

$$R = R_0 + k \cdot \left| \overset{*}{V} \right|, \qquad (2)$$

gdzie: R_0 – składnik stały (liniowy) $[cmH_2O \cdot s \cdot l^{-1}]$,

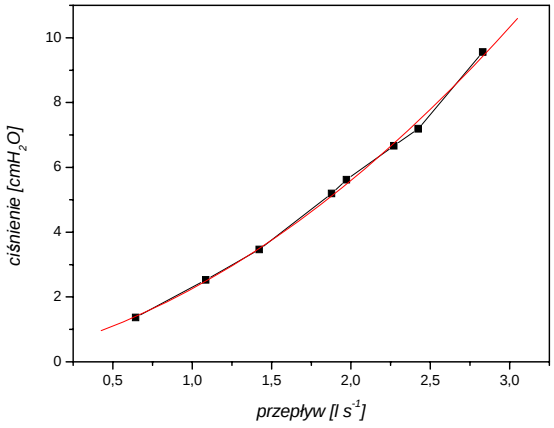
k – stały współczynnik $[cmH_2O \cdot s^2 \cdot l^{-2}]$,

$\dot{V}$ – przepływ $[l \cdot s^{-1}]$.

Uwzględniając fakt, że rezystancja dróg oddechowych R_{aw} ma największy udział (80-90%) [5] w całkowitej oporności układu oddechowego, nieliniowy model zastosowany został wyłącznie do tej rezystancji.

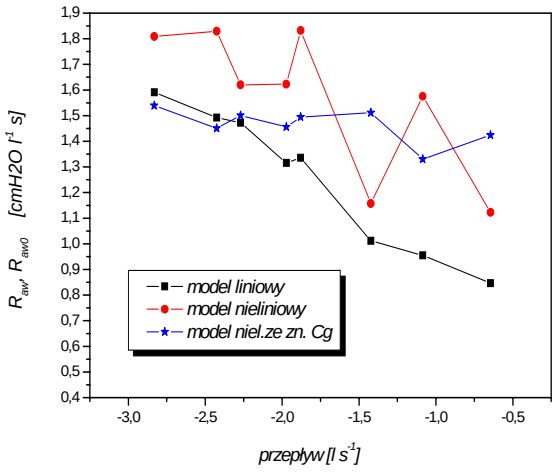
Eksperyment pomiarowy przeprowadzono na zdrowym pacjencie w ten sposób, że w zakresie od *1* do *10 [cmH₂O], co* daje wartość zmian przepływu w granicach od *0.5* do *3 [l·s⁻¹],* wytypowano osiem punktów, w których dokonano pomiaru i rejestracji sygnałów. W sumie zarejestrowano 80 sygnałów (po 10 dla każdej wartości amplitudy impulsu ciśnienia), które wykorzystano do identyfikacji parametrów obiektu.

Na rysunku 10 przedstawiono zależność ciśnienia od przepływu. Zależność ta – jak widać – jest nieliniowa, można więc założyć, że rezystancja dróg oddechowych jest zależna od przepływu zgodnie z równaniem (2).



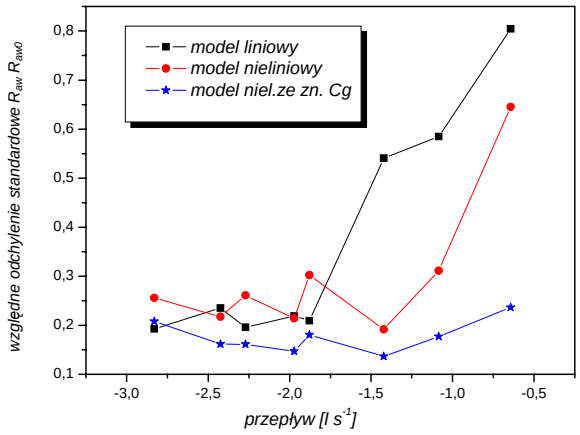
Rys. 10. Wyznaczony nieliniowy charakter zależności ciśnienia od przepływu przy użyciu metody ujemnych impulsów ciśnienia

Na rysunku 11 przedstawiono wyniki identyfikacji oporności dróg oddechowych. Dla modelu liniowego widać wyraźną zależność wartości rezystancji R_{aw} od wartości przepływu, a tym samym od amplitudy impulsu wymuszającego. Zastosowanie modelu nieliniowego niweluje tę zależność, co jest widoczne w szczególności w modelu ze znaną wartością parametru C_g [8], [14].



Rys. 11. Wartości średnie parametru R_{aw} dla liniowego i R_{aw0} dla nieliniowego modelu rezystancji uzyskane w wyniku identyfikacji prowadzonej z sześcioelementowym modelem dróg oddechowych przy znanej i nieznanej wartości parametru C_g, dla różnych wartości przepływu powietrza

Wartość odchylenia standardowego średnich wartości estymat oporności dróg oddechowych jest najmniejsza dla tego modelu i wynosi do 23% (rys. 12) [8], [14].

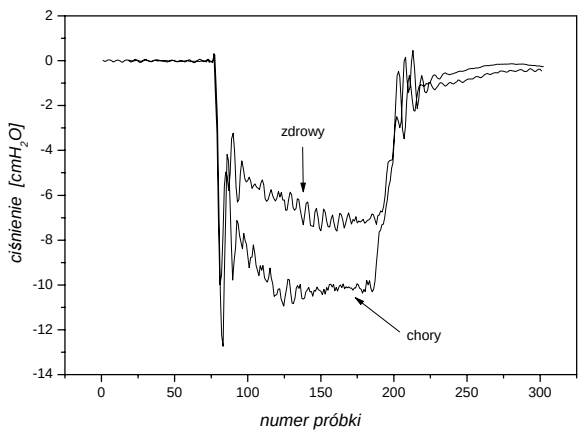


Rys. 12. Względne odchylenia standardowe wartości średnich parametru R_{aw} dla liniowego i R_{aw0} dla nieliniowego modelu rezystancji uzyskane w wyniku identyfikacji prowadzonej z sześcioelementowym modelem dróg oddechowych przy znanej i nieznanej wartości parametru C_g, dla różnych wartości przepływu powietrza

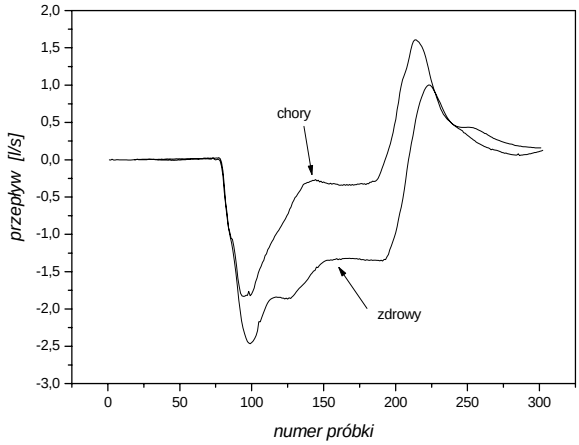
6. Wnikliwość diagnostyczna proponowanej metody

W celu sprawdzenia wnikliwości diagnostycznej proponowanej metody wykonano eksperyment polegający na identyfikacji parametrów układu oddechowego w grupie osób zdrowych i chorych. W sumie badaniu poddano czternaście osób z czego siedem stanowili zdrowi mężczyźni w wieku od 25 do 50 lat natomiast pozostałe siedem to pacjenci IGiChP w wieku 14 do 20 lat (w tym trzy kobiety i czterech mężczyzn) chorych na mukowistydozę. Dla każdego pacjenta zarejestrowano po 10 sygnałów ciśnienia oraz przepływu powietrza, które wykorzystano do identyfikacji parametrów obiektu.

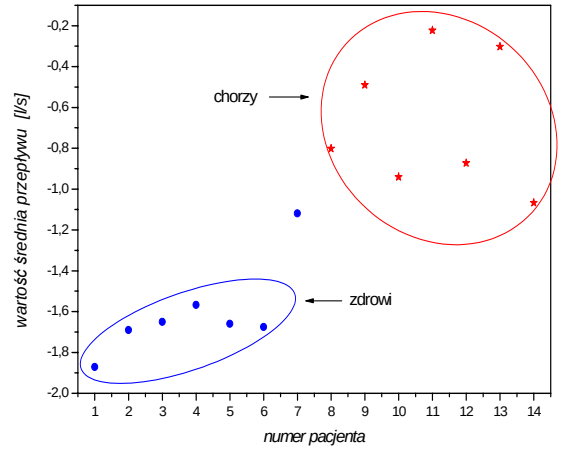
Eksperyment pomiarowy przeprowadzono w celu sprawdzenie słuszności tezy, że metodą ujemnego impulsu ciśnienia można dokonać klasyfikacji pacjentów na chorych i zdrowych. Podstawą tej klasyfikacji są wyznaczone w procesie identyfikacji parametry układu oddechowego.



Rys. 13. Przykładowe przebiegi sygnału ciśnienia zarejestrowanego u wylotu dróg oddechowych zdrowego i chorego pacjenta



Rys. 14. Przykładowe przebiegi sygnału przepływu powietrza zarejestrowanego u wylotu dróg oddechowych zdrowego i chorego pacjenta

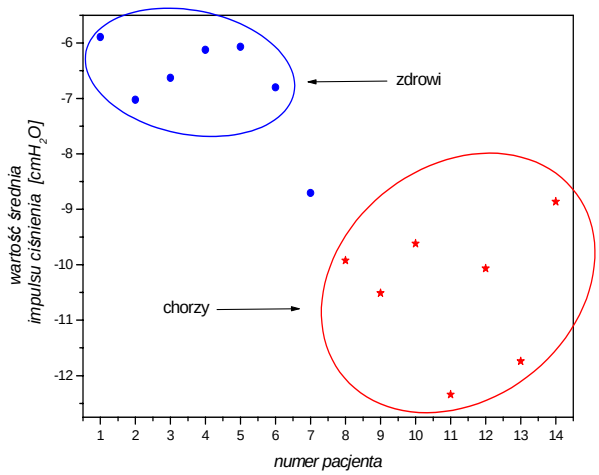


Rys. 16. Wartości średnie sygnału przepływu powietrza zmierzonego u wylotu dróg oddechowych dla osób zdrowych i chorych

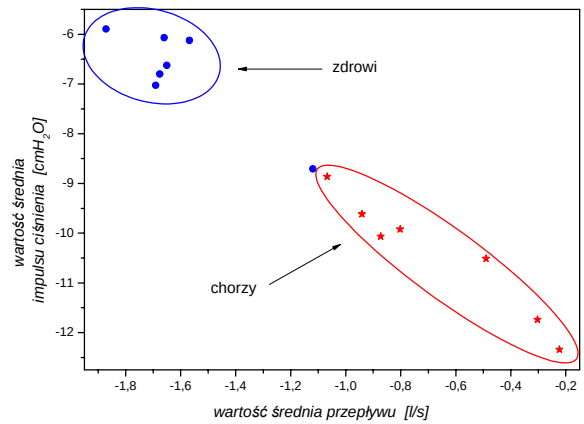
Na rysunkach 13 i 14 przedstawiono przebiegi sygnałów ciśnienia oraz przepływu powietrza zarejestrowane u wylotu dróg oddechowych pacjentów zdrowego i chorego, wywołane ujemnym impulsem ciśnienia. Można zauważyć, że różnice występują nie tylko w kształcie sygnału przepływu, ale także w wartościach minimalnych sygnału.

W celu porównania w sposób ilościowy sygnałów ciśnienia oraz przepływu powietrza, mierzonych u wylotu dróg oddechowych i pochodzących od pacjentów zdrowych i chorych, wyznaczono wartości średnie tych sygnałów za czas trwania impulsu wymuszającego, to jest 50 ms. Otrzymane wyniki, przedstawione na rysunkach 15 i 16, pozwalają w sposób jednoznaczny wydzielić grupę pacjentów z wyraźnie zmniejszoną wartością przepływu od grupy pacjentów zdrowych.

Rysunek 17 przedstawia dodatkowo wykres rozdzielenia wyników grupy osób zdrowych i chorych względem średnich wartości przepływu i ciśnienia mierzonych u wylotu dróg oddechowych.



Rys. 15. Wartości średnie sygnału ciśnienia powietrza zmierzonego u wylotu dróg oddechowych dla osób zdrowych i chorych

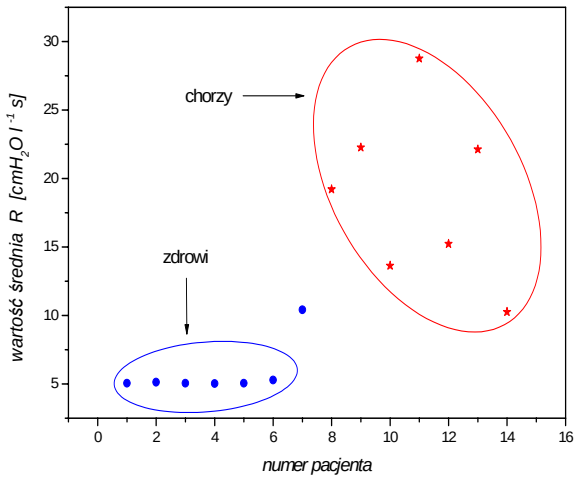


Rys. 17. Rozdzielenie grupy osób zdrowych i chorych ze względu na średnie wartości przepływu i ciśnienia

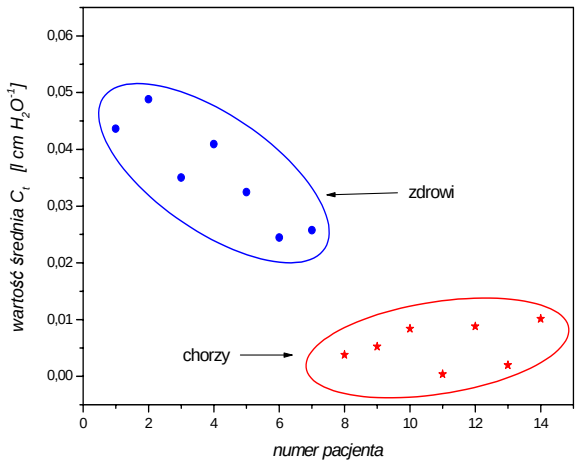
Rysunek 18 przedstawia wyniki identyfikacji parametru *R*, przyjętego jako suma oporu dróg oddechowych R_{aw} i oporu tkanek R_t. Wartości średnie tego parametru, wyliczone dla każdego pacjenta na podstawie wyników identyfikacji z 10 eksperymentów, dla osób zdrowych utrzymują się na stałym poziomie około 5 [$cmH_2O \cdot s \cdot l^{-1}$]. Wartość tego oporu dla grupy osób ze stwierdzonymi zmianami chorobowymi układu oddechowego jest kilkakrotnie większa i przekracza 10 [$cmH_2O \cdot s \cdot l^{-1}$].

Zróżnicowane wartości parametru *R* w tej grupie mogą świadczyć o stopniu zmian chorobowych. Niepewność wyników identyfikacji obrazuje odchylenie standardowe średnich wartości parametru R, które dla osób zdrowych nie przekraczają 33% a dla osób chorych lokują się poniżej 24%.

Na rysunku 19 przedstawiono wyniki identyfikacji drugiego parametru, który w sposób jednoznaczny różnicuje pacjentów zdrowych i chorych, to jest podatności C_t. Średnie wartości tego parametru dla grupy pacjentów zdrowych zawierają się w przedziale od 0.02 [$cmH_2O \cdot s \cdot l^{-1}$] do 0.05 [$l \cdot cmH_2O^{-1}$], natomiast dla grupy pacjentów ze stwierdzonymi zmianami w układzie oddechowym są wyraźnie obniżone osiągnęły wartość poniżej 0.001 [$l \cdot cmH_2O^{-1}$].



Rys. 18. Wartości średnie oporu *R* dla pacjentów zdrowych i chorych uzyskane w wyniku identyfikacji prowadzonej z sześcioelementowym modelem dróg oddechowych



Rys. 19. Wartości średnie parametru C_t dla grupy pacjentów zdrowych i chorych uzyskane w wyniku identyfikacji prowadzonej z sześcioelementowym modelem dróg oddechowych

Odchylenia standardowe średnich wartości tego parametru sięgają 50% dla osób zdrowych i 65% dla osób chorych. Pozostałe parametry (inetrancje L_{aw} i L_t) nie wykazują zmiany swoich wartości pod wpływem zmian chorobowych układu oddechowego i w praktyce klinicznej nie mają większego znaczenia diagnostycznego.

Podsumowanie

Metoda ujemnych impulsów ciśnienia, polegająca na wywołaniu przejściowych stanów dynamicznych u wylotu dróg oddechowych pacjenta, jest nową metodą pomiarową przeznaczoną do diagnozowania układu oddechowego i może być metodą alternatywną w stosunku do klasycznych metod częstotliwościowych. Zaletą opisywanej metody, w porównaniu z metodami częstotliwościowymi, jest krótki czas trwania pomiaru, łatwość przeprowadzania samego badania, jak i prosta struktura systemu pomiarowego.

Stabilne wyniki identyfikacji uzyskuje się dla czasów trwania sygnału pomiarowego przekraczających 200 ms. Zastosowanie filtru wygładzającego sygnały pomiarowe nie ma wpływu na wyniki identyfikacji jeżeli częstotliwość graniczna nie jest mniejsza niż 160 – 170 Hz.

Najlepsze wyniki identyfikacji otrzymuje się stosując model sześcioelementowy, który jest wystarczająco skomplikowany, by uwzględnić podstawowe zjawiska pneumatyczne i mechaniczne, a zarazem jest dość prosty, aby estymacja jego współczynników była wiarygodna.

Zależność między ciśnieniem a przepływem w tej metodzie ma charakter nieliniowy, co powoduje, że wartość rezystancji dróg oddechowych zależy od zmian przepływu, a tym samym od amplitudy wymuszenia. Uwzględnienie nieliniowości w przyjętym modelu poprawia wyniki identyfikacji tego parametru.

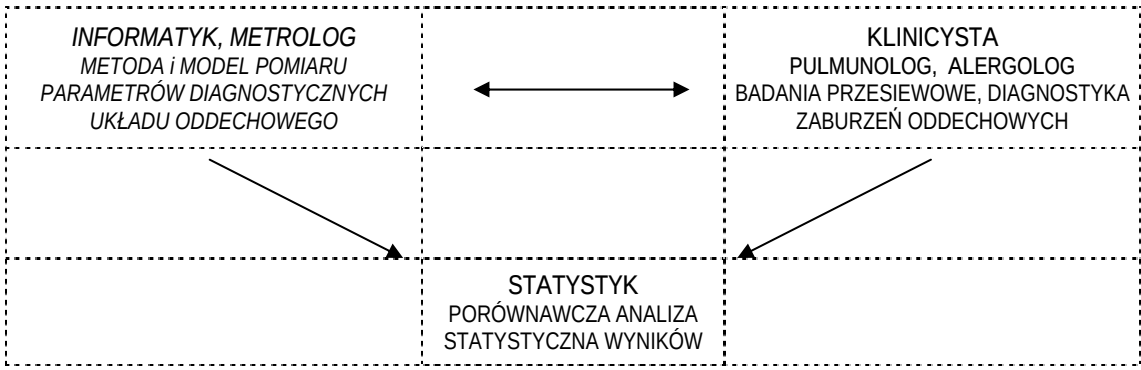
Lepsze wyniki identyfikacji parametrów uzyskuje się przyjmując dodatkowo za znaną wartość parametru C_g. Dobre wyniki identyfikacji parametrów uzyskuje się dla przepływów większych niż *2* [$l \cdot s^{-1}$], (większa stałość wartości estymowanych parametrów, jak i odchylenia standardowego w funkcji przepływu), co może być związane z tym, że w zakresie mniejszych wartości amplitud wymuszenia dla tej metody identyfikacji obiekt może nie być dostatecznie pobudzany.

Badania prowadzone tą metodą w grupie osób zdrowych oraz chorych ze stwierdzoną mukowistydozą wykazały, że parametrami różnicującymi obie te grupy są głównie zwiększone – nawet kilkakrotnie – opory oddechowe R_{aw} oraz tkankowe R_t, a także obniżona wartość podatności tkankowej C_t. Są to znane skutki obserwowalne przy chorobach obstrukcyjnych i restrykcyjnych układu oddechowego.

Literatura cytowana

1. Otis A. B., McKenow C. B., Bartlett R. A., Mead J., McIlroy M. B., Selverstone N. J., Radford E. P.: Mechanical factors in distribution of pulmonary ventilation, Journal of Applied Physiology, No. 8, 1956.
2. Nagels J., Landser F. J., Van der Linden L., Clement J., Van de Woestijne K. P.: Mechanical properties of lungs and chest wall during spontaneous breathing, Journal of Applied Physiology, No. 49, 1980.
3. Du Bois A. B., Brody A. W., Leuis D. H., Burgess B. F.: Oscillation mechanics of the lung and chest in humans, Journal of Applied Physiology, No. 8, 1956.
4. Gajda J.: A New Identyfication Algorythm of the Human Respiratory System, IEEE Instrumentation and Measurement Technology Conference, Budapest 2001.
5. Bullok J., Boyle J., Wag M. B.: Fizjologia, Wydawnictwo Medyczne, rozdz. 16, s. 239, Wrocław 1997.
6. Piwowar P.: Metoda wymuszenia ujemnych impulsów ciśnienia w identyfikacji pneumatycznych parametrów układu oddechowego, Międzyuczelniana Konferencja Metrologii MKM'2003, Kraków.
7. Buchała G., Gajda J.: Identyfikacja nieliniowego modelu dróg oddechowych, Modelowanie i Pomiary w Medycynie MPM'2000, Krynica.
8. Piwowar P.: Identyfikacja parametrów układu oddechowego metodą ujemnych impulsów ciśnienia przy użyciu nieliniowego modelu dróg oddechowych, VI Sympozjum Modelowanie i Pomiary w Medycynie MPM'2004, Krynica.
9. Tomalak W.: Modele układu oddechowego do interpretacji odechowej impedancji skrośnej, Materiały Sympozjum MiSSP'98, Krynica G. 1998.

10. Buchała G., Gajda J.: Identyfikacja impedancji wejściowej dróg oddechowych metodą czasową – badania symulacyjne, I Sympozium Modelowanie i Pomiary w Medycynie, Krynica 1999.
11. Peslin R., Papon J., Duvivier C., Richalet J.: Frequency response of the chest: modelling and parameter estimation, Journal of Applied Physiology, No. 35, 1975.
12. Buchała G., Twardowski T.: Parameter Identifiability of human airways model, MS'2000, International Conference on Modeling and Simulation, Las Palmas 2000.
13. Tomalak W., Mazurek H.: Technika oscylacji wymuszonych (FOT) w badaniu układu oddechowego. I. Podstawy teoretyczne. Interpretacja wyników pomiaru. Sprzęt., Pneumonol. Alergol. Pol., 63: 11-12, 1995.
14. Piwowar P.: Identification of Pneumatic Parameters of the Human Respiratory System by Negative Pressure Impulses, IMTC 2004 – Instrumentation and Measurement Technology Conference Como, Italy, 18-20 May 2004.
15. Rohrer R.: Der Stromungswiderstand in den menschlichen Atemwegen und der Einfluss der unregelmassigen. Verzweigung des bronchial Systems auf den Atmungsverlauf in vershiedenen Lungenbezirken Pfluegers Arch. Gesamte Physiol. Menschen Tiere 162: 225-99, 1915.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 191-196.

APLIKACJA OPARTA NA BIBLIOTECE VTK WSPOMAGAJĄCA ZABIEGI BRONCHOSKOPOWE

VTK-BASED APPLICATION SUPPORTING BRONCHOFIBEROSCOPIC PROCEDURES

MIROSŁAW SOCHA[*], MARIUSZ DUPLAGA[**]

[*] *Katedra Metrologii, Akademia Górniczo-Hutnicza, Kraków,*
[**] *II Katedra Chorób Wewnętrznych Collegium Medicum Uniwersytetu Jagiellońskiego, Kraków*

Streszczenie. Badanie bronchofiberoskopowe posiada podstawowe znaczenie w diagnostyce chorób płuc. Metody pobierania materiału do oceny histopatologicznej i mikrobiologicznej wymagają niejednokrotnie od endoskopisty dużej wprawy i dobrej orientacji w przestrzennej lokalizacji struktur klatki piersiowej. Takie techniki, jak transbronchialna igłowa biopsja aspiracyjna lub transbronchialna szczypczykowa biopsja płuca są zwykle poprzedzone oceną obrazów uzyskanych w tomografii komputerowej klatki piersiowej. Dane z tomografii komputerowej są najczęściej dostępne w postaci sekwencji kolejnych dwuwymiarowych warstw i wnioskowanie co do układu przestrzennego narządów może być trudne dla lekarza wykonujące bronchofiberoskopię.

W celu zwiększenia skuteczności technik pobierania materiału w trakcie bronchofiberoskopii autorzy przygotowali aplikację udostępniającą endoskopiście dane z tomografii komputerowej w postaci interaktywnej trójwymiarowej przestrzeni z opcją wizualizacji wirtualnej bronchoskopii. Aplikacja może służyć do oceny warunków anatomicznych w klatce piersiowej u pacjenta zakwalifikowanego do zabiegu bronchofiberoskopowego przed, w jego trakcie i po wcześniejszym wykonaniu tomografii komputerowej klatki piersiowej. Program pozwala na znaczną interakcję z danymi z tomografii komputerowej dostępnymi w standardzie DICOM. W pełni interaktywna wirtualna bronchoskopia jest generowana na podstawie serii obrazów MPR (*multiplanar reformation)*. Rekonstrukcja powierzchni drzewa oskrzelowego odbywa się algorytmem *marching cube* zaimplementowanym w bibliotece VTK. Aplikacja pozwala na rekonstrukcję przekrojów MPR w dowolnej płaszczyźnie, a także interaktywne sterowanie sposobem prezentacji danych.

Program został napisany w języku C++ i jest przeznaczony do pracy w systemie Windows 2000/XP na platformie PC. Wykorzystuje on dostępną na zasadach *open source* specjalizowaną bibliotekę VTK (Visualization ToolKit). Był on testowany w Pracowni Inwazyjnej Diagnostyki Chorób Klatki Piersiowej II Katedry Chorób Wewnętrznych CMUJ w Krakowie i uzyskał pozytywną ocenę lekarzy wykonujących zabiegi bronchofiberoskopowe.

Słowa kluczowe: inwazyjna pulmonologia, bronchofiberoskopia, transbronchialna aspiracyjna biopsja igłowa, interaktywna wirtualna bronchoskopia

Abstract. Bronchofiberoscopy is one of the basic procedures employed in pulmonary diagnostics. Tissue sampling methods for pathomorpholical and microbiological examination frequently require high skills and good recognition of spatial location of thoracic structures. Such procedures as transbronchial needle aspiration or transbronchial lung biopsy are usually preceeded by browsing of computer tomography (CT) scans of the thorax. CT data are routinely available as films with printed two-dimensional strata, which this may result in some difficulties for physician performing bronchofiberoscopy in obtaining the insight into three-dimensional location of thoracic structures.

To enhance effectiveness of sampling techniques applied during bronchofiberoscopy, the authors developed an application which enables presentation of data from CT in the form of interactive, three-dimensional environment with the option of virtual bronchoscopy visualiation. The application may be used for the analysis of anatomical conditions in thorax in specific patients before and during bronchofiberoscopy. High interactivitiy with CT DICOM data was assured. Fully interactive virtual bronchoscopy is generated on the basis of multiplanar reformation (MPR) series. The reconstruction of tracheobronchial tree surface is carried out through marching cube algorithm available in VisualisationToolkit library. The application may also be used for MPR reconstructions in any plane.

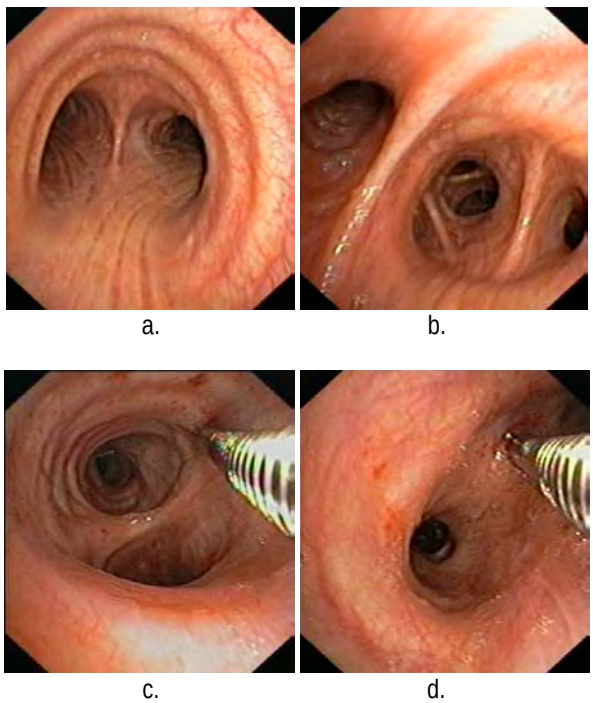
The software was designed for Windows 2000/XP system installed on PC. It was created in C++ with the use of *open source* VTK library, underwent tests in Laboratory of Invasive Diagnotics of Chest Disorders, 2nd Department of Medicine, Jagiellonian University Medical College in Krakow and received positive assessment from experienced endoscopists performing at least 200 procedures/year.

Key words: invasive pulmonology, bronchofiberoscopy, transbronchial needle aspiration, interactive virtual bronchoscopy

1. Wprowadzenie

Transbronchialna cienkoigłowa biopsja aspiracyjna jest wykorzystywana do diagnostyki zmian patologicznych zlokalizowanych w obrębie śródpiersia oraz w celu określenia stadium zaawansowania niedrobnokomórkowego raka płuca (rys. 1) [3]. Pomimo stosunkowo małej inwazyjności metoda ta nie jest powszechnie stosowana ze względu na jej ograniczoną czułość [4].

Od momentu wprowadzenia transbronchialnej igłowej biopsji aspiracyjnej do diagnostyki pulmonologicznej podejmowano próby zwiększenia czułości tej metody. W chwili obecnej zabieg jest wykonywany najczęściej po wcześniejszej lokalizacji zmian w tomografii komputerowej (TK). W znacznej mierze skuteczne pobranie materiału z guza nowotworowego czy też powiększonych węzłów chłonnych zależy od sprawności i doświadczenia badającego [4]. Przy wyborze miejsca nakłucia drzewa tchawiczo-oskrzelowego kieruje się on wyobrażeniem struktur anatomicznych na podstawie TK. W celu zwiększenia czułości transbronchialnej biopsji aspiracyjnej wykorzystuje się także inne metody obrazowania [8], np. ultrasonografię wewnątrz-oskrzelową. Ponadto przydatne mogą być w tym kontekście trójwymiarowe rekonstrukcje drzewa tchawiczo-oskrzelowego uzyskiwane poprzez zastosowanie zaawansowanych technik komputerowych [1].



Rys. 1. Obraz drzewa oskrzelowego (a,b) oraz moment transbronchialnej biopsji aspiracyjnej (c,d)

Właściwe określenie miejsca wprowadzenia igły bioptycznej w czasie bronchofiberoskopii pozwala znacząco zwiększyć czułość tej metody. Jest to szczególnie istotne w przypadku diagnostyki procesów chorobowych przebiegających z zajęciem węzłów chłonnych zlokalizowanych w obrębie śródpiersia oraz u chorych z niedrobnokomórkowym rakiem płuca, u których dąży się do określenia stadium zaawansowania nowotworu i podjęcia adekwatnych decyzji terapeutycznych.

Odpowiednia prezentacja danych uzyskanych przy pomocy technik obrazowych przed i w trakcie zabiegu bronchofiberoskopowego niesie nowe możliwości poprawy skuteczności pobierania materiału. Coraz większa moc stanowisk typu PC pozwala dzisiaj myśleć o prezentacji danych z tomografii komputerowej zapisanych w standardzie DICOM przy pomocy odpowiedniego oprogramowania na przenośnym stanowisku PC w trakcie zabiegu bronchofiberospowego. Ponadto analogia realnego obrazu endoskopowego i obrazu tzw. wirtualnej bronchoskopii, generowanego przy pomocy odpowiednich algorytmów z danych z tomografii komputerowej, daje szansę na dalszą poprawę doboru miejsca transbronchialnej biopsji aspiracyjnej [6].

Te postulaty realizuje opisany w tej pracy program przeznaczony do pracy w systemie Windows 2000/XP na platformie PC. Został on napisany w języku C++ i wykorzystuje dostępną na zasadach open source specjalizowaną bibliotekę VTK (Visualization ToolKit).

2. Standard DICOM

Wraz z rozwojem tomografii komputerowej w latach siedemdziesiątych konieczne stało się opracowanie standardu umożliwiającego wymianę danych obrazowych oraz towarzyszących im danych medycznych między urządzeniami i systemami informatycznymi. Pierwszy zarys standardu został opracowany w 1983 roku przez American College of Radiology (ACR) i National Electrical Manufacturers Association (NEMA) (ACR-NEMA Standards Publication No. 300-1985). W 1993 roku została przygotowana już trzecia wersja tego dokumentu, znacznie bardziej rozbudowana i uzupełniona o nowe aspekty. Zmieniono wówczas nazwę standardu na Digital Imaging and Communications in Medicine – DICOM [2].

Standard DICOM powstał w celu popularyzacji cyfrowej wymiany danych, ułatwienia tworzenia oraz rozbudowy systemów PACS (Picture Archiving and Comunication Systems), archiwizacji obrazów oraz wymiany danych medycznych z innymi systemami informatycznymi stosowanymi w ochronie zdrowia.

DICOM przystosowano do sieciowej wymiany informacji (*on-line*) poprzez protokół TCP/IP oraz do współpracy z nośnikami wymiennymi (*off-line*), takimi jak dyski CD czy MOD. Wymiana danych na tych nośnikach odbywa się poprzez systemy plików ISO 9660 oraz FAT16.

W standardzie DICOM wykorzystano koncepcję Service Classes. Definiuje on składnię komend oraz sposób łączenia danych. Dokładnie określony jest poziom zgodności – wyszczególniono, jakie informacje muszą być zawarte w konkretnych przypadkach (druga wersja standardu określała tylko minimalny poziom zgodności). Wprowadza także jasno sprecyzowany model informacji (Information Objects) dla różnych typów danych takich, jak obrazy, przebiegi czasowe, obiekty graficzne, raporty i wydruki. Ponadto pozwala na jednoznaczną identyfikację danych (Information Objects), a przez to na sprecyzowanie zależności między danymi, np. podczas transmisji poprzez sieć. Standard nie definiuje szczegółów implementacji, wymogów dotyczących możliwości, funkcji urządzeń oraz sposobu testowania czy zatwierdzania zgodności konkretnej implementacji.

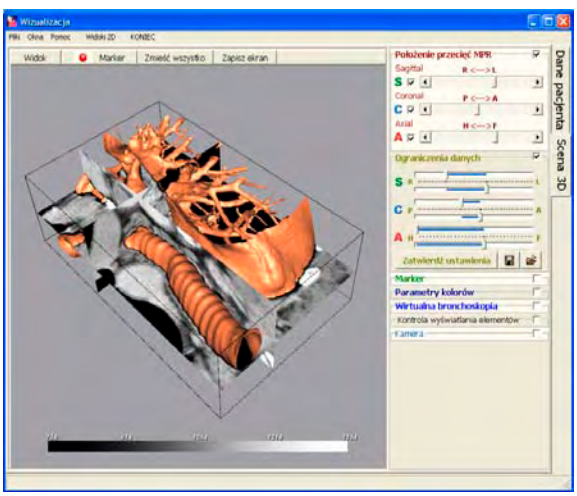
3. Biblioteka wizualizacji danych VTK

The Visualization Toolkit (VTK) jest dostępną na zasadach *open source* obiektowo zorientowaną, wieloplatformową biblioteką do wizualizacji, analizy oraz przetwarzania danych graficznych [7]. VTK zawiera ponad 700 klas (350000 linii kodu, w tym 110000 kodu wykonywalnego) napisanych w języku C++. Biblioteka współpracuje z biblioteką OpenGL oraz z językami programowania: C++, Tcl, Jawa oraz Python. Zawiera wiele algorytmów wizualizacji opartych o takie dane, jak skalary, wektory, tensory, tekstury czy dane objętościowe (wolumetryczne). Posiada zaawansowane narzędzia do modelowania: od modelowania parametrycznego, modelowania na poziomie pojedynczych punktów, przycinania, wykrywanie konturów po decymację siatek, wygładzanie powierzchni oraz triangulację [5].

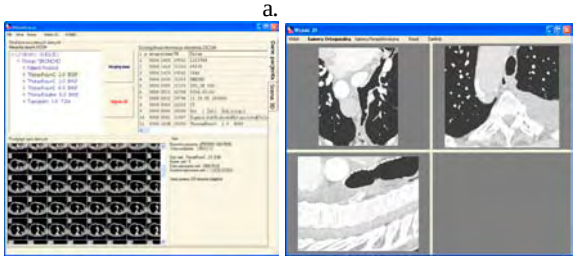
4. Program do wspomagania zabiegów bronchoskopowych

4.1. Zakres funkcjonalny programu

Program składa się z dwóch części: pierwsza służy do wczytywania danych (rys. 2b), druga zaś umożliwia wizualizację wczytanych danych (rys. 2a). Program przystosowany jest również do pracy wielomonitorowej – dostępne jest dodatkowe okno programu (rys. 2c), w którym prezentowane są wyłącznie ortogonalne przekroje MPR.



a.



b. c.

Rys. 2. Interfejs programu do interaktywnej wizualizacji danych CT: główne okno programu interfejs wizualizacji (a), główne okno programu moduł wczytywania danych (b) oraz dodatkowe okno widoków MPR (c)

Danymi wejściowymi dla programu jest płyta z danymi CT zapisana w standardzie DICOM. Program umożliwia wczytanie zawartości płyty do programu. Następnie prezentowane są odczytane zależności między danymi w postaci drzewa hierarchii. Korzeniem drzewa są informacje o pacjencie, następnie znajdują się informacje o wykonanym badaniu, serii oraz konkretnych obrazach serii. Dla serii danych generowane są podglądy zawartości poszczególnych slajdów. Lekarz ma możliwość wybrania serii zapewniającej najlepszą wizualizację. Duża liczba przekrojów w serii, mała odległość między przekrojami oraz miękki filtr rekonstrukcji są głównymi parametrami wpływającymi na jakość wizualizacji. Po wybraniu serii, program przechodzi do zakładki interfejsu – wizualizacji.

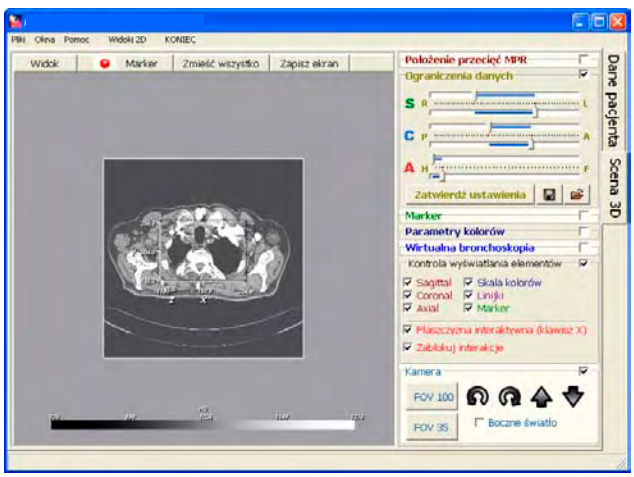
Prezentowane oprogramowanie udostępnia cztery płaszczyzny wizualizacji przekrojów MPR oraz umożliwia generowanie interaktywnego obrazu wirtualnej bronchoskopii. Trzy płaszczyzny są ortogonalne względem siebie oraz względem osi wczytanych danych.

Sterowanie płaszczyznami odbywa się przy pomocy suwaków interfejsu. Dodatkowa, czwarta płaszczyzna sterowana jest przy pomocy wskaźnika myszki i umożliwia wizualizację dowolnego przekroju MPR pod dowolnym kątem. Rysunek 6 przedstawia przykładową wizualizację z widocznymi płaszczyznami MPR oraz przestrzenną rekonstrukcją powierzchni drzewa tchawiczo-oskrzelowego.

Poza sterowaniem położeniem płaszczyzn w przestrzeni aplikacja umożliwia określenie zakresu wartości jednostek Hounsfielda prezentowanych na płaszczyznach MPR oraz wybranie palety monochromatycznej lub kolorowej.

4.2. Wizualizacja danych TK

Po wybraniu przez użytkownika konkretnej serii obrazów CT oraz zapisaniu danych o obrazach (pikseli) bez informacji DICOM w oddzielnych plikach, uruchamiany jest proces wizualizacji (rys. 3).



Rys. 3. Okno wizualizacji po wczytaniu danych DICOM

Algorytmy wizualizacji zawarte w bibliotece VTK opierają się na mechanizmach strumieniowego przetwarzania danych. Podstawowym elementem wizualizacji jest więc główny strumień wizualizacji zawierający odpowiednio połączone różne obiekty VTK, odczytujące, a następnie przetwarzające dane z tomografii komputerowej. Aby zapewnić możliwość interakcji użytkownika z procesem wizualizacji, strumień obiektów VTK obudowany został interfejsem obiektów sterujących w postaci suwaków, przełączników i innych

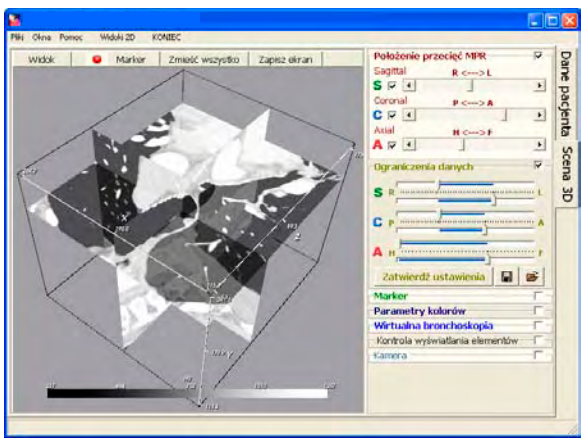
obiektów graficznych API Windows. Wykorzystanie środowiska programistycznego Borland C++ Builder, zawierającego własne klasy wykorzystujące API Windows, znacznie ułatwiło i przyspieszyło tworzenie interfejsu użytkownika.

W celu zapewnienia maksymalnej jakości wizualizacji do generowania wirtualnych obrazów drzewa oskrzelowego wykorzystuje się serie danych TK, w których odległości między kolejnymi obrazami są rzędu jednego milimetra. Zastosowanie tak gęstego skanowania pociąga za sobą poważny problem ze znaczną objętością przetwarzanych danych. Rozdzielczość pojedynczego obrazu CT wynosi 512 na 512 pikseli, każdy opisany 16-bitową wartością całkowitą. Jedna seria zawiera zazwyczaj około 350 obrazów, co daje 512x512x2x350 = 183500800 bajtów, czyli około 178 MB danych.

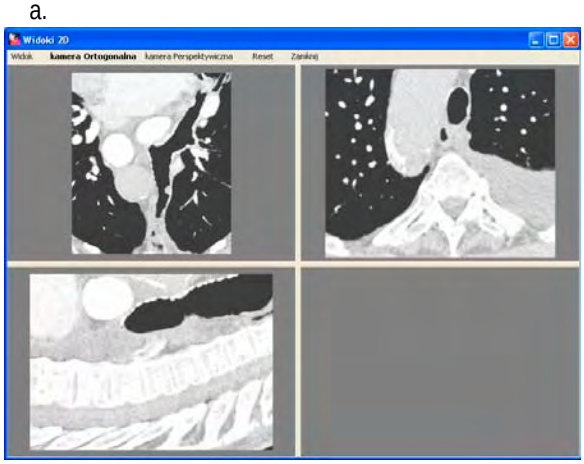
Znaczna część informacji zawartych we wczytywanych danych jest niepotrzebna do generowaniu wirtualnej bronchoskopii. Najprostszą metodą ograniczenia danych jest wybranie jedynie fragmentu danych zawierającego zakres niezbędny do dalszej obróbki. Dla każdej z trzech osi (x, y, z) określa się wartość maksymalną oraz minimalną, które definiują zawężony fragment danych.

4.3. Interfejs użytkownika

Program przystosowany jest do pracy dwumonitorowej (rys. 4) i składa się z dwóch okien: głównego okna programu oraz okna pomocniczego, przeznaczonego do wyświetlania na drugim monitorze dwuwymiarowych przekrojów MPR.



a.



b.

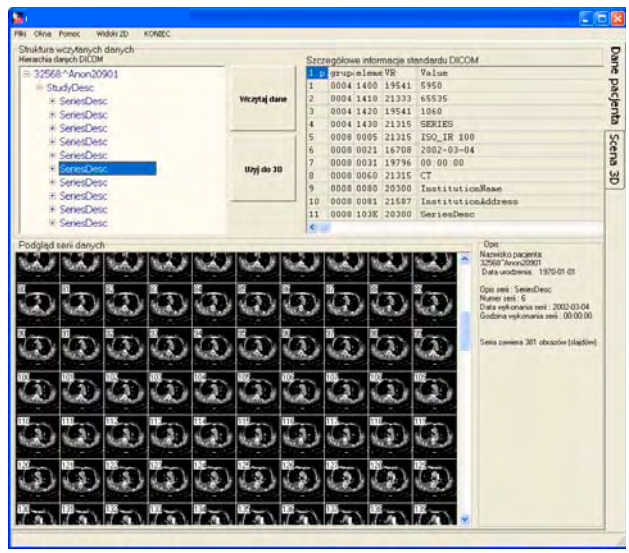
Rys. 4. Prezentacja danych na dwóch monitorach: główne okno wizualizacji (a) oraz dodatkowe okno z dwuwymiarowymi przekrojami (b).

Główne okno programu przedstawione jest na rysunku 3a. Zawiera ono menu programu, pasek stanu wyświetlający podpowiedzi oraz opisy elementów interfejsu, a także dwie zakładki: „Dane pacjenta" – zawierającą interfejs wczytywania danych oraz „Scena 3D"– służącą do prezentowania oraz sterowania wizualizacji danych.

Zakładka „Dane pacjenta" jest interfejsem wczytywanych danych (rys.5). Procedura wczytywania danych kończy się po wybraniu pliku DICOMDIR z odpowiedniego katalogu na dysku lub płycie CD – dane DICOM wczytywane są do programu. Drzewo Hierarchia danych DICOM prezentuje zależności między grupami danych DICOM odczytanych z płyty CD-R. Rozwijając drzewo oraz wybierając konkretne węzły, można zapoznać się z zawartością nośnika. Każdy węzeł drzewa prezentuje grupę elementów danych DICOM. Nie wszystkie serie danych zawierają dane nadające się do przestrzennej wizualizacji.

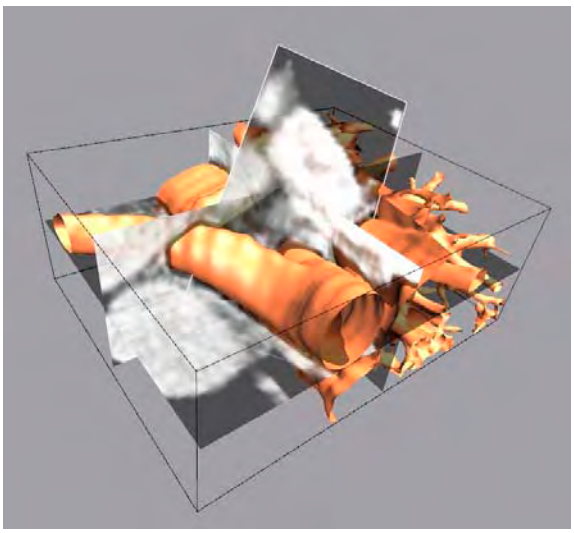
Przycisk „Użyj do 3D" umożliwia przesłanie wybranej serii danych do wizualizacji. Na podstawie wybranej serii generowany jest wirtualny obraz drzewa oskrzelowego. Program nie sprawdza, czy wybrana seria zawiera wszystkie obrazy zrekonstruowane w jednej płaszczyźnie. Kontrola, czy seria nadaje się do wizualizacji, spoczywa na użytkowniku. Otrzymanie wizualizacji o dobrej jakości wymaga serii danych zrekonstruowanych w jednej płaszczyźnie. Rekonstrukcja powinna być wykonana przy użyciu miękkiego filtra rekonstrukcji. Odległość między slajdami powinna być jak najmniejsza, co zazwyczaj wymaga dużej liczby slajdów w serii.

Zakładka „Scena 3D" służy do wizualizacji danych (rys. 4a). Zakładka zawiera okno wizualizacji, w którym przedstawiany jest widok z wirtualnej kamery obserwującej trójwymiarowe obiekty reprezentujące dane. Sterowanie kamery odbywa się przy pomocy trójprzyciskowej myszki, klawiatury oraz przycisków interfejsu. Funkcja „Marker" udostępnia opcje zapisu i odczytu pozycji znacznika.



Rys. 5. Główne okno programu – zakładka wczytywania danych

Boczne menu narzędzi zawiera elementy interfejsu umożliwiające sterowanie wizualizacją, m.in. pozwalające na kontrolę prezentowanych pozycji oraz określanie przestrzennego zakresu wizualizowanych danych.



Rys. 6. Wizualizacja danych TK – cztery płaszczyzny MPR oraz wirtualna powierzchnia drzewa tchawiczo-oskrzelowego

Możliwy jest także dowolny dobór zakresu prezentowanych danych – okna wizualizacji. Poza klasyczną, monochromatyczną skalą kolorów używaną do prezentacji danych TK na kliszach, możliwe jest zastosowanie skali kolorowej.

Grupa „Wirtualna bronchoskopia" służy do generowania trójwymiarowego modelu drzewa oskrzelowego oraz sterowania opcją przezroczystości, a grupa „Kamera" pozwala na zmianę kąta patrzenia kamery pomiędzy 100º i 35º. Obracanie kamery wokół osi prostopadłej do płaszczyzny obrazu jest możliwe poprzez użycie odpowiednich przycisków. Ponadto poprzez użycie opcji „Boczne światło" można zmienić sposób oświetlenia obiektów.

4.4. Interaktywna wirtualna bronchoskopia

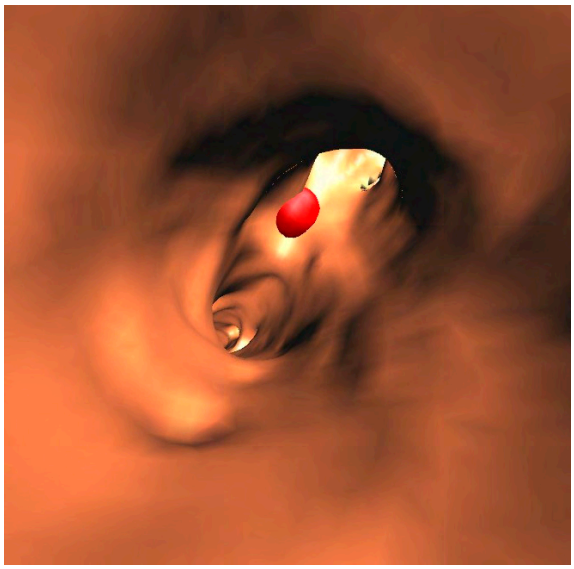
Klisze z seriami TK nie zapewniają bronchoskopiście możliwości interaktywnego przeglądania danych, np. powiększenia wybranego fragmentu obrazu czy też zmiany zakresu wartości HU wydrukowanych na kliszy danych.

Wprowadzenie obok światłoczułych klisz nośników elektronicznych, zawierających zapis wyników badania TK w standardzie DICOM, znacznie poszerzyło możliwości wykorzystania wyników tych badań przez klinicystów. Dostępne oprogramowania dla komputerów klasy PC, tzw. przeglądarki danych DICOM, umożliwiają wygodne przeglądanie danych. Ponieważ dane w standardzie DICOM zapisywane są w jednostkach HU, lekarz może zmieniać zakres gęstości w prezentowanych obrazach. Możliwa jest również zmiana palety kolorów, w której prezentowane są dane, powiększanie wybranych fragmentów danych oraz dokonywanie pomiarów.

Prezentowany program został stworzony w odpowiedzi na zapotrzebowanie wyrażone przez specjalistów zajmujących się pulmonologią interwencyjną w celu zapewnienia im jeszcze większej interakcji z danymi. Prezentowane oprogramowanie pozwala na generowanie w pełni interaktywnej wirtualnej bronchoskopii na podstawie jednej serii obrazów MPR odczytanej z danych DICOM, na rekonstrukcję przekrojów MPR w dowolnej płaszczyźnie, a także interaktywne sterowanie sposobem prezentacji danych.

Do rekonstrukcji powierzchni drzewa oskrzelowego wykorzystywany jest algorytm *marching cubes* zaimplementowany w bibliotece VTK. Algorytm wyznacza izopowierzchnię dla żądanego poziomu gęstości tkanek. W wyniku działania algorytmu powstaje ciągła siatka trójkątów, która jest następnie sprzętowo wizualizowana przy pomocy karty graficznej [5].

Program łączy dwuwymiarową wizualizację MPR z wizualizacją powierzchni drzewa tchawiczo-oskrzelowego. Dzięki temu możliwe jest wykorzystanie obrazów MPR do lokalizacji położenia węzłów chłonnych względem wirtualnej powierzchni drzewa tchawiczo-oskrzelowego oraz innych struktur. Na przecięciu ortogonalnych powierzchni MPR lekarz może ustawić znacznik w postaci sfery w celu zaznaczenia optymalnego miejsca wykonania biopsji (rys. 7).



Rys. 7. Wirtualna bronchoskopia – widok wewnętrznej powierzchni drzewa oskrzelowego oraz wskaźnika wyznaczającego dogodny punkt wykonania biopsji

Obrazy wirtualnej bronchoskopii generowane przez stworzone oprogramowanie są podobne do obrazów oglądanych podczas bronchoskopii. Zachowane są główne kształty oraz detale, takie jak: zagłębienia, bruzdy między chrząstkami oraz inne charakterystyczne kształty. Dzięki podobieństwu kształtu możliwe jest dokładne planowanie miejsca nakłucia. Przez analizę i porównanie kształtu powierzchni wirtualnego modelu oraz kształtu drzewa oskrzelowego widzianego w bronchoskopie lekarz może umieścić końcówkę bronchoskopu w taki sposób, by igła wychodząca z kanału roboczego trafiała w miejsce zaznaczone na wirtualnej bronchoskopii.

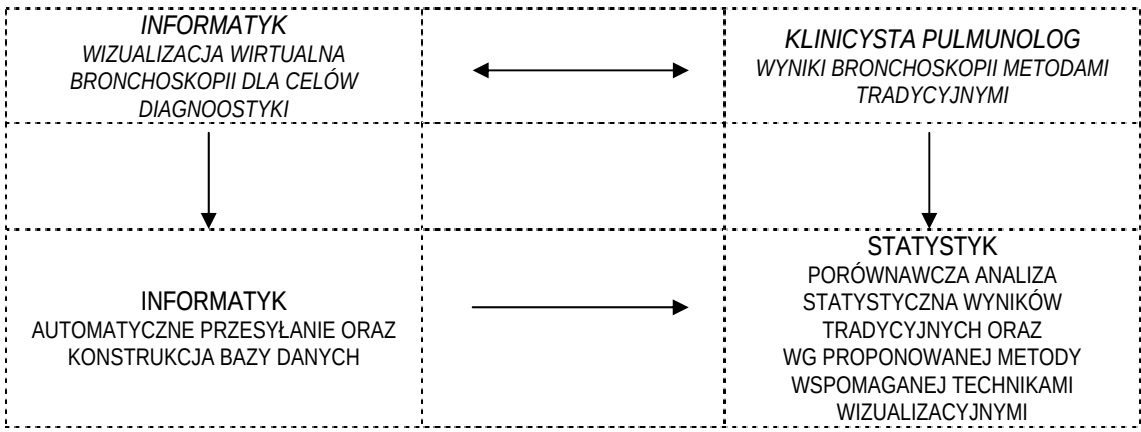
Program został poddany testom w Pracowni Inwazyjnej Diagnostyki Chorób Klatki Piersiowej II Katedry Chorób Wewnętrznych Collegium Medicum UJ w Krakowie. Wstępna ocena programu dokonana przez pulmonologów wykonujących zabiegi bronchofiberoskopowe była pozytywna. W chwili obecnej aplikacja jest wykorzystywana do udostępniania obrazów wirtualnej bronchoskopii w celu wspomagania zabiegów igłowej biopsji transbronchialnej w ramach projektu badawczego realizowanego przez II Katedrę Chorób Wewnętrznych i Katedrę Radiologii Collegium Medicum Uniwersytetu Jagiellońskiego.

4. Podsumowanie

Zaprezentowany w tej pracy program został opracowany w celu zapewnienia interaktywnego środowiska wizualizacyjnego dla pulmonologii interwencyjnej. Dane pacjenta uzyskane w wyniku badania metodą tomografii komputerowej i zapisane w standardzie DICOM mogą być następnie przeglądane przed lub w trakcie zabiegów bronchofiberoskopowych. Potencjalne korzyści wynikające z zastosowania tego typu programu to lepsza orientacja w warunkach anatomicznych u indywidualnego pacjenta oraz większe bezpieczeństwo technik pobierania materiału do badania histopatologicznego, takich jak transbronchialna igłowa biopsja aspiracyjna.

Literatura cytowana

1. Boiselle P. M., Ernst A.: Recent Advances in Central Airway Imaging, 121(5): 1651-1660, Chest 2002.
2. DICOM home page: http://medical.nema.org/
3. Gay P., Brutinel W.: Transbronchial needle aspiration in the practice of bronchoscopy, Mayo Clin Proc, 64: 158-162, 1989
4. Haponik E. F., Cappellari J. O., Chin R., et al.: Education and experience improve transbronchial needle aspiration performance, Am J Respir Crit Care Med, 151: 1998-2002, 1995.
5. Schroeder W., Martin K., Lorensen B.: The Visualization Toolkit, an object-oriented approach to 3D Graphics, Prentice Hall PTR 1996.
6. Socha M., Duplaga M., Turcza P.: Methods of Bronchial Tree Reconstruction and Camera Distortion Corrections for Virtual Endoscopic Environments, Studies in Health Technology and Informatics vol.105, Transformation of Health Care with Information Technologies, IOS Press 2004.
7. VTK – The Visualizarion ToolKit, www.vtk.org
8. Wang K.: Continued Efforts To Improve the Sensitivity of Transbronchial Needle Aspiration, Chest, 114(1): 4-5, 1998.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 197-204.

STABILNOŚĆ ENDOPROTEZ CEMENTOWYCH STAWU BIODROWEGO W ŚWIETLE BADAŃ HISTOPATOLOGICZNYCH, NUMERYCZNYCH I EKSPERYMENTALNYCH

STABILITY OF THE HIP JOINT CEMENT ENDOPROSTHESIS IN HISTOPATHOLOGICAL, NUMERICAL AND EXPERIMENTAL TESTS

JERZY WŁODARSKI

Zakład Inżynierii Biomedycznej, Politechnika Częstochowska, Częstochowa, wlodarski@iop.pcz.czest.pl

Streszczenie. Podczas operacji wymiany obluzowanych trzpieni endoprotez stawu biodrowego w Klinice Ortopedii UM w Łodzi pobierano materiał do badań histopatologicznych z okolicy pseudotorebki endoprotezy i okolicy ziarniny otaczającej wymieniany trzpień implantu. Preparaty oceniono w mikroskopie świetlnym i elektronowym, zwracając uwagę na charakter tkanki łącznej, obecność martwicy, martwiaków, liczbę makrofagów oraz ciał typu obcego. Stwierdzono, że oprócz czynników mechanicznych w procesie obluzowania współuczestniczy mechanizm immunologiczny. Dokonano konfrontacji uzyskanych wyników z wynikami badań numerycznych i doświadczalnych.

Słowa kluczowe: endoproteza, strefa graniczna, preparaty

Abstract. During operations of exchange of loose endoprosthetic mandrels In the Orthopedic Teaching Hospital of Medical University in Łódź the material for histopathology tests has been collected from the region of endoprosthesis articular pseudo capsule and the region of granular tissue around the exchanged mandrels. The samples have been examined with both light and electron microscope regarding the connective tissue, presence of necrosis, number of macrophage and foreign bodies. It has been stated that the endoprosthesis loosening process is caused not only by the mechanical factors but by the immunologic mechanism as well. Numerical and experimental verification of the obtained results has also been performed.

Key words: endoprosthesis, contact area, abacterials

1. Wstęp

Implantacja sztucznego stawu biodrowego wiąże się z wprowadzeniem do środowiska wewnętrznego człowieka ciał biologicznie obcych z założeniem, że spełniając swoją funkcję, będą biologicznie obojętne.

W przypadku endoprotezoplastyki całkowitej wprowadza się do organizmu chorego staw metalowy z głową i panewką (z tworzywa sztucznego), które to obie części są osadzone w kości za pomocą cementu kostnego.

Są to jednak implanty statyczne, które oddziaływują całą swą nieuszkodzoną powierzchnią na otaczające tkanki. Jednak te same materiały w wyniku wieloletniej eksploatacji stawu biodrowego i powstałych produktów zużycia w postaci drobnych cząstek lub opiłek będą reagować z żywymi tkankami na znacznie większym czynnym polu powierzchni [2], [7].

W wyniku tarcia pomiędzy metalową głową i polietylenową panewką uwalniają się drobne opiłki metalu i tworzywa sztucznego oraz odrywają się stopniowo drobne cząsteczki cementu kostnego. Liczba tych produktów zużycia stale się zwiększa i bardzo się nasila przy obluzowywaniu endoprotezy [9], [10].

Procesy wzajemnej tolerancji tkanek z różnymi rodzajami protez są problemami ciągle nurtującymi tak transplantologów, jak i wytwórców materiałów użytych do ich produkcji [3], [5], [6], [7], [13], [19].

Przyczyną procesów powodujących aseptyczne obluzowanie endoprotezy jest szereg ciągle jeszcze nie w pełni poznanych czynników, w tym produkty zużycia [3], [7], [8], [12], [15], [16]. Poznanie tych procesów może stanowić podstawę stosowania różnych metod zapobiegających rozwojowi niekorzystnych zjawisk w endoprotezoplastyce stawu biodrowego.

Przeprowadzone badania mikroskopowo świetlne i elektronowe miały na celu ocenę zjawisk morfologicznych zachodzących w strefie granicznej łączącej implant z kością. Uzyskane wyniki skonfrontowano z wynikami badań numerycznych i eksperymentalnych.

Materiał do badań pobrano od 12 chorych różnej płci, wiek wahał się w granicach od 46 do 68 lat, ze średnią wieku 62 lata. Tkanki pobierano podczas zabiegu wymiany endoprotezy z powodu jej obluzowania. W celu prześledzenia zmian zachodzących w obszarach styku kość-cement-implant pobierano tkanki z 7 stref wg Gruena [20].

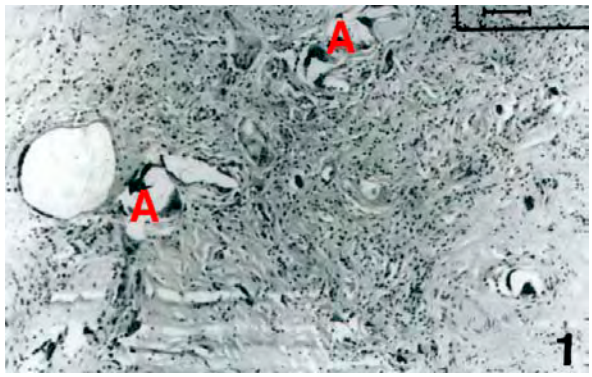
W strefie 4 materiał pobierano w odległości około 5 mm poniżej końca endoprotezy. Dodatkowo pobierano tkanki z okolicy panewki.

2. Wyniki badań mikroskopowo-świetlnych

We wszystkich badanych preparatach w zachowanych szczelinach jamy szpikowej stwierdzono utkanie szpiku żółtego, w którym układ komórek szeregu czerwono- i białokrwinkowego był prawidłowo zachowany.

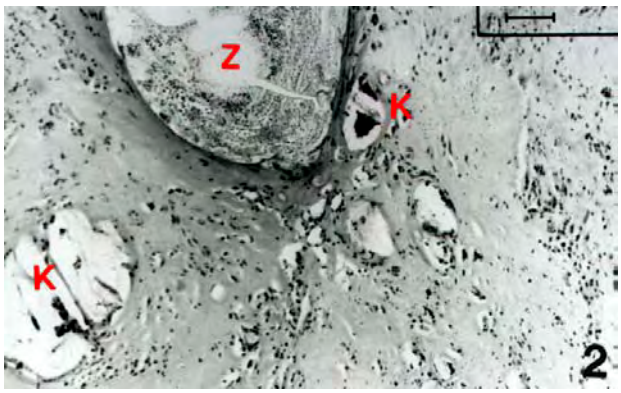
Większość badanych miała zmniejszoną liczbę megakriocytów w obrębie szpiku. Bezpośrednio do części gąbczastej kości przylegały odcinkowo bezpostaciowe kwasochłonne masy bądź też pasmo tkanki łącznej.

W tej ostatniej obok pól szkliwienia i dość masywnej kolagenizacji stwierdza się obszary utworzone przez nieswoistą ziarninę. Utworzona jest ona z komórek wrzecionowatych typu fibroblasta oraz proliferujących angioblastów, a ponadto widoczne są w niej bardziej liczne limfocyty i komórki plazmatyczne oraz miejscami bardzo bogato reprezentowane komórki olbrzymie typu ciała obcego (rys. 1).



Rys. 1. Obraz mikroskopowy tkanki okołoprotezowej z licznymi ziarniniakami typu ciała obcego (A). Barw. H+E. Pow. 75x

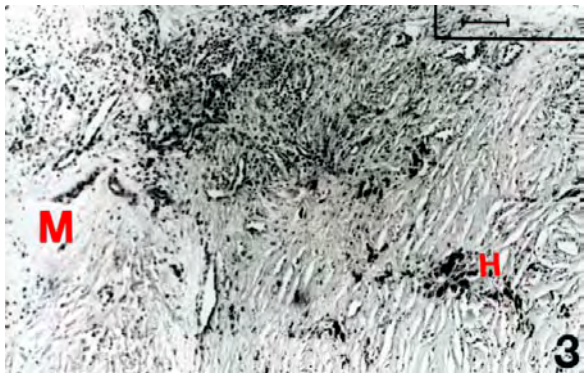
Te ostatnie były obserwowane z odmienną częstotliwością u poszczególnych chorych. Czasem towarzyszyły one ziarnistym złogom ciała obcego, a niekiedy spostrzegano w ich cytoplaźmie twory krystaloidne (rys. 2). Ponadto w obrębie tak uformowanej ziarniny spostrzegano dość licznie rozsiane ziarenka hemosyderyny (rys. 3).



Rys. 2. W strefie ziarniniaków olbrzymiokomórkowych widoczne są struktury krystaloidne (K) i ziarniste złogi (Z). Barw. H+E. Pow. 100x

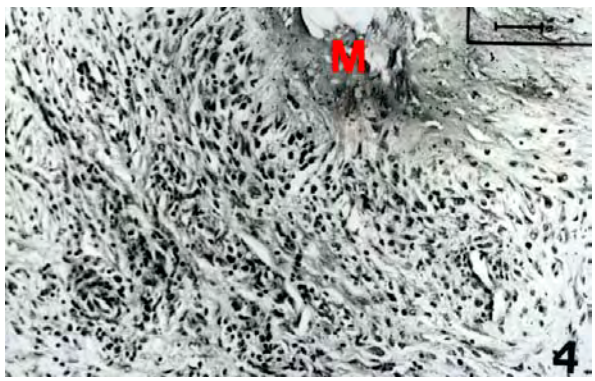
Dodatkowo u niektórych chorych widoczne były również w tym obszarze pola martwicy, chociaż odczyn komórkowy dookoła nich był śladowy.

Nierzadko w obszarach idących w głąb kości udowej widoczne były zgrupowania osteoklastów oraz komórek wrzecionowatych typu osteoblasta.

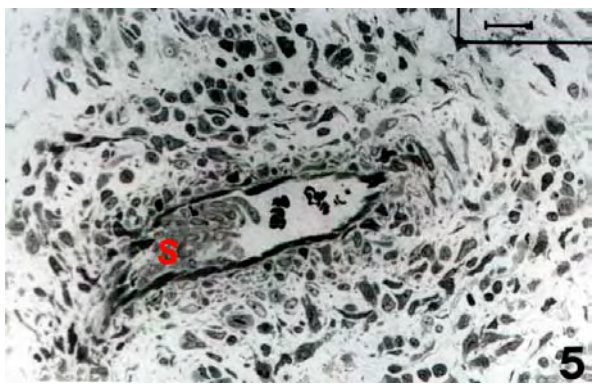


Rys. 3. Obraz mikroskopowy tej samej strefy okołoprotezowej. Różnej wielkości pola martwicy bez odczynu komórkowego (M) oraz różnie reprezentowane złogi hemosyderyny (H). Barw. H+E. Pow. 75x

W preparatach mikroskopowych pochodzących z tkanki pobranej z dna osadzenia endoprotezy stwierdzono różnie nasilone pola martwicy otoczone przez obszary tkanki ziarninowej (rys. 4). Skład ziarniny był identyczny jak ten opisany w strefach bocznych endoprotezy, jedynie waskularyzacja tej tkanki bardziej była nasilona. W samych zaś naczyniach spostrzegano proliferację śródbłonków w obrębie drobnych kapilarów (rys. 5).

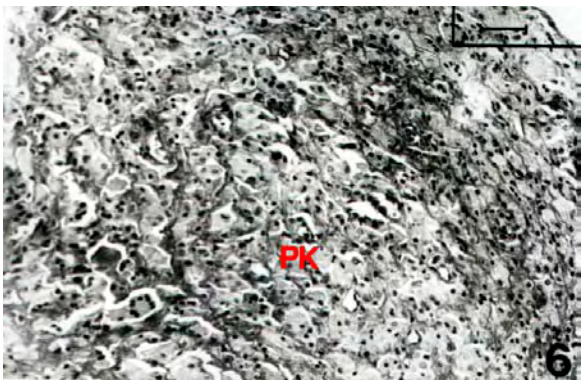


Rys. 4. Obszar dna protezy kości udowej. Pola martwicy (M) otoczone tkanką ziarninową dobrze unaczynioną. Barw. H+E. Pow. 150x

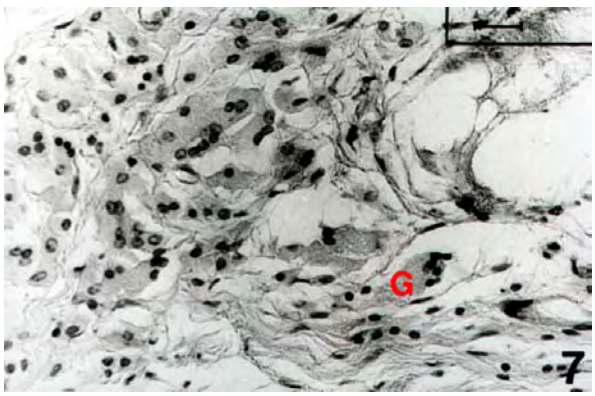


Rys. 5. Obraz tej samej strefy dna endoprotezy. Wyraźnie podkreślona proliferacja elementów śródbłonkowych (S) w nowo uformowanych naczyniach. Preparat półcienki. Pow. 400x

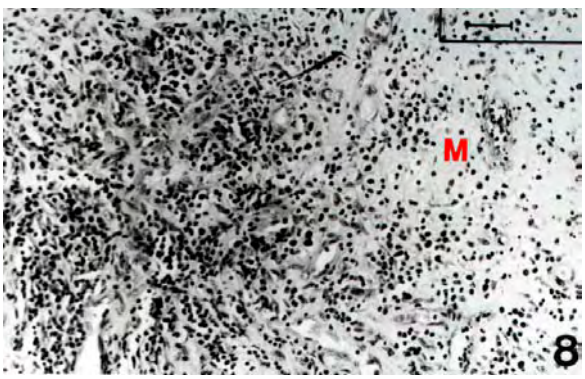
W tkankach uzyskanych z panewki stawowej stwierdzono obok fragmentaryzacji i rozpadu struktur kostnych różnej wielkości obszary nieswoistej ziarniny zbudowanej ze skupisk komórek olbrzymich i ksantomatycznych, tworzących czasami lite pola (rys. 6). Niekiedy obszary tych komórek były porozdzielane przez pasma wzajemnie przeplatających się włókien kolagenowych, a w innych przez mniej lub bardziej rozległe skupiska ziarniaków typu ciała obcego (rys. 7). W komórkach olbrzymich, w ich cytoplazmie widoczne były szczeliny po strukturach krystaloidnych, bądź też złogi ziarniste. Tylko w pojedynczych miejscach obserwowano niewielki naciek zapalny, składający się z limfocytów i komórek plazmatycznych (rys. 8).



Rys. 6. Rozległe pola utworzone przez komórki ksantomatyczne (PK) otoczone niewielkimi skupiskami limfocytów. Barw. H+E. Pow. 150x



Rys. 7. Komórki ksantomatyczne przeplatane pasmami tkanki łącznej, wśród których obecne są komórki olbrzymie (G) z pustą, wypłukaną cytoplazmą. Barw. Odczyn PAS wg McManusa. Pow. 400x



Rys. 8. Pola martwicy (M) objęte głównie naciekiem zapalnym z komórek plazmatycznych i limfocytów. Barw. H+E. Pow. 150x.

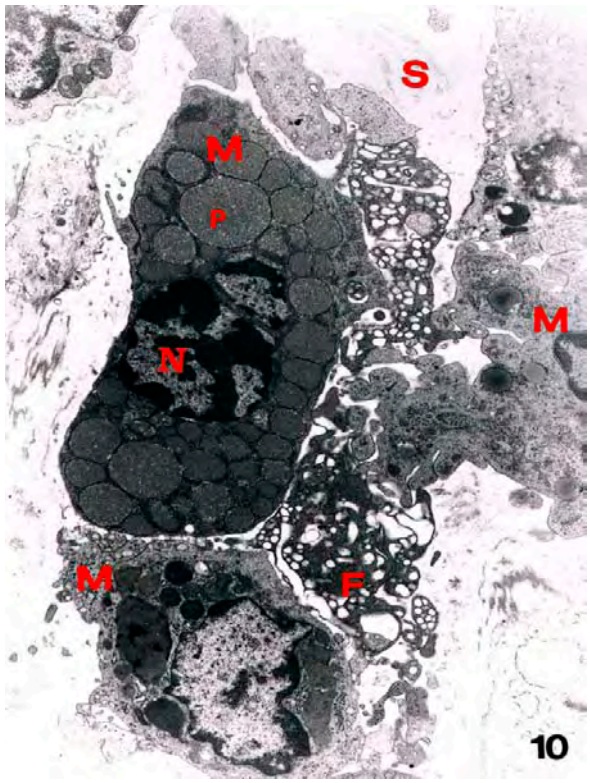
W pojedynczych przypadkach obserwowano duże pola osteolizy i destrukcji substancji podstawowej tkanki łącznej z tworzeniem się rzekomo-brodawkowatych struktur pokrytych komórkami maziówki, a w ich wnętrzu widoczne były mniej lub bardziej nasilone ogniska zwapnienia.

3. Wyniki badań mikroskopowo-elektronowych

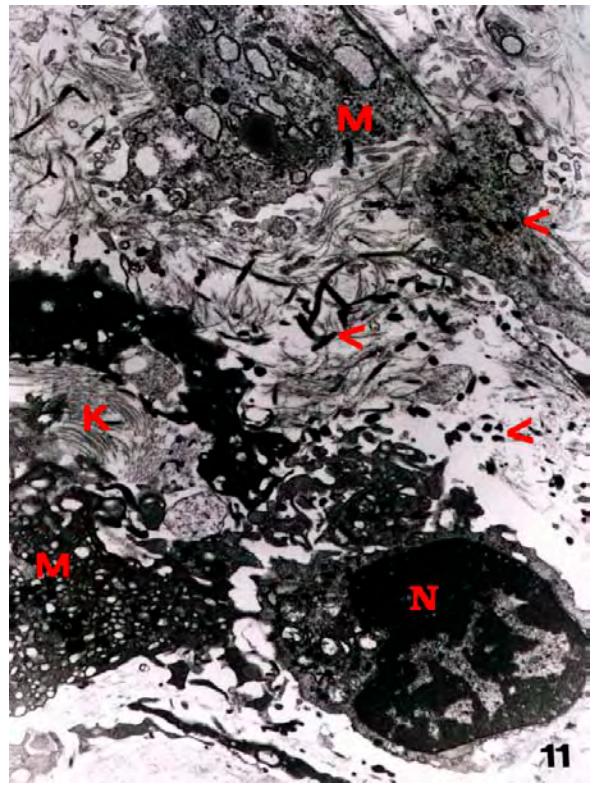
W preparatach mikroskopowo-elektronowych pochodzących z tkanek miękkich panewki ujawniono różnie wyrażone cechy aktywnej fagocytozy, w której makrofagi i pozostałe komórki mezenchymalne wykazują różnie nasilone oznaki zwyrodnienia (rys. 9).

Rys. 9. Fibroblasty (F) o znacznie rozszerzonych kanałach siatki śródplazmatycznej i wyraźnie euchromatynowym jądrem (J) z aktywnym jąderkiem. Makrofagi (M) zawierają liczne różnie wysycone elektronowo autofagozomy. Osteoblasty (O) o rozbudowanej siatce śródplazmatycznej otoczone przez pojedyncze limfocyty (L) i ciała makrofagów. Pow. 3000x

Makrofagi w swojej cytoplazmie zawierają dość licznie reprezentowane autofagozomy, zawierające substancje lipidowe o dużej gęstości elektronowej oraz fragmenty rozpadłych organelli. Niekiedy prawie całą cytoplazmę wypełniały struktury pęcherzykowate otoczone dość wyraźnymi błonami, zawierające w swoim świetle ziarniste, elektronowo gęste złogi (rys. 10).



Rys. 10. W cytoplazmie histiocytów (M) liczne pęcherzyki (P) obładowane elektronowo gęstym materiałem. W innych histiocytach liczba pęcherzyków różnie zaakcentowana. Fibroblasty (F) o znacznie rozbudowanej siatce śródplazmatycznej, której kanały są poszerzone. Przestrzeń międzykomórkowa zawiera niewielkie skupiska rozfragmentowanego kolagenu. Pow. 3000x



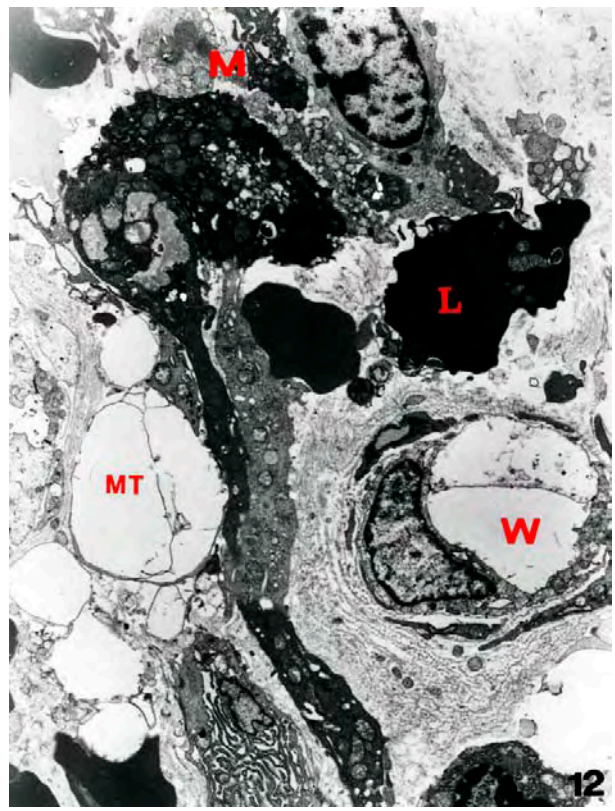
Rys. 11. W obszarze międzykomórkowym z licznymi włóknami kalagenu (K) i w cytoplazmie pojedynczych makrofagów (M) widoczne struktury krystaloidne (V) oraz znacznie poszerzone kanały siatki śródplazmatycznej. Jądra komórek plazmatycznych (J) o grubych złogach heterochromatyny. Pow. 3000x

Znajdowały się one często w sąsiedztwie fibrocytów i fibroblastów, wykazujących oznaki poszerzenia kanałów siatki śródplazmatycznej i osteoblastów o dość licznych wolnych rybosomach w cytoplazmie. W innych obszarach, między strukturami makrofagów zawierających zfagocytowane lipidy i oznaki destrukcji mitochondriów oraz poszerzenie kanałów siatki śródplazmatycznej obserwowano twory krystaloidne o dużej gęstości elektronowej. Znajdują się one między włóknami uformowanego kolagenu, a niekiedy i w cytoplazmie makrofagów (rys. 11).

W preparatach uzyskanych z miejsca styku trzpienia endoprotezy z kością udową stwierdzono dość żywą angiogenezę, a formujące drobne kapilary komórki śródbłonka objęte są różnie nasilonym zwyrodnieniem wodniczkowym. Złącza międzykomórkowe są prawidłowo utrzymane, a jedynie w komórkach śródbłonka ujawniono w cytoplazmie autofagozomy obładowane lipidami. Dookoła tak ukształtowanych naczyń widoczne są obok histocytów komórki plazmatyczne i limfocyty.

Nierzadko w tym obszarze spostrzegano mikrotorbielowate puste przestrzenie, które łączyły się dość często z błoną komórkową przylegającego makrofaga (rys. 12).

We wszystkich ocenianych preparatach nie wykazano osteoporozy w strefie styku protezy z kością, a jedynie różnie bogatą tkankę ziarninową. Stwierdzone między komórkami wykształcone włókna kolagenowe formowały dość chaotyczny układ.



Rys. 12. Nowo uformowane kapilary (W) ze śródbłonkiem wykazującym znaczne przewodnienie cytoplazmy. W przestrzeni międzykomórkowej pojedyncze złogi lipidowe (L). W przestrzeni międzykomórkowej uformowane mikrotorbielki (MT) ze struktur błoniastych, otoczone przez aktywnie fagocytujące makrofagi (M). Pow. 3000x

4. Omówienie i dyskusja

Przeprowadzone badania mikroskopowe wykazały, że u podłoża obluzowania endoprotezy leżą zjawiska odczynu morfologicznego na toczące się w tym obszarze procesy postępującego uprzątania ciał obcych o różnym stopniu oddziaływania na zjawiska proliferacji komórek mezenchymalnych. Wśród tych procesów można wyróżnić dwa zjawiska: jedno, o przypuszczalnie immunologicznym podłożu, wyrażające się obecnością różnie nasilonego nacieku zapalnego z komórek plazmatycznych i limfocytów. Drugie natomiast polega na formowaniu się w miejscach destrukcji kości przewlekłego zapalenia ziarniniakowego, w którym dominują komórki olbrzymie typu ciała obcego oraz makrofagi obładowane lipidami – komórki ksantomatyczne. Według Yanga i wsp. [19] w procesie obluzowania endoprotez szczególnie ważną rolę odgrywają procesy resorpcji tkanki kostnej w następstwie działania enzymów proteolitycznych, w tym głównie katepsyn lizozomalnych i metataloprotez. Są one uwalniane z makrofagów, w tym głównie z histiocytów i granulocytów obojętnochłonnych. Niektórzy autorzy sugerują nawet odstąpienie od ponownego wszczepiania endoprotezy, jeżeli w badaniu mikroskopowym w tkance okołoprotezowej stwierdza się już niewielką liczbę leukocytów [1].

Dotychczasowe badania szeregu autorów [4], [6], [7], [8], [12], [15], [16] wykazały, że wśród czynników mobilizujących odpowiedź mezenchymalną w strefie okołoprotezowej istotną rolę odgrywają substancje starte z powierzchni protezy, a więc cząsteczki metali, polietylenu i inne składowe cementu. Są one obok innych ważnym aktywatorem odpowiedzi komórkowej w miejscu styku protezy z tkankami właściciela. Innym dość ważnym czynnikiem aktywującym te procesy jest wzrost ciśnienia wewnątrzstawowego, które powoduje przepychanie się płynu do powstałych szczelin w miejscach osteolizy. Znajdujące się w tym płynie enzymy proteolityczne oraz inne mediatory zapalenia potęgują procesy niszczenia kości i rozwoju proliferacji mezenchymalnej [14], [16]. W tych odczynach proliferacyjnych wiodącą rolę odgrywają histiocyty, które obdarzone są szczególną właściwością fagocytarną, prowadzącą nierzadko do formowania ziarniniaków typu ciała obcego oraz obrazów zapalenia ksantomatycznego, co było szczególnie często obserwowane w naszym materiale w tkance pochodzącej z panewki stawu biodrowego. Na ten masywny rozwój odczynu ksantomatycznego mogą składać się dwa procesy.

Pierwszy to ciągła mobilizacja czynności fagocytarnej przez obecne w przestrzeni pozakomórkowej ciała obce z protezy, a drugi to aktywna fagocytoza głównie fosfolipidów błon biologicznych rozpadających się struktur komórkowych. Ta ciągła mobilizacja do transformacji komórkowej w strefie okołoprotezowej może mieć – co sugerują niektórzy autorzy [4], [11] – swój udział w możliwości rozwoju procesu nowotworowego, rozwijającego się w tej strefie u pojedynczych chorych. Jest to jednak problem nie w pełni udokumentowany, a pojedynczo opisane przypadki mogą być zjawiskiem incydentalnym. Również warte podkreślenia są obserwacje poczynione odnośnie obecności w wysięku zapalnym komórkowym w strefie okołoprotezowej limfocytów i komórek plazmatycznych. Stanowią one istotną odpowiedź komórkową tkanek na toczące się w tej strefie zjawiska immunologiczne, które mogą stanowić dodatkowy ważny element w poznaniu mechanizmów obluzowania wszczepionych protez.

Badania histopatologiczne materiału uzyskanego z pseudotorebki i tkanek otaczających obluzowany trzpień endoprotezy wykazywały charakterystyczne zmiany we wszystkich pobranych próbkach. Różnice polegały jedynie na liczbie makrofagów i ciał typu obcego. Tkanka łączna była w podobny sposób rozwinięta u każdego chorego, niezależnie od miejsca jej pobrania. W badaniach jednak nie stwierdzono zależności jej stopnia zorganizowania od: wieku pacjenta, płci, czasu obecności endoprotezy w kości.

Różne nasilenie martwicy i obecność martwiaków notowano w pseudotorebce, jak i w ziarninie z okolic udowej części implantu we wszystkich badanych przypadkach. Najbardziej istotną cechą badań histopatologicznych wydaje się być potwierdzenie podobnych do notowań w literaturze [2], [5], [11], [15] obecności makrofagów i ciał typu obcego, wykrytych w podobnych proporcjach we wszystkich przypadkach.

Ciała typu obcego były widoczne w trzech postaciach morfologicznych: jako ziarniste, pasmowate oraz sporadycznie widoczne (tylko w dwóch przypadkach) ciała czarnobarwiące, w piśmiennictwie opisywane jako cząstki metali. Drobiny te znajdowały się wewnątrz makrofagów, jak i na zewnątrz tych komórek.

Szczególnie duże ich nagromadzenie zauważono tam, gdzie radiologiczne objawy obluzowania endoprotezy były daleko zaawansowane.

Zmiany radiologiczne w postaci strefy osteolizy wokół trzpienia endoprotezy obserwowano w różnym nasileniu we wszystkich analizowanych przypadkach.

Największe zmiany osteolityczne były kolejno w VII, III, IV i V strefie wg Gruena. Nie zauważono żadnej korelacji pomiędzy stanem implantu a czasem pozostawania endoprotezy w organizmie, stopniem zmian histopatologicznych, jak i wieku chorych czy ich płci.

Willert i Semlitsch [17], [18] na podstawie swoich badań stwierdzili, że późną przyczyną niepowodzenia endoprotezoplastyki stawu biodrowego są substancje uwalniane z elementów implantu. Cząstki polietylenu, metalu i cementu są przyczyną reakcji komórek olbrzymich, które poprzez rozwój ziarniny zapalnej doprowadzają do obluzowania protezy. W badaniach histologicznych ziarniny uzyskanej podczas wymieniania endoprotezy zaobserwowali obecność makrofagów wielojądrzastych we wszystkich 34 przypadkach. Stwierdzili oni zależność liczby makrofagów od stopnia zaawansowania procesów osteolizy wokół usuniętej endoprotezy.

Podobne wyniki badań ogłosili De Lee i Maloney [13]. Dzisiaj wielu autorów zgodnie potwierdza stałe i charakterystyczne zmiany mikroskopowe ziarniny pobranej śródoperacyjnie podczas wymiany endoprotezy.

Zaobserwowano również, że zwiększeniu obszaru osteolizy kości otaczającej implant odpowiadał wzrost liczby histiocytów i makrofagów w badaniu histopatologicznym.

Ogólnie rzecz biorąc, przyjmuje się, że generowanie cząstek pochodzących ze zużywającej się endoprotezy i ich migracja do jamy maziowej i przestrzeni okołostawowej stymuluje pojawienie się oraz rozmnażanie makrofagów, które odpowiedzialne są za procesy osteolizy tkanki kostnej wokół implantu [20].

Badania, które przeprowadzono w WAM na własnym materiale wykazały podobne proporcje liczby makrofagów i ciał typu obcego w pseudotorebce oraz ziarniny z okolicy trzpienia.

Odpowiedź na pytanie, jaką drogą dostają się cząstki do miejsc, w których być nie powinny, próbuje dać Maloney [16]. Podaje on, że polietylen dostaje się do okolicy trzpienia poprzez pseudotorebkę, działającą jako kanał dla tych odpadów. Innego zdania są Willert i Semlitsch twierdząc, że cząstki te przemieszczają się drogą okołonaczyniowego systemu limfatycznego [18].

5. Symulacja numeryczna a wyniki eksperymentu i badań histopatologicznych

Mechanizm powstawania defektów w alloplastyce stawu biodrowego jest nadal z punktu widzenia mechaniki słabo rozeznany.

Badania numeryczne poparte eksperymentem na modelach fizycznych (preparatach) oraz badania histopatologiczne tkanek i analiza kliniczna mogą tworzyć podstawy do polioptymalizacji zabiegu alloplastyki. Praktyka kliniczna wykazała skłonność do utraty stabilności trzpieni. Nie wyjaśnia jednak znaczenia wielkości oraz rozkładu odkształceń i naprężeń występujących w układzie „kość-cement-implant" („kość-implant") pod wpływem obciążeń na osiągnięcie, a następnie utrzymanie stabilności trzpienia po implantacji.

Istotne znaczenie ma tutaj zachowanie ciągłości odkształceń i naprężeń oraz zapobieganie ich koncentracji.

Aseptyczne obluzowanie jest procesem stopniowym, prowadzącym do zaburzenia mechanicznej integralności wszczepu z kością. Jednocześnie na granicy fazowej między materiałem trzpienia a tkanką kostną tworzy się tkanka włóknista [21]. Jej stopniowy przyrost jest z czasem główną przyczyną wszystkich komplikacji.

Dla konfrontacji wyników symulacji numerycznej obciążenia stawu biodrowego z rezultatami przeprowadzonych badań doświadczalnych [21] i badań histopatologicznych tkanek pobranych podczas wymiany obluzowanej endoprotezy opracowano zestawienie (tabela 1).

Z zestawienia tego wynika, że dla analizowanego układu „kość-cement-implant" w górnej części endoprotezy występują większe naprężenia zredukowane, w przypadku zaś kości największe naprężenia zredukowane występują w dolnej części. Kość jest zatem w górnej części odciążona, obciążenie przenosi endoproteza.

W przypadku cementu kostnego strefa wyższych naprężeń zredukowanych występuje wzdłuż powierzchni przylegania do endoprotezy, a w dolnej części w całym przekroju płaszcza cementowego.

W badaniach doświadczalnych obciążenia układu „kość-cement-implant" na symulatorze, w obrazach radiologicznych zwiększony obszar radioluscencji tkanki kostnej i zmiany odpowiadające zmianom osteolitycznym kości żywej pojawiały się w górnej części połączenia (strefy 7 i 1) oraz stopniowo w dolnej części (strefy 3, 4 i 5). Taki obraz radiologiczny jest typowy dla obluzowanej endoprotezy.

Analizując przy tym obrazy histopatologiczne tkanek pobranych podczas wymiany endoprotezy z powodu jej obluzowania w strefach przylegania „kość-endoproteza", obserwowano makrofagi i komórki olbrzymie typu ciała obcego (polietylen, cement, metal). Taki obraz histopatologiczny jest typowy dla obluzowanej endoprotezy. Zwiększone pola martwicy i tkanka ziarninowa występowały także w dolnej części osadzonej endoprotezy [20].

Tab. 1. Zestawienie wyników badań numerycznych, badań doświadczalnych i histopatologicznych [21]

Strefy wg Gruena i wsp.	1	2	3	4	5	6	7
W stawie biodrowym [1] - endoproteza	wysoki	wysoki		niski		wysoki	wysoki
- w cemencie kostnym	wysoki	wysoki			wysoki		wysoki
- w kości		podwyższony			wysoki	podwyższony	
Poziom naprężeń zredukowanych							
Badania radiologiczne [2]	Zwiększony obszar radioluscencji tkanki kostnej		Zmiany odpowiada-jące zmianom osteolity-cznym w kości żywej	Zwiększony obszar fadiolu-scen-cji tkanki kostnej	Zmiany odpowiada-jące zmianom osteolity-cznym w kości żywej		Zmiany odpowiadające zmianom osteolity-cznym w kości żywej
Badania histopatologiczne [3]	Makrofagi + komórki olbrzymie typu ciała obcego	Komórki olbrzymie, ziarnina rys. 1	Makrofagi + komórki olbrzymie typu ciała obcego	Pola martwicy + tkanka ziarninowa rys. 4		Komórki olbrzymie, ziarnia, tkanka łączna rys. 3	Makrofagi + komórki olbrzymie typu ciała obcego rys. 7

[1] obciążenia dla t = 1 s,
[2] oceniano rozmieszczenie cementu kostnego, jego brak lub widoczne pęknięcia, jak też stopień przejaśnień wokół implantu,
[3] oceniano tkankę pobraną od 12-tu chorych, podczas wymiany endoprotezy z powodu jej obluzowania i oceniano w mikroskopie świetlnym i elektronowym.

Wysokiemu poziomowi naprężeń w endoprotezie towarzyszą obszary radioluscencji tkanki kostnej w obszarze radiologicznym i charakterystyczny wynik badania histopato-

logicznego (mikrofagi, ciała typu obcego, tkanka łączna, ziarnina, komórki olbrzymie).

Należy podkreślić, że analiza numeryczna dotyczyła określonego typu endoprotezy, a mianowicie zmodyfikowanego modelu Wellera 2, tj. endoprotezy TT-1, którą stosowano także w badaniach doświadczalnych na symulatorze [21].

Natomiast spośród analizowanych 12-tu przypadków wymiany endoprotezy u chorych, 7 dotyczyło ww. endoprotezy, pozostałe zaś były innych typów. Jednak prezentowane wyniki badań histopatologicznych były typowe dla wszystkich przypadków obluzowania endoprotez.

Objawem fizykalnym niskiego poziomu naprężeń zredukowanych w górnej części kości jest resorpcja kostna (stosownie do zasad przedstawionych przez Wolfa – *„struktura trabekularna tkanki kostnej w warunkach równowagi dostosowuje się do kierunków naprężeń głównych”*). Jest to związane z tym, że współczesne endoprotezy są sztywniejsze od kości korowej i dlatego indukują „*stress shielding*”, pozbawiając pewne obszary tkanki kostnej właściwych naprężeń, będących mechanicznym stymulatorem adaptacji kości do zmieniających się warunków środowiskowych [21].

„*Stress shielding*” doprowadza do ubytku tkanki kostnej bliższego odcinka kości udowej. Cząsteczki polietylenu (z panewki), wytworzone wskutek procesów tarcia, przedostają się do powstałych mikroprzestrzeni i inicjują obluzowanie trzpienia. Obluzowany implant na skutek mikroruchów staje się elementem „szlifującym” tkankę kostną i potęguje jej ubytek. Występujący w połączeniu „kość-cement-implant” cement kostny jest ogniwem najsłabszym i ulega pękaniu i wykruszaniu. Prowadzi to do zaniku cementu kostnego, który jest procesem niebezpiecznym, ponieważ może prowadzić do złamania kości, implantu.

Z pewnością poza resorpcyjnym wpływem „*stress shielding*”, inne czynniki muszą determinować ilość pozostałej tkanki kostnej, ponieważ w przeciwnym wypadku można by się spodziewać jej całkowitego zaniku!

Harris [21] uważa, że nie można tłumaczyć wszystkich reakcji wyłącznie prawami mechaniki. Stan tkanki kostnej wokół endoprotezy jest uwarunkowany wieloma czynnikami, być może natury genetycznej, metabolicznej, endokrynologicznej czy jej lokalnymi właściwościami osobniczymi. Jednakże z przedstawionych rozważań wynika, że w totalnej alloplastyce stawu biodrowego nie sposób pominąć roli czynnika mechanicznego.

6. Wnioski

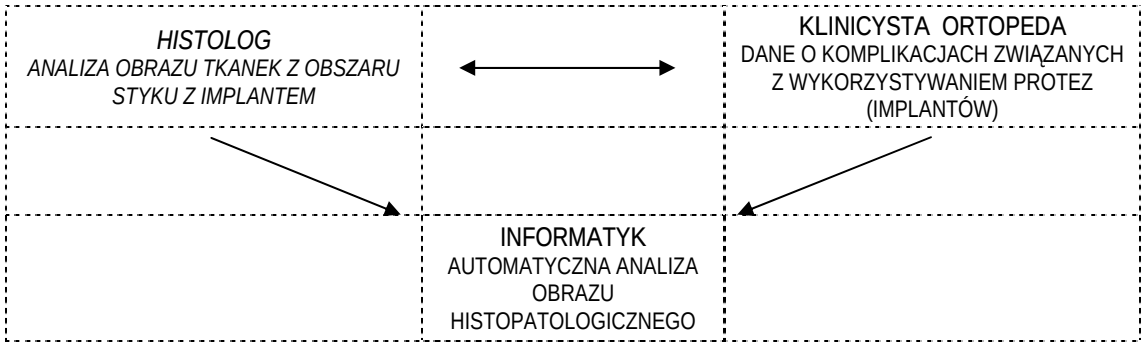
1. Implant, przenosząc obciążenie, powoduje odciążenie części bliższej odcinka kości udowej.
2. Wyższe wartości naprężeń w cemencie kostnym występują na powierzchni jego przylegania do endoprotezy.
3. Rozkłady naprężeń w elementach układów „kość-cement-implant” i „kość-implant” nie wykazują znaczących różnic, chociaż ze względu na odmienne warunki przenoszenia obciążenia nie można ich porównywać bezpośrednio.
4. Analiza numeryczna oraz badania eksperymentalne i histopatologiczne mogą stanowić przyczynek do poznania i wyjaśnienia niektórych problemów procesu aseptycznego obluzowania endoprotezy.

Literatura cytowana

1. Abdul Karim F. W., McGinnis M. G., Kraay M., Emancipator S. N., Golberg V.: Frozen section biopsy assessment for the presence of polimorphonuelear leukocytes in patients undergoing revisin of arthroplasties, Mod Pathol., 11(5): 427-431, 1998.
2. Albores - Saavadera J., Vuiteh F., Dekgado R., Wiley E., Hagler H.: Sinus histiocyrtosis of pelvis lymph nodes after hip replacement. A histiocytic proliferation inducted by cobalt-chromium and titanum, Am. J. Surg. Pathol., 18(1): 83-90, 1994.
3. Amsutz H. C., Grogoris P., Dorey F. J.: Evolution and future of surfaee replacement of the hip, J. Orthop Scj., 3(3): 169-186, 1998.
4. Bell R. S., Hopyan S., Davis A. M., Kandel R., Gross A. E.: Sarcoma of bone -cement membrane: a case report and review of the literature, Can J. Surg., 40: 51-55, 1997.
5. Devane P. A., Horne J. G., Martin K., Colgdham G., Krause B.: Tree-dimensional polyethylene wear of press-fit titanium prosthesis. Factor influencing generation of poliethylene debris, J. Arthroplasty, 12(3): 256-266, 1997.
6. Kadoya Y., Kobayasi A., Ohashi H.: Wear end osteolysis in total joint replacement, Acta orthiop. Scand, Suppl. 278: 1-16, 1998.
7. Kim K. J., Rubash H. E.: Large amounts of poliethylene debris in the interface tissue surrounding bipolar endoprotheses. Comparison on total hip prosthesis, J. Arthoplasty, 12(1): 32-39, 1997.
8. Lerouge S., Huk O., Yahia L., Witvoet J., Sedel L.: Ceramic-ceramic and metal poliethylen total hip replacement: comparison of pseudomembranes after loosening, J. Bone Joint Surg. Br. 79(1): 135-139, 1997.
9. Masshof i wsp.: Arch.Orthop. Unfall-Chir., t. 78, 1974.
10. Mirna i wsp.: Clin. Orthop., t. 117, 1976.
11. Murphy S. T., Parker R. B., Woodard J. C.: Osteosarcoma following total hip arthoplasty in a dog, J. Small Anim Pract., 38(6): 263-267, 1997.
12. Nizegorodcew T., Gasparini G., Maccauro G., Todesca A., De Santis E.: Massive osteolvsis inducted by high molecural weight poliethylene wear debris, Int. Orthop., 21(1): 14-18, 1997.
13. Okafor B., MacLellan G.: Postoperative changes of eryrtrocyte sedimentation rate, plasma viscosity and C reaktive protein levels after hip surgerv, Acta orthop. Belg., 64(1): 52-56, 1998.
14. Robertsson O., Wingstrand H., Kesteris U., Jonsson K., Onnerflat R.: Intracapsular pressure and loosening of hip prosthesis. Preoperative measurments in 18 hips. Acta orthop. Scand., 1997, 68(3): 231-234.
15. Stachowiak G. W., Campbell P.: Application of numerical descriptors to the charakterization of wear partieles obtained from joint repalcement, Proc. Inst. Mech. Eng H., 211(1): 1-10, 1997.
16. Szuszczewicz E. S., Schmalzried T. P., Petersen T. D.: Progressive bilateral pelvic osteolysis in patient with MeKee-Farrar metal-metal total hip prostheses. J. Arthroplasty, 12(7): 819-824, 1997.
17. Willert H. G. i wsp.: Osteolysis in alloarthroplasty of the hip. The role of ultra – high molecular weight polyethylene wear particles, Clinical Orthop., 258, 1990.
18. Willert H. G., Ludwig J., Semlitsch M.: Reaction of bone to methocrylate after hip arthroplasty. A long- term gross, light microscopic and scanning electron microscopic study, J. of Bone and Joint Surg., 56-A, 1974.
19. Yang R. S., Liu T. K.: Immunohistochemical analysis of matrix proteolityc enzymes in the periprosthetic tissue in the patients with loosening prostheses, Tahoku J. Exp. Med., 184(2): 99-111, 1998.

20. Włodarski J., Szram S.: Badanie zmian biologicznych i fizykochemicznych zachodzących w obszarze styku: kość-cement-implant w procesie destabilizacji endoprotezy, Acta of Bioingineering and Biomechanics, Vol. 2, Sup. 1, 601-606, 2000.

21. Włodarski J.: Analiza biologiczna i fizykochemiczna zmian zachodzących w obszarze styku kość-cement-implant w aseptycznym obluzowaniu endoprotezy cementowej stawu biodrowego, Przegląd Wojskowo-Medyczny, Sup. 4, Łódź, 1999.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 205-208.

COMPARISON OF DIALYSIS ADEQUACY INDICES CALCULATED ACCORDING TO ONE AND TWO COMPARTMENT MODEL

PORÓWNANIE WSPÓŁCZYNNIKÓW ADEKWATNOŚCI DIALIZY WYZNACZANYCH Z WYKORZYSTANIEM MODELI JEDNOKOMPARTMENTOWEGO I DWUKOMPARTMENTOWEGO

MAŁGORZATA DĘBOWSKA[*], JACEK WANIEWSKI[* **]

[*] *Instytut Biocybernetyki i Inżynierii Biomedycznej PAN, Warszawa, mdebowska@ibib.waw.pl*
[**] *Divisions of Baxter Novum and Renal Medicine, Department of Clinical Sciences, Karolinska Institutet, Stockholm, Sweden, jacekwan@ibib.waw.pl*

Abstract. Assessment of the dialysis efficiency is important, especially when different treatment modalities are applied or when a patient shifts from one modality to another (e.g. from peritoneal dialysis, PD, to hemodialysis, HD). The aim of this study was to derive formulas for dialysis adequacy indices, as fractional solute removal, FSR, and equivalent urea clearance, EKR, that are valid for HD and PD and for any patient characteristics, according to one compartment model and to reveal differences between one and two compartment calculations for wide range of dialysis parameters. It was shown, using one compartment mathematical model for urea distribution in the body that FSR is proportional to EKR and depends on three dimensionless parameters. Charts of the surfaces of FSR and EKR calculated according to one and two compartment model and drawn as a function of chosen dimensionless parameters showed that one compartment model is not sufficient for short-time, highly effective hemodialysis (relative error was 15% for FSR and 22% for EKR), but may be used for low effective dialysis, especially for PD, where for both indices relative error was up to 1.3%. FSR and EKR may be equivalently applied for the assessment of dialysis dose in hemodialysis and peritoneal dialysis, although the choice of model: one or two compartments, should depend on the values of parameters describing dialysis modality and patient.

Key words: urea kinetic model, KT/V, fractional solute removal, equivalent renal clearance

Streszczenie. Ocena wydajności dializy jest ważna, szczególnie, gdy są stosowane różne typy leczenia lub gdy pacjent jest kierowany z jednego typu dializy na inny (np. z dializy otrzewnowej, DO na hemodializę, HD). Celem niniejszego opracowania było wyprowadzenie wzorów dla współczynnika względnego usuwania substancji, FSR i równoważnego klierensu mocznika, EKR, które byłyby poprawne dla HD i DO i dla każdego typu pacjenta zgodnie z modelem jednokompartmentowym oraz wykazanie różnic między modelami jedno- i dwukompartmentowym dla szerokiego zakresu parametrów dializy. Używając modelu jednokompartmentowego dla opisu rozkładu substancji rozpuszczonych w ciele pacjenta, zostało pokazane, że wskaźnik FSR jest proporcjonalny do EKR i zależy od trzech bezwymiarowych parametrów. Wykresy powierzchni FSR i EKR, wyznaczane zgodnie z jedno- i dwukompartmentowym modelem i wykreślane jako funkcje wybranych bezwymiarowych parametrów pokazały, że modelowanie jednokompartmentowe nie jest wystarczające dla krótkotrwałych, wysoce efektywnych hemodializ (błąd względny wyniósł 15% dla FSR i 22% dla EKR), ale może być stosowane dla niskoefektywnych dializ, jak dializa otrzewnowa, gdzie dla obu wskaźników błąd względny był do 1.3%. FSR i EKR mogą być równoważnie stosowane do oceny dawki dializy, chociaż wybór modelu jednokompartmentowego lub dwukompartmentowego powinien zależeć od wartości parametrów opisujących dializę i pacjenta.

Słowa kluczowe: kinetyczny model mocznika, KT/V, współczynnik względnego usuwania substancji, równoważny klirens mocznika

1. Introduction

To assess dialysis efficacy a few different indices were introduced. The most popular index is KT/V (K – clearance of removal device, T – treatment time, V – solute distribution volume in the body). This index has also a different role: it is a non-dimensional parameter, which describes the kinetics of solute removal during dialysis. However, the solute removal depends also on other parameters, as for example, in metabolically stable patients, on the ratio of the treatment time to the total time of the cycle, i.e. time of treatment plus time of break between treatments, or, in peritoneal dialysis, on the ratio of dialysis fluid volume to the solute distribution volume (see below). Thus, alternative parameters of dialysis adequacy, fractional solute removal, FSR [4], [6], and equivalent urea clearance, EKR [1], [3], depend on more parameters than KT/V alone. Closed (analytical) formulas can be derived for FSR and EKR as functions of three non-dimensional parameters using a one compartment constant volume model for the solute distribution in the body. If the solute distribution has to be described by a two compartment model then some other parameters need to be taken into account, as the inter-compartment transport coefficient or sizes of compartments. The objective of our study was to compare the outcomes of urea kinetic modeling using a simple one compartment model and the closed formulas versus more sophisticated two compartment model solved by numerical simulations, as dependent on the three basic non-dimensional parameters.

The main difference between HD and PD is the fact that PD solute accumulates in dialysis fluid, whereas in HD fresh dialysate is delivered during whole treatment. Dialyzer clearance may attain the value of 400 mL/min [2], while the mass transport parameter of peritoneal membrane, MTAC, is up to 30 mL/min for urea [7] and therefore HD is more efficient than PD. For fast hemodialysis, usage of two compartment model for the simulation of the treatment and evaluation of dialysis indices is recommended.

2. Methods

Urea kinetic models. In compartment models, the patient's body can be considered as one compartment (thick line in Fig. 1) or as two compartments, intracellular and extracellular (broken line in Fig. 1).

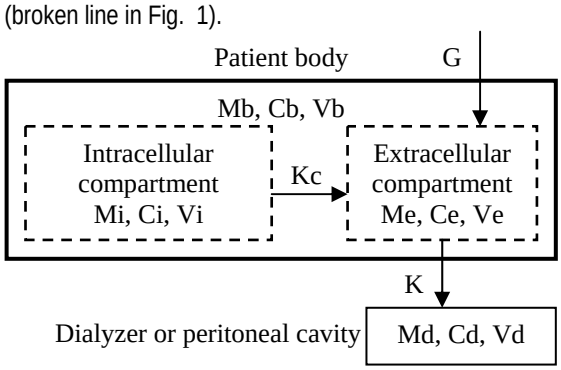


Fig. 1. Diagram of one compartment model (thick line) and two compartment model (broken line), for explanations of symbols – see Abbreviations

The one compartment urea kinetic model assumes that urea of mass Mb is distributed in the body in a single, homogenous pool of volume Vb, with concentration Cb. This model does not take into account the compartmental effects, characteristic for highly efficient dialysis. The two compartment model more precisely describes the function of the patient – dialyzer system, assuming that body fluids are divided into two parts: those directly and those indirectly accessible to dialyzer. Urea generation, at the rate G, occurs only in the extracellular space. The removal of urea with clearance K is a function of the urea concentration in the extracellular compartment, Ce, but indirectly depends on intercompartmental mass transport parameter Kc.

Assuming constant volume of solute distribution, no residual kidney function, and steady state of the patient, the one and the two compartment urea kinetic models can be described by the following systems of ordinary differential equations, respectively:

$$\begin{cases} \frac{dMb}{dt} = G - K(\frac{Mb}{Vb} - \frac{Md}{Vd}) \\ \frac{dMd}{dt} = K(\frac{Mb}{Vb} - \frac{Md}{Vd}) \end{cases}, \quad (1)$$

$$\begin{cases} \frac{dMe}{dt} = G + Kc(\frac{Mi}{Vi} - \frac{Me}{Ve}) - K(\frac{Me}{Ve} - \frac{Md}{Vd}) \\ \frac{dMi}{dt} = -Kc(\frac{Mi}{Vi} - \frac{Me}{Ve}) \\ \frac{dMd}{dt} = K(\frac{Me}{Ve} - \frac{Md}{Vd}) \end{cases} \quad (2)$$

Urea may accumulate in dialysis fluid of constant volume Vd as in peritoneal dialysis, PD, or fresh dialysis fluid without urea (Cd = 0) is continuously provided (formally Vd = ∞), as in hemodialysis, HD.

Dialysis adequacy indices. Fractional solute removal, FSR, is defined as the total amount of solute removed from the body, ΔM_R, normalized by the initial mass of the solute, Mb_0, [4], [6], [8]:

$$FSR = \frac{\Delta M_R}{Mb_0}, \quad (3)$$

where $\frac{dM_R}{dt} = K\frac{Mb}{Vb}$, (4)

according to one compartment model, and

$$\frac{dM_R}{dt} = K(\frac{Me}{Ve} - \frac{Md}{Vd}), \quad (5)$$

according to two compartment model. FSR is based on direct quantification and is a function of treatment parameters (clearance of dialyzer or diffusive mass transport parameter of peritoneal membrane, K, time of treatment, Tt, and the volume of dialysis fluid, Vd) and patient parameters (urea distribution volume, Vb, generation rate, G, and urea concentration before the treatment, Cb_0)

Equivalent urea clearance, EKR, is determined as the urea generation rate, G, over the time average concentration (TAC) [1], [3]:

$$EKR = \frac{G}{TAC}, \quad (6)$$

EKR is a hypothetical continuous clearance necessary to keep a target urea concentration equal to TAC.

Relative error. To evaluate differences between one (1C) and two compartment (2C) calculations of dialysis adequacy indices, the relative error of them can be calculated:

$$\text{Relative error}_X = \frac{|X_{1C} - X_{2C}|}{X_{2C}}, \quad (7)$$

Two compartment model is used as a standard reference one.

3. Results

Theoretical analysis. The ratio of FSR to EKR is:

$$\frac{FSR}{EKR} = \frac{Tc}{Vb}, \quad (8)$$

Equation (8) demonstrates that FSR and EKR are strictly correlated, and the coefficient of proportionality depends on solute distribution volume Vb during the treatment cycle.

Using one compartment model presented in Methods and the definition of FSR (3), the general formula of FSR as a function of three dimensionless parameters v, k and τ, for the treatment time, was derived as:

$$FSR = v\frac{1 - e^{-k}}{1 - v\tau}, \quad (9)$$

where

$$v = \frac{Vz}{Vb}, \quad (10)$$

$$Vz = \frac{Vb \cdot Vd}{Vb + Vd}, \tag{11}$$

$$k = \frac{K \cdot Tt}{Vz}, \tag{12}$$

$$\tau = \frac{Tt}{Tc}\left(1 - \frac{1 - e^{-k}}{k}\right). \tag{13}$$

Equations from (9) to (13) can be used for calculation of FSR in both hemodialysis and peritoneal dialysis.

Simulations. Typical treatment parameters were used: urea generation rate G = 7 mg/min, urea distribution volume Vb = 35 L; for the two compartment model it was assumed that the intercomparemental clearance is Kc = 600 mL/min and water volumes of extracellular and intracellular compartment are related to total body volume, Vb, Ve = 0.35Vb, Vi = 0.65Vb.

Defining the ratio of treatment time, Tt, to time of dialyses cycle, Tc, as θ:

$$\theta = \frac{Tt}{Tc}, \tag{14}$$

where cycle time is the time after that the patient parameters as concentration and distribution volume of urea, come back to their initial values, figures of FSR as a function of θ and k were drawn. Figure 2 presents FSR according to one (1C) and two compartment model (2C) in hemodialysis (Cd = 0) for Tt of range from 30 min to Tc = 56 h, and dialyzer clearance K, from 2 to 2300 mL/min. FSR according to one compartment model may be calculated using equation (9) or system of differential equations (1). FSR according to two compartment model was calculated directly from equations (2).

EKR for one compartment model may be also expressed by previously defied parameters:

$$EKR = \frac{2(-1 + e^{k})kVz}{Tc(v(2\theta(-2 + v\theta) + k(1 - 2\theta + v\theta^2)) + \ldots}, \tag{15}$$

$$\ldots + e^{k}(-2v\theta(-2 + v\theta) + k(2 - v - 2v\theta + v^2\theta^2)))$$

The surfaces of EKR according to one (1C) and two (2C) compartment model for Cd = 0 are shown in Fig. 3.

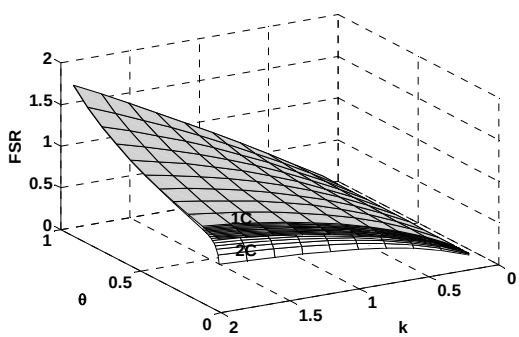


Fig. 2. FSR in hemodialysis according to one compartment model (1C) and two compartment model (2C) as a function of two non-dimensional parameters θ and k

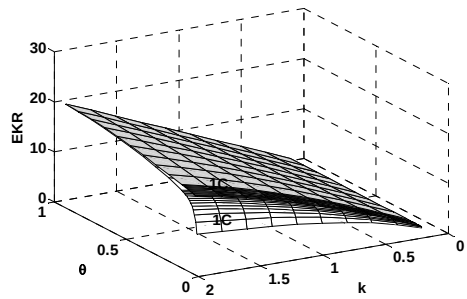


Fig. 3. EKR in hemodialysis according to one compartment model (1C) and two compartment model (2C) as a function of two non-dimensional parameters θ and k

The highest values of FSR and EKR were for $\theta \approx 1$, which corresponds to treatment time equal cycle time (continuous treatment) and $k \approx 2$ (kinetic KT/V equal to 2). The biggest discrepancies between one and two compartment model calculations were for short-time, high-efficient dialysis. Figure 4 presents relative errors for FSR and EKR calculated according to one compartment model in hemodialysis, assuming calculation according to two compartment model as a standard.

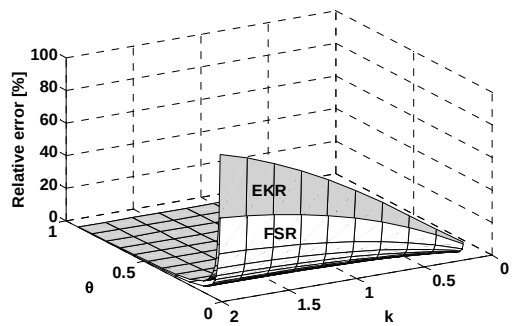


Fig. 4. Relative errors of FSR and EKR in hemodialysis, defined as absolute value between one and two compartment calculations over index value calculated according to two compartment model, equation (7)

Because clearance of value higher than 400 mL/min is not attainable in dialyzers, the relative errors of one compartment calculation for K ranged from 2 to 400 mL/min were examined and reached 15% for FSR and 22% for EKR.

An example of FSR surfaces for peritoneal dialysis is shown in Fig. 5. Figure 6 presents the relative error of FSR and EKR one compartment calculation for closed loop system as PD.

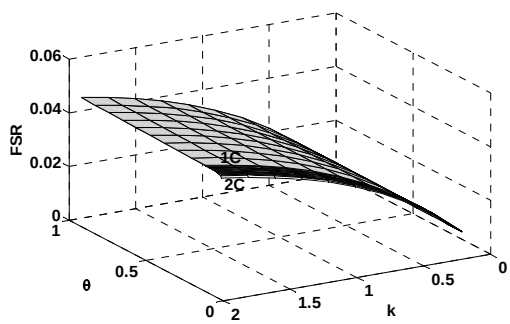


Fig. 5. Surfaces of FSR according to one compartment model (1C) and two compartment model (2C) for peritoneal dialysis

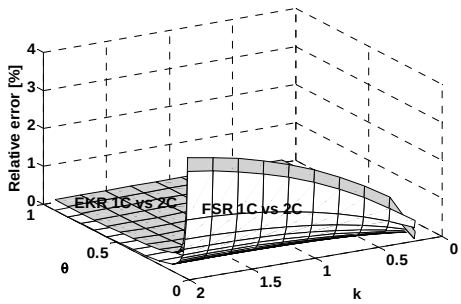


Fig. 6. Relative errors of FSR and EKR in peritoneal dialysis, defined as absolute value between one and two compartment calculations over index value calculated according two compartment model, equation (7)

For clearance meaning diffusive mass transport parameter of peritoneal membrane ranged from 0.1 to 30 mL/min the relative error of one compartment calculations was up to 1.3% for FSR and EKR.

Special cases of FSR and EKR. For specific conditions of different treatment modalities and time schedules of dialysis, the formulas of FSR and EKR occur in following forms:

1) No accumulation of the solute in dialysis fluid (i.e. fresh dialysis fluid is used at any moment of the treatment, $Vd \to \infty \;\Rightarrow\; Vz = Vb \;\Rightarrow\; v = 1$):

$$FSR = \frac{1-e^{-k}}{1-\tau} \tag{16}$$

$$EKR = \frac{2(-1+e^{k})kVb}{Tc((1+e^{k})k(-1+\theta)^{2} - \ldots} \;, \tag{17}$$

$$\ldots 2(-1+e^{k})(-2+\theta)\theta)$$

2) Short highly effective treatment ($Tt << Tc \Rightarrow \theta = 0$):

$$FSR = v(1-e^{-k}) \;, \tag{18}$$

$$EKR = \frac{2(-1+e^{k})Vz}{Tc(e^{k}(-2+v)-v)} \;, \tag{19}$$

3) Hemodialysis/hemofiltration (K is clearance of dialyzer/ hemofilter, $v = 1, Tt << Tc \Rightarrow \theta = 0$):

$$FSR = 1-e^{k} \,, \tag{20}$$

$$EKR = \frac{2(-1+e^{k})Vz}{(1+e^{k})Tc} \;, \tag{21}$$

4) Continuous treatment, $Tt = Tc \;\Rightarrow\; \theta = 1$, for example CAPD where K is the diffusive mass transport coefficient:

$$FSR = v\frac{1-e^{-k}}{1-v\tau} \;, \tag{22}$$

$$EKR = \frac{2(-1+e^{k})kVz}{Tc(e^{k}(k(-1+v)-2v)(-2+v)+\ldots} \;, \tag{23}$$

$$\ldots + (2(-2+v)+k(-1+v))v)$$

Using equations from (16) to (23), which are valid for indicated conditions one can easily calculate values of FSR and EKR according to one compartment, constant volume model.

4. Discussion

Compartmental models and solute kinetic analysis may be used for the description of prescribed and delivered dose of dialysis and provide support for assessment of dialysis efficiency.

The removed solute mass according to one compartment model is higher than according to two compartment model, because of the role of the compartment mass transfer (Kc), which precludes an immediate intercompartmental flow. Simplifications of one compartment model do not let this model be used for simulation of fast solute removal from the body, because it does not take into account the compartmental effects, characteristic for high efficient dialysis, as inbound and rebound [5]. The simulations were carried out assuming the constant value of Kc = 600 mL/min, for higher value of Kc the discrepancies between one and two compartment calculations are smaller.

Using a fixed dwell volume or recirculation of dialysis fluid decreases the overall efficiency of the treatment compared to continuous use of fresh dialysis fluid because of accumulation of urea in dialysis fluid, thus the values of FSR and EKR for peritoneal dialysis are much smaller than these for hemodialysis.

The relative errors of FSR and EKR calculated according to one compartment model with reference to two compartment calculations, for hemodialysis, were 15% and 22%, respectively. For peritoneal dialysis, the relative error for both indices was up to 1.3%. The biggest discrepancies between one and two compartment calculations occur for short-time (30 min), high-effective (clearance about 400 mL/min) hemodialysis, where compartmental effects are the most visible.

The choice between the dialysis adequacy indices is not obvious. In general, FSR is proportional to EKR and both can be equivalently applied for the description and comparison of dialysis efficacy for different time schedules and treatment modalities of dialysis, in contrast to KT/V, which has different meaning in hemodialysis and peritoneal dialysis [8].

Abbreviations

Patient characteristics:

Mb, Cb, Vb – solute mass, concentration and volume distribution volume of solute within the body, respectively,

Me, Ce, Ve – solute mass, concentration and volume of extracellular compartment, respectively,

Mi, Ci, Vi – solute mass, concentration and volume of intracellular compartment, respectively,

Kc – mass transfer parameter of peritoneal membrane,

G – urea generation rate.

Dialysis and dialyzer parameters:

Md, Cd, Vd – solute mass, concentration and volume of dialysis fluid, respectively,

Tt – treatment time; Tc – time of dialyses cycle ; K – dialyzer clearance.

References

1. Casino F. G., Lopez T.: The equivalent renal urea clearance: a new parameter to assess dialysis dose, Nephrol Dial Transplant, 11: 1574-1581, 1996.
2. Daugirdas J. T., Blake G., Ing T. S.: Handbook of Dialysis. Lippincott Williams & Wilkins, USA 2001.
3. Gotch F. A., Sargent J. A., Keen M. L.: Whither goest Kt/V? Kidney Int Suppl, 76: S3-18, 2000.
4. Henderson L. W.: Critical interpretation of adequacy parameters in peritoneal dialysis and hemodialysis, Perit Dial Int, 19 Suppl 2: S38-44, 1999.
5. Schneditz D., Daugirdas J T.: Compartment effects in hemodialysis, Semin Dial, 14: 271-277, 2001.
6. Verrina E., Brendolan A., Gusmano R., Ronco C.: Chronic renal replacement therapy in children: which index is best for adequacy? Kidney Int, 54: 1690-1696, 1998.
7. Waniewski J., Heimburger O., Werynski A., Park M. S., Lindholm B.: Diffusive and convective solute transport in peritoneal dialysis with glucose as an osmotic agent, Artif Organs, 19: 295-306, 1995.
8. Waniewski J., Lindholm B.: Fractional solute removal and KT/V in different modalities of renal replacement therapy, Blood Purif, 22: 367-376, 2004.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 209-212.

COMPARTMENTAL MODEL OF FLUID AND SOLUTE ABSORPTION IN PERITONEAL DIALYSIS

KOMPARTMENTOWY MODEL ABSORPCJI PŁYNU I SUBSTANCJI W DIALIZIE OTRZEWNOWEJ

MAGDA GAŁACH, ANDRZEJ WERYŃSKI

Instytut Biocybernetyki i Inżynierii Biomedycznej PAN, Warszawa, magda@ibib.waw.pl

Abstract. Solute and fluid transport during peritoneal dialysis can be predicted using various models developed in the past. They were very helpful in understanding the principles of peritoneal dialysis as well as in clinical patient treatment. However, these models have specific tasks or explain specific phenomena, for example membrane model is used for estimation of the peritoneal transport parameters and the three-pore model was very helpful in explaining sodium sieving, yet these models do not cover processes in the peritoneal interstitium. Also in three-pore model the fluid uptake is described inadequately by assuming that fluid absorption is by systemic lymphatics only. Numerous studies with macromolecular volume marker have shown that the fluid absorption rate is well above 1 ml/min, whereas rate of fluid absorption via lymphatics according to physiological investigations is approximately 0.3 ml/min. In this study we propose a theoretical model in which interstitium was included as a separate compartment, and beside the lymphatic absorption, the absorption due to Starling forces was taken into account. With the help of this model the discrepancy mentioned above could be explained.

Key words: mathematical model, lymphatics, absorption, interstitium, peritoneal transport

Streszczenie. W przeszłości powstało wiele modeli opisujących transport substancji oraz płynu podczas dializy otrzewnowej. Modele te były bardzo przydatne dla zrozumienia zasad działania dializy otrzewnowej. Wykorzystywane też były w trakcie leczenia pacjentów. Jednakże modele te mają określone zadania i opisują pewne wybrane aspekty dializy otrzewnowej. Na przykład model membranowy używany jest do estymacji parametrów opisujących transport płynu i substancji, a model trójporowy pomógł w wyjaśnieniu tzw. przesiewania sodu. Te modele jednakże nie opisują procesów zachodzących w tkance otrzewnej. Dodatkowo model trójporowy zakłada, że absorpcja płynu z przestrzeni otrzewnowej zachodzi jedynie w procesie absorpcji limfatycznej. Jednak wiele studiów z wykorzystaniem znacznika objętości pokazało, że szybkość absorpcji płynu wynosi ponad 1 ml/min, podczas gdy szybkość absorpcji limfatycznej wynosi jedynie 0.3 ml/min. W pracy proponujemy model, w którym tkanka otrzewnowa potraktowana jest jako dodatkowy kompartment. Dodatkowo oprócz absorpcji limfatycznej brana jest pod uwagę również absorpcja za pomocą sił Starlinga. Dzięki temu prezentowany model pozwala na wyjaśnienie wspomnianych wcześniej niezgodności.

Słowa kluczowe: model matematyczny, układ limfatyczny, absorpcja, śródmiąższe, transport otrzewnowy

1. Introduction

The mathematical models of the physiological systems are of great importance, because they can be used to improve our understanding of the behavior of these complex systems. Important parts of them are models describing functions of the kidney or renal replacement therapies.

Kidneys perform many functions. Healthy kidneys clean blood by removing excess fluid, minerals, metabolic waste products and toxins, help control blood pressure by releasing enzyme rennin, participate in making red blood cells and help maintain the proper balance of acid and minerals. When a kidney fails, one of renal replacement therapy can be a life-saving solution. The main types of this therapies, beside kidney transplantation, are hemodialysis and peritoneal dialysis.

Hemodialysis removes waste products, toxins and excess of fluid from the blood in an extracorporeal circuit in which blood flows through a filtering system called dialyzer and returning it, cleaned, to the body. During peritoneal dialysis the blood is cleaned inside the body, through the peritoneum (membrane which surrounds the outside of the organs in the abdomen). Dialysis fluid is infused into the peritoneal space, the peritoneum allows waste and fluid to pass from the blood into the dialysate (in the processes of diffusion and convection), and then dialysate is drained out.

Nowadays, peritoneal dialysis, in depending of the country, comprises from 5% to 45% of the population of renal failure patients. Physiological mechanisms of water and solute transport through the peritoneum are rather complex. Many mathematical models describing these mechanisms have been developed, and were very helpful in understanding the

principles of peritoneal dialysis as well as in assessment of the efficacy and adequacy of peritoneal dialysis. Two main models are applied for the evaluation of peritoneal transport: the membrane model [3] and the three-pore model [6].

Membrane model is based on thermodynamic principles of transport of fluid and solutes across semi permeable membrane. This model is rather simple and is predominanty used for the estimation of fluid and solutes transport parameters using experimental data.

The three-pore model is more complex and based on capillary physiology. The basic idea of this model is an assumption that transport of the solutes and fluid is through cylindrical pores inside walls of peritoneal blood capillaries. In spite of its complexity the three-pore model does not include processes in the interstitium.

2. Model

Recently, peritoneal microcirculation and transport through the peritoneum has become an important subject of various study [7], [8], [9].

The model that we propose in this study is partly based on the three-pore model, but one additional compartment (beside dialysate and blood) is taken into account.

The basic assumption of this model is the following idea: when the solute leaves the peritoneal cavity, it must first enter the interstitium to be then sieved by the capillary wall (Fig. 1).

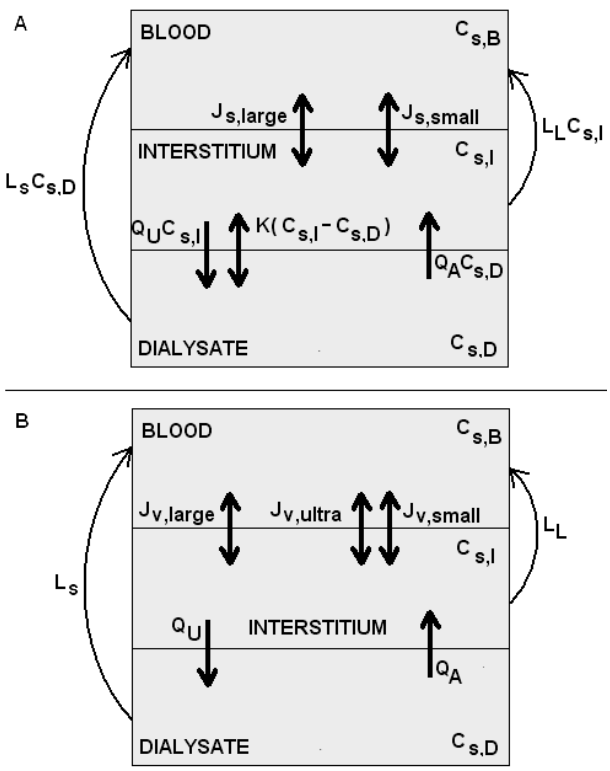


Fig. 1. Visualization of the compartmental model describing transport of the solutes (A) and fluid (B) between dialysate and blood during peritoneal dialysis

The solute flow between dialysate and interstitium is caused by diffusion, convection, fluid absorption due to Starling forces and systemic lymphatic absorption (Fig. 1a). Thus, the mass balance equation for dialysate can be written as follows:

$$\frac{dV_D C_{s,D}}{dt} = K_s(C_{s,I} - C_{s,D}) - Q_A C_{s,D} + S_s Q_U C_{s,I} + - L_s C_{s,D} \quad (1)$$

where: V_D – dialysate volume, K_s – permeability surface area coefficient, Q_A – fluid absorption rate due to Starling forces, Q_U – ultrafiltration flow rate, L_s – systemic lymphatic absorption rate, S_s – sieving coefficient of solute *s*, $C_{s,c}$ – concentration of solute *s* in compartment *c* (c: 'D' – dialysate, 'I' – interstitium, 'B' - blood).

As ahown in Fig. 1b, the dialysate volume depends basically on following factors: ultrafiltration flow rate, fluid absorption rate due to Starling forces and systemic lymphatic flow rate. Therefore:

$$\frac{dV_D}{dt} = Q_U - Q_A - L_s\,, \quad (2)$$

Change of the solute mass in the interstitium can be caused by solute flow between dialysate and interstitium and solute flow between interstitium and blood. The three-pore model of peritoneal transport has been used for description of the transport of fluid and solutes between interstitium and blood. It is assumed, that capillaries on the peritoneal membrane are heteroporous with three sizes of pores: large pores (radius 250 Ä), small pores (radius 43 Ä), and ultrasmall pores (radius 2 Ä). A large number of small pores makes the membrane permeable to most small solutes, whereas low number of large pores allows the transport of macromolecules from blood to peritoneal cavity. Ultrasmall pores (denoted aquaporins) reject solute transport, and play an important role in osmotic water transport. Thus:

$$\frac{dV_I C_{s,I}}{dt} = Q_A C_{s,D} - S_s Q_u C_{s,I} - K_s(C_{s,I} - C_{s,D}) + - L_L C_{s,I} + J_{s,small} + J_{s,large} \quad (3)$$

where: V_I – interstitial volume, L_L – local lymphatic absorption rate, $J_{s,p}$ - solute *s* flow through pores *p* according to three-pore model [10].

For simplification, it is assumed that interstitial volume accessible to the solute, V_I, is constant and equals 350 ml [11] and ultrafiltration flow rate, Q_u, is:

$$Q_u = ae^{-bt} + c \quad (4)$$

for certain parameters *a*, *b* and *c* (a, b > 0). Therefore, the equation (1) is dependent only on time and it is possible to find explicit solution of the equation (2).

Intraperitoneal dialysate volume during peritoneal dialysis is assessed using the macromolecular solute called volume marker. A macromolecular volume marker, usually radioisotopically labelled albumin (RISA) or dextran is added to fresh dialysis fluid. Dialysate volume is determined using the dilution principle for this volume marker.

Consideration about fluid and volume marker transport was performed with assumption that volume marker is transported to blood only by lymphatic absorption.

3. Methods

Presented mathematical model was implemented in Matlab. To evaluate the model, fluid transport was studied in single six – hour dwell study using standard glucose 3.86% dialysis fluid in patients on continuous ambulatory peritoneal dialysis in Sweden. In this paper, data of two patients were used. One patient with relatively low marker elimination coefficient (K_e = 0.75 ml/min, patient 1) and one with high marker elimination coefficient (K_e = 2.5 ml/min, patient 2)

Parameters of the model were taken from literature or were fitted to minimize differences between simulation results and experimental data (using relative error).

Tab. 1. The fluid and solute transport parameters used in computer simulations.

Parameters	Patient 1	Patient 2
V_I [ml]	350 ml**	
L_s [ml/min]	0.3 ml/min ***	
L_l [ml/min]	0.1096*	0.1108*
K [ml/min]	0.0017*	0.5139*
Q_a [ml/min]	0.5012*	1.0594*
a [ml/min]	15.5815*	17.6618*
b [1/min]	0.0189*	0.0154*
c [ml/min]	0.8168*	-0.7477*

* Parameters estimated with model; Parameters taken from literature: ** Struijk et al. [11], *** Rippe et al. [10]

3. Results

As shown in Figs. 2 – 5, there was a good agreement between simulations of the model and experimental data. Moreover, the estimated parameters of the fluid and solute transport (Tab. 1) coincided with the parameter values taken from literature.

With the help of the model presented it is possible to observe the behavior of the macromolecular volume marker during passage from peritoneum to blood.

A volume marker from peritoneal cavity moves to the interstitium and is then transported to the blood. Because of the high molecular weight of the marker it may be assumed that volume marker is transported to blood only via lymphatic absorption and therefore the has tendency to accumulate in the interstitium. It was confirmed by computer simulations of the model. In the patient with low marker elimination coefficient this accumulation is rather slow (Fig. 2), while in the patient with high elimination coefficient accumulation is well noticeable, and the concentration of the marker in interstitium exceeded concentration in dialysate (Fig. 4).

The model allows differentiating the nature of the fluid uptake: lymphatic absorption or absorption of the fluid due to Starling forces. In three-pore model [10] the fluid uptake is described inadequately by assuming that fluid absorption is by systemic lymphatics only. Physiological investigations show that a rate of fluid absorption via lymphatics is about 0.3 ml/min. However, various studies with volume marker have shown that the fluid absorption rate is well above 1 ml/min. This is confirmed further by the simulations of the model (Q_a, Tab. 1).

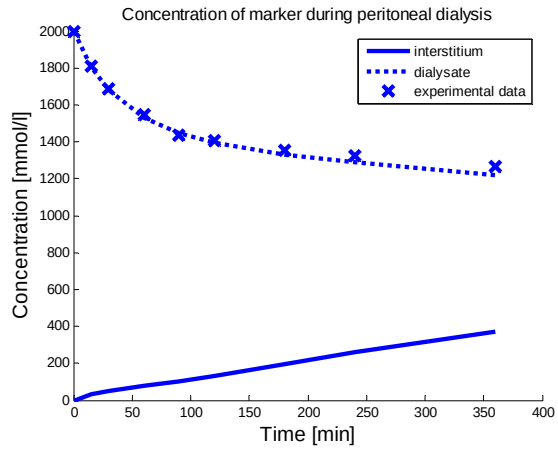


Fig. 2. Simulation of the marker concentrations during peritoneal dialysis for patient 1: — marker in interstitium, marker in dialysate, X – experimental data.

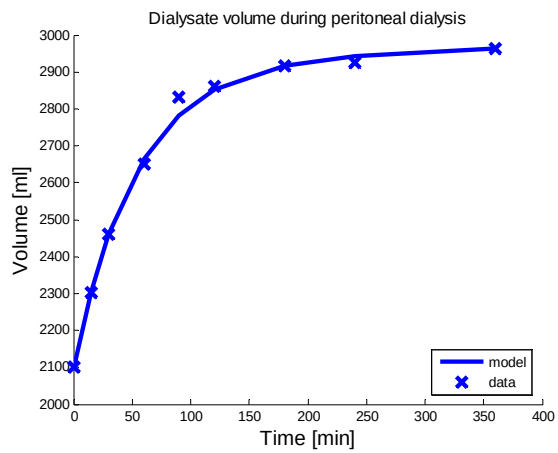


Fig. 3. Dialysate volume during peritoneal dialysis for patient 1: — simulation, X – experimental data.

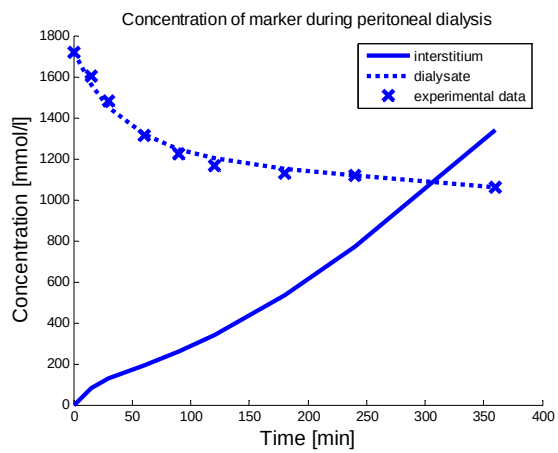


Fig. 4. Simulation of the marker concentrations during peritoneal dialysis for patient 2: — marker in interstitium, marker in dialysate, X – experimental data.

3. Discussion

The model examined in this paper seems to provide efficient description of the fluid and solute transport during peritoneal dialysis. However, this model still requires verification and confirmation. First of all, the experimental data of solute concentration in the interstitium during treatment are necessary, because then the fluid and solutes transport parameters estimated with the model will be more reliable.

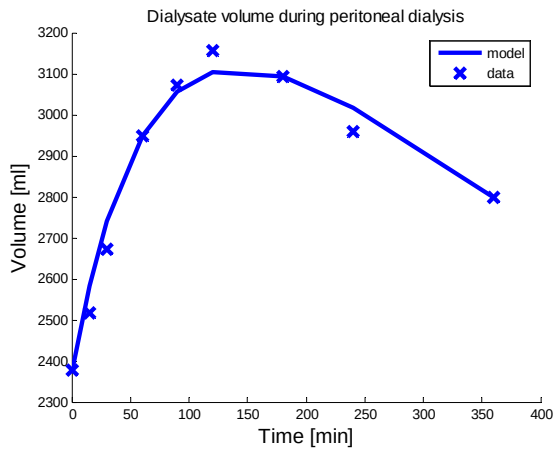
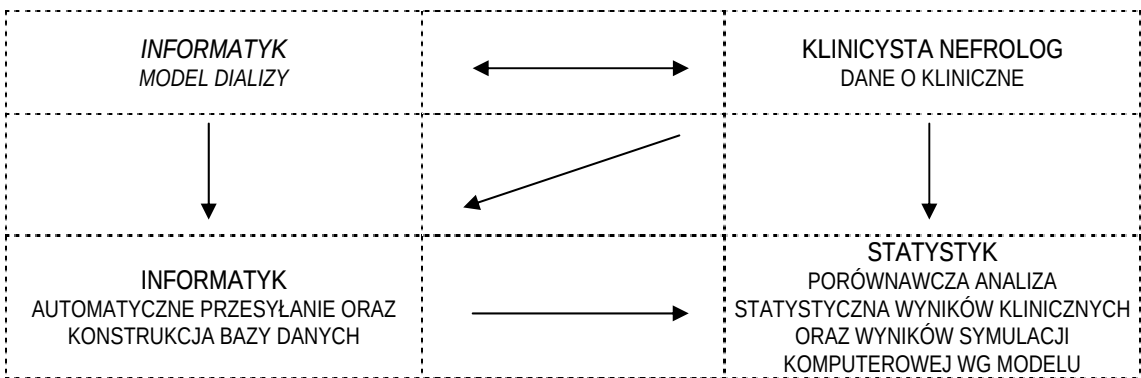


Fig.5. Dialysate volume during peritoneal dialysis for patient 2: — simulation, X – experimental data.

In summary, the model explains not only the discrepancy between fluid uptake and lymphatic absorption, but also between disappearance of the marker from dialysate and appearance of the marker in blood. In the previous models, without the additional compartment describing interstitium, it was assumed that these rates are equal. This did not agree with physiological investigations and experimental data. The simulations of the presented model show the reason of this disagreement as a result of the marker accumulation in the tissue.

Literature

1. Rippe B., Stelin G., Haraldsson B.: Computer simulations of peritoneal fluid transport in CAPD. Kidney Int, 40: 315-325, 1991.
2. Vonesh E.F., Lysaght M.J., Moran J., Farrell P.: Kinetic modeling as a prescription aid in peritoneal dialysis. Blood Purif, 9: 246-270, 1991.
3. Waniewski J., Werynski A., Heimburger O., Lindholm B.: Simple membrane models for peritoneal dialysis. Evaluation of diffusive and convective solute transport. Asaio J, 38: 788-796, 1992.
4. Waniewski J., Heimburger O., Werynski A., Lindholm B.: Simple models for fluid transport during peritoneal dialysis. Int J Artif Organs, 19: 455-466, 1996.
5. Haraldsson B.: Assessing the peritoneal dialysis capacities of individual patients. Kidney Int, 47: 1187-1198, 1995.
6. Rippe B., Simonsen O., Stelin G.: Clinical implications of a three-pore model of peritoneal transport. Adv Perit Dial, 7: 3-9, 1991.
7. Rippe B., Rosengren B.I., Venturoli D.: The peritoneal micro-circulation in peritoneal dialysis. Microcirculation, 8: 303-320, 2000.
8. Venturoli D., Rippe B.: Transport asymmetry in peritoneal dialysis: application of a serial heteroporous peritoneal membrane model. Am J Physiol Renal Physiol, 280: F599-606, 2001.
9. Rosengren B. I., Rippe B., Tenstad O., Wiig H.: Acute peritoneal dialysis in rats results in a marked reduction of interstitial colloid osmotic pressure. J Am Soc Nephrol, 15: 3111-3116, 2004.
10. Rippe B., Levin L.: Computer simulations of ultrafiltration profiles for an icodextrin-based peritoneal fluid in CAPD. Kidney Int, 57: 2546-2556, 2000.
11. Struijk D. G., Koomen G. C., Krediet R. T., Arisz L.: Indirect measurement of lymphatic absorption in CAPD patients is not influenced by trapping. Kidney Int, 41: 1668-1675, 1992.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 213-216.

SYMULACYJNY MODEL RODZINY PSZCZELEJ

SIMULATION MODEL OF A BEE COLONY

ANDRZEJ IZWORSKI*, ANDRZEJ MIGACZ**, MARCIN MICHAŁEK***

Katedra Automatyki Akademii Górniczo-Hutniczej, 30-059 Kraków, Al. Mickiewicza 30,
*e-mail: * izwa@agh.edu.pl,** ami@agh.edu.pl,*** michal@biocyb.ia.agh.edu.pl*

Streszczenie. W pracy przedstawiono opis modelu symulacyjnego złożonego systemu biologicznego, jakim jest rodzina pszczela (*Apis mellifica L.*). Zaprezentowano metody weryfikacji modelu oraz szacowania dostępnych pożytków pszczelich. Opisano również rozszerzenia modelu uwzględniające zjawisko konkurencji pomiędzy rodzinami pszczelimi. Prezentowane rozszerzenia podstawowego modelu funkcjonowania rodziny pszczelej mogą stać się narzędziem dla lepszego poznania powiązań rodziny pszczelej ze środowiskiem oraz stanowić bardzo elastyczny i uniwersalny sposób badania całokształtu zjawisk w układzie: pszczoła – gospodarka pasieczna. Oprócz aspektu poznawczego stworzony model symulacyjny ma również aspekt dydaktyczny, stanowiąc znaczącą pomoc podczas kształcenia zarówno studentów kierunków pszczelarskich, jak i pszczelarzy praktyków.

Słowa kluczowe: model rodziny pszczelej, przetwarzanie obrazów, symulacja

Abstract. A necessary condition for gaining practical profits from computer modeling is a quantitative and qualitative consistence of the results of simulation with real functioning of the system submitted to modeling. The paper concerns the computer model of simulated functioning of a bee colony (*Apis Mellifera L.*) and problem of verification of the model. The paper also presents an attempt at modeling description of the passive competition phenomenon among the bee families concerning a common source of food. This phenomenon has been analysed to show the rules of dividing the limited food reserve and the use of this reserve. Moreover, we disscused strength of a bee family, efficiency, location and the attraction of a food source based on photographs and image, influences of random factors, weather, rain, etc. Effect of absence of the bee queen in bee colony on the number of worker bees was analysed too. For investigation purposes a simulation model implemented on a PC has been used.

Key words: bee colony, computer modeling, simulation, image recognition

1. Wstęp

Przedmiotem modelowania cybernetycznego zazwyczaj są wybrane fragmenty systemów biologicznych (poszczególne narządy, organy i mechanizmy). Niniejsza praca poświęcona jest modelowi o większej skali, przedstawiającemu system złożony z wielkiej liczby osobników połączonych zależnościami typu społecznego, a konkretnie modelowi rodziny pszczelej (Apis mellifica L.). Rozważany system funkcjonuje w powiązaniu ze środowiskiem, z którego pobiera niezbędne do życia produkty, podlegając jednocześnie zakłócającym oddziaływaniom zlokalizowanym w środowisku (temperatury, opadów i warunków fenologicznych), w związku z czym model komputerowy rozważanego systemu musi zawierać także element symulacji środowiska.

W poprzednich pracach autorów zaprezentowano założenia symulacyjne komputerowego modelu rodziny pszczelej (*Apis mellifera L.)* [5] oraz dokonano przeglądu aspektów dydaktycznych tego modelu [2]. Prezentowana wówczas wersja modelu charakteryzowała się tekstowym interfejsem, a grafika stanowiła jedynie minimalne i konieczne uzupełnienie modelu. W chwili obecnej model uzyskał już nowy graficzny interfejs użytkownika [3]. Wersja ta jest odpowiedzią na wzrastające możliwości współczesnych systemów komputerowych w zakresie wizualizacji i interakcji graficznej. Właśnie metody wizualizacji i interakcji graficznej dzięki syntetycznej prezentacji wyników wielu pomiarów czy eksperymentów dokonywanych na modelu symulacyjnym pozwalają na dokonanie szybkiej oceny i ewentualne wskazanie istotnych własności prezentowanych modeli. Może to nawet prowadzić do formułowania nowych hipotez badawczych.

Tworząc nowy model, zdecydowano się na umiejscowienie go w środowisku programowym MATLAB. Wydaje się, że nowa wersja modelu, uwzględniająca w znacznie większym stopniu metody wizualizacyjne oraz techniki interakcji graficznej, powinna znaleźć jak najszersze zastosowanie w szeroko pojętej dydaktyce pszczelarskiej.

2. Model rodziny pszczelej

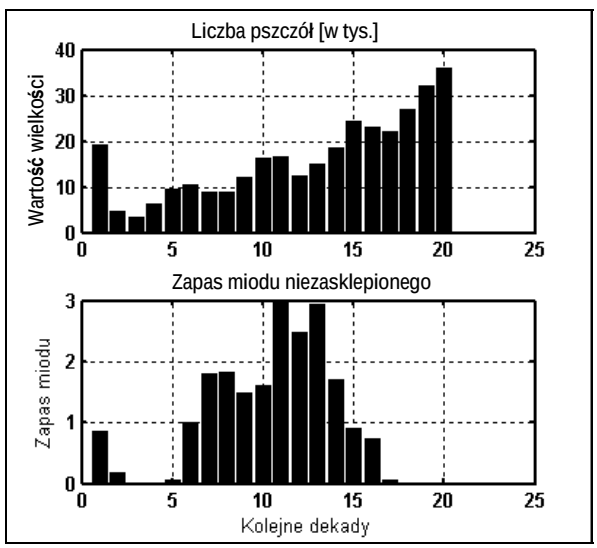
Budując model funkcjonowania rodziny pszczelej, ustalono następujące cechy, jakie powinien on posiadać:

- Stanowić adekwatny opis jakościowy i ilościowy funkcjonowania rodziny pszczelej z dokładnością wystarczającą do:
  - Zrozumienia i przebadania prawidłowości stymulujących funkcjonowanie rodziny pszczelej,
  - Ułatwienia decyzji pszczelarzowi.
- Posiadać możliwość korekty struktury i parametrów w oparciu o aktualnie dostarczone informacje.

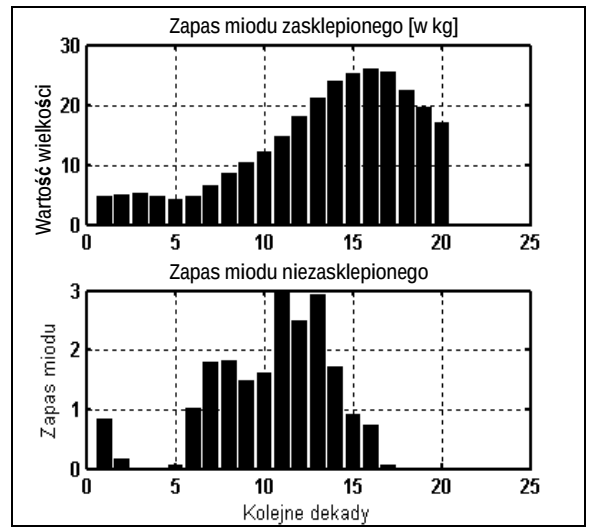
Zasadniczy schemat funkcjonalny (struktura wewnętrzna modelu) opisany został szczegółowo w pracy [4].

Podstawowym modułem modelu jest blok symulujący zjawiska rozrodu, dojrzewania i wymierania owadów. Oprócz reprodukcji zasadniczą funkcją warunkującą przetrwanie modelowanego obiektu jest pozyskiwanie ze środowiska nektaru, spadzi i pyłku. Na ogół przetrwanie rodziny pszczelej możliwe jest pod warunkiem istnienia zapasów, dlatego funkcje magazynowe odgrywają w modelu ważną rolę. Miód gromadzony jest w postaci niezasklepionej i zasklepionej, natomiast pyłek w postaci pierzgi. Na stan tych zapasów ma także wpływ ilość produktów zjedzona przez pszczoły oraz działalność człowieka, który może dodawać lub odejmować odpowiednie ich ilości. Część zgromadzonych zapasów ulega spożyciu, przy czym proces ten w modelu obsługują aż trzy bloki. Elementem domykającym model systemu jest model procesu budownictwa.

Zasadniczą trudność poza zaprogramowaniem modelu stanowiło oszacowanie wartości parametrów omówionych w modelu stałych. Wartości tych parametrów przyjęto na podstawie danych literaturowych, a także oparto na oszacowaniach i indywidualnie prowadzonym eksperymencie identyfikacyjnym. Rysunki 1 i 2 przedstawiają przebiegi uzyskane z modelu – odpowiednio liczbę pszczół oraz zapas miodu zasklepionego na tle zapasu miodu niezasklepionego.



Rys. 1. Liczba pszczół w tysiącach

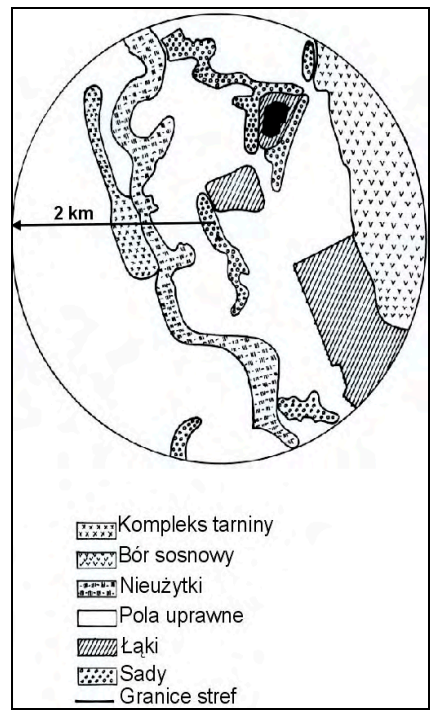


Rys. 2. Zapas miodu zasklepionego

Osobny program służy do opisu warunków atmosferycznych oraz pożytków pszczelich dla potrzeb modelu.

3. Szacowanie pożytków pszczelich na podstawie zdjęć lotniczych

Rodzina pszczela żyje w swoim naturalnym otoczeniu i jest z nim silnie związana. Środowisko dysponuje zapasami nektaru, spadzi oraz pyłku kwiatowego, które to surowce są pobierane przez pszczoły, a następnie spożywane, przerabiane na inne produkty oraz magazynowane. Pod pojęciem otoczenia należy rozumieć praktycznie powierzchnię koła (wokół ula czy pasieki) o promieniu 2 km wraz ze źródłami surowców interesujących pszczoły, co zostało przedstawione na rys. 3. Promień ten przyjmuje się jako ekonomiczny zasięg lotu pszczół.



Rys. 3. Rozmieszczenie obszarów pożytkowych w terenie. Skala 1:26000

Na potrzeby modelu podjęto próbę wykorzystania metod przetwarzania i analizy cyfrowych obrazów kolorowych do szacowania ilości pożytków pszczelich, co z kolei pozwala na określenie warunków bytowych rodziny pszczelej i prognozowanie możliwości rozwoju populacji takiej rodziny. Szacowanie jest dokonywane na podstawie wstępnie zinterpretowanego przez człowieka zdjęcia lotniczego (rys. 4).

Rys. 4. Zdjęcie lotnicze pokazujące teren wokół pasieki

Interpretacja polega na określeniu klas poszczególnych obszarów widocznych na zdjęciu, a jej rezultatem jest kolorowy obraz z naniesionymi granicami poszczególnych klas. Obraz ten jest następnie skanowany i już w postaci cyfrowej przetwarzany (m.in. dokonywana jest zmiana sposobu reprezentacji barw z RGB na model CIE, detekcja krawędzi, szkieletyzacja, uciąglanie krawędzi, wypełnianie obszarów, a końcowym efektem jest obraz z wydzielonymi obszarami odpowiadającymi poszczególnym klasom obszarów (pożytków)). Wyznaczone pola powierzchni znalezionych obszarów pozwalają już na określenie względnej ilości rozpatrywanych pożytków na analizowanym obszarze.

4. Modelowanie zjawiska konkurencji rodzin pszczelich

Rozwój każdego gatunku jest nierozerwalnie związany ze środowiskiem, w którym on żyje i zależy od całokształtu warunków w obszarze jego działalności. Prawie zawsze w naturze występują zależności konkurencyjne związane z walką o osiągnięcie środków koniecznych do podtrzymania życia lub z podziałem zasobów. Prowadzone eksperymenty dotyczyły próby modelowego ujęcia zjawisk biernej konkurencji rodzin pszczelich względem wspólnego źródła wziątku. Zjawisko konkurencji rozpatrzono z punktu widzenia podziału ograniczonych zasobów pomiędzy określoną liczbą rodzin pszczelich oraz w aspekcie wykorzystania źródła wziątku. Ujęto wpływ następujących czynników określających konkurencję:

- Siłę rodziny pszczelej, wydajność, lokalizację i „atrakcyjność" źródła wziątku,
- Poszukiwanie, przekazywanie i wymianę informacji,
- Wpływ niedeterminizmu oraz „pamięć" lokalizacji w trakcie poszukiwania źródła wziątku,
- Wpływ warunków atmosferycznych.

W eksperymentach przyjęto liczbę konkurujących ze sobą rodzin pszczelich wynoszącą: IK = 5.

Rozpatrywano dwa modele konkurencji – pierwszy z nich uwzględniał „pamięć" (opisaną niżej), drugi natomiast już jej nie uwzględniał.

Jeżeli pszczoły i-tej rodziny znalazły źródło surowca w określonym kroku symulacji, to trafiały do niego z prawdopodobieństwem równym jeden na każdym kolejnym kroku, o ile nie wystąpiły czynniki zacierające pamięć. Jak wynika zarówno z literatury, jak i z własnych doświadczeń, do czynników zacierających pamięć zaliczyć należy warunki atmosferyczne (opady, silne obniżenie temperatury) uniemożliwiające wykonywanie lotów przez określony czas (1-3 doby). Z chwilą poprawy warunków atmosferycznych poszukiwanie celu rozpoczynało się od nowa.

Uwzględnienie w modelu dynamiki procesu, warunków środowiska oraz zachowań pszczół w zakresie pozyskiwania surowców pokarmowych pozwoliło na wnikliwszą analizę zjawiska konkurencji niż oszacowania globalne. Stwierdzono również ważność czynnika pamięci jako znaczącego w ocenie ewolucyjnej miary przystosowania gatunku. W warunkach globalnej oceny, sugerującej całkowite wykorzystanie dostępnego źródła surowca, w praktyce niejednokrotnie część surowca nie zostanie wykorzystana. Zatem na podstawie globalnych oszacowań wydajności poszczególnych źródeł surowca nie można z góry przesądzić o ilości pozyskanego miodu towarowego. Z badań modelowych wynika też potrzeba tworzenia i wzbogacania źródeł pokarmowych, szczególnie ważnych w związku z nasilającą się obecnie chemizacją rolnictwa.

Dodatkową korzyścią jest fakt, iż na podstawie przeprowadzonych eksperymentów można dobrać optymalną rasę pszczół oraz liczebność rodzin pszczelich, dopasowane do oszacowanych terminów i wydajności dostępnych dla nich na terenie źródeł pokarmowych.

5. Weryfikacja modelu

Warunkiem koniecznym uzyskania z komputerowego modelowania jakichkolwiek korzyści praktycznych jest dobra ilościowa i jakościowa zgodność wyników symulacji z rzeczywistym funkcjonowaniem podlegającego modelowaniu systemu. Weryfikacja modelu polegała na porównaniu wyników obliczeniowych z oszacowaniami uzyskanymi dla wybranej rodziny lub grupy rodzin pszczelich. Stan biologiczny rodziny pszczelej określony jest dużą liczbą parametrów, a wybór ich w formie jednoznacznego kryterium jest dosyć trudny. W niniejszej pracy jako parametry charakteryzujące rodzinę pszczelą z punktu widzenia biologicznego przyjęto:

- Liczbę pszczół robotnic w rodzinie,
- Dobową produkcję czerwiu przez matkę pszczelą,
- Liczbę komórek czerwiu zasklepionego.

Wybrane wielkości były szacowane dwukrotnie (w różnych latach) w pasiece dla wybranej rodziny pszczelej, a badania weryfikacyjne prowadzono w ciągu całego sezonu pasiecznego. Liczbę pszczół robotnic szacowano metodą ważenia samych pszczół. Liczbę komórek z czerwiem zasklepionym szacowano metodą fotogrametryczną w 11-dniowych odstę-

pach czasu. Dobową produkcję czerwiu przez matkę pszczelą obliczono z oszacowań czerwiu zasklepionego.

Dla tych samych warunków pożytkowych i atmosferycznych uruchomiono model symulacyjny. W trakcie eksperymentu zginęła matka pszczela, co w modelu również zostało uwzględnione. Tak jak rodzina pszczela, również i model przez 20 dni był pozbawiony dopływu czerwiu. Pszczoły wychowały matkę z mateczników ratunkowych, złożonych na młodych larwach pszczelich. Zaczęła ona składać jaja dopiero po upływie 20-23 dni.

Średni błąd procentowy symulacji (odnoszony do wartości oszacowanej) zawierał się w granicach od +4,72% do –21,42% dla liczby komórek czerwiu zasklepionego, od +5,25% do –25% dla produkcji jaj przez matkę pszczelą, i od +2% do –16% dla liczby pszczół. Wielkości te należy uznać za zadowalające.

6. Wnioski

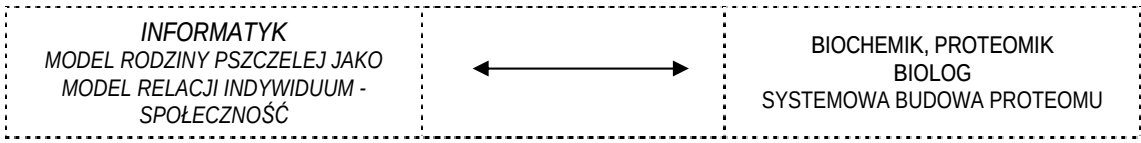
Matematyczny model rodziny pszczelej w jego komputerowej realizacji może być wykorzystany do badania struktury i zależności występujących w rodzinie pszczelej. Zastosowanie modelu pozwala na podjęcie określonych decyzji w aktualnych warunkach atmosferycznych i pożytkowych. Przeprowadzone eksperymenty potwierdzają znany pogląd, że najbezpieczniejszymi sposobami wymiany matek pszczelich są te, które nie powodują dłuższej niż 10 dni przerwy w produkcji czerwiu.

Prezentowane modele rozszerzające podstawowy model funkcjonowania rodziny pszczelej mogą stać się narzędziem dla lepszego poznania powiązań rodziny pszczelej ze środowiskiem oraz stanowić bardzo elastyczny i uniwersalny sposób badania całokształtu zjawisk w układzie: pszczoła – gospodarka pasieczna.

Praca wykonana w ramach badań statutowych AGH, nr 11.11.120.249.

Literatura cytowana

1. Bubliński Z., Migacz A.: Szacowanie pożytków pszczelich na podstawie wstępnie zinterpretowanego zdjęcia lotniczego, Materiały Seminarium: Przetwarzanie i analiza sygnałów w systemach wizji i sterowania, Słok k/Bełchatowa 2002.
2. Migacz A.: Dydaktyczne aspekty komputerowego modelu rodziny pszczelej, Materiały XXXVIII Naukowej Konferencji Pszczelarskiej, Puławy 2001.
3. Migacz A., Izworski A., Michałek M.: Multimedialne funkcje symulacyjnego modelu rodziny pszczelej, Materiały XL Naukowej Konferencji Pszczelarskiej, Puławy 2003.
4. Migacz A., Tadeusiewicz R.: Komputerowy model rodziny pszczelej, Elektrotechnika, tom 1, zeszyt 1, str. 67-79, 1982.
5. Migacz A., Izworski A.: Cybernetyczny model złożonego systemu biologicznego na przykładzie rodziny pszczelej, Materiały XXXI Naukowej Konferencji Pszczelarskiej, Puławy 1994.
6. Migacz A.: Zastosowanie zdjęć lotniczych do szacowania wydajności miodowo-pyłkowej „pastwisk" pszczelich, Materiały XXXIX Naukowej Konferencji Pszczelarskiej, Puławy 2002.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 217-220.

FUNKCJONALNY MODEL SYMULACYJNY ZESPOŁU SYNAPS ZAPAMIĘTUJĄCYCH KOMÓRKI PIRAMIDOWEJ HIPOKAMPA

THE FUNCTIONAL MODEL OF MEMORIZING PYRAMIDAL CELL SYNAPSES COMPLEX IN HIPPOCAMPUS

EDWARD JACEK GORZELAŃCZYK [*], MICHAŁ HUFLEJT [**], JAN KNIAT [**],
EWA LASKOWSKA [*], JANUSZ MURAKOWSKI [***], PIOTR WOŹNIAK [****]

[*] *Zakład Neuropsychologii i Genetyki Behawioralnej, Akademia Bydgoska, Bydgoszcz, medsystem@medsystem.com.pl*
[**] *Instytut Informatyki, Politechnika Poznańska, Poznań,*
[***] *Department of Electrical and Computer Engineering University of Delawar,*
[****] *SuperMemo Research*

Streszczenie. Zaproponowano model komórki piramidowej hipokampa umożliwiający przeprowadzenie obliczeń w środowisku rozproszonym. W modelu uwzględniono: aktywność elektryczną pojedynczych synaps i receptorów glutamatergicznych oraz aktywność jądra komórkowego i transport receptorów. Podstawą modelu jest koncepcja alternatywnego składania genów. Symulator umożliwia na przykład modelowanie potencjałów czynnościowych, efektu przerwy, zjawiska długotrwałego wzmocnienia synaptycznego.

Słowa kluczowe: model biologiczny, synapsa glutamatergiczna, długotrwałe wzmocnienie synaptyczne, alternatywne składanie genów

Abstract. The model of pyramidal cell is proposed. With this model it is possible to make calculations in diffuse environment. This model considers electrical activity of the individual synapses and glutamate receptors, and cell nucleus activity as well as the transport of receptors. The basis of this model is the alternative splicing of genes. The simulator allows for example modeling action potentials, spacing effect, and long term synaptic potentiation.

Key words: biological model, glutamate synapse, long term potentiation, alternative splicing

Wstęp

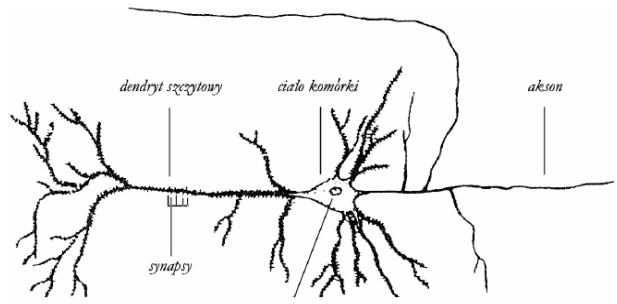
Funkcja komórki piramidowej hipokampa wiązana jest z procesem zapamiętywania [1]. Istniejące środowiska symulacyjne, choć przydatne, nie pozwalają na stworzenie modelu działania komórki piramidowej z uwzględnieniem procesu zapamiętywania oraz zaimplementowanie symulacji w postaci umożliwiającej przeprowadzenie obliczeń w środowisku rozproszonym. Oprogramowanie Genesis [2] pozwala na zbadanie aktywności elektrycznej rozbudowanych połączeń, jednak nie pozwala na symulowanie populacji receptorów błonowych, co konieczne jest, by symulować uczenie się neuronu. Program Neuron [3] umożliwia symulacje aktywności elektrycznej uwzględniające receptory postsynaptyczne, lecz nie pozwala na symulację komunikacji między błoną neuronu a jądrem komórkowym, konieczną dla implementacji koncepcji udziału alternatywnego składania genów w procesie uczenia się synapsy [4].

Oprogramowanie symulacyjne

Konstruowanie modelu działania komórki piramidowej oraz wytworzenie oprogramowania realizującego model obliczeniowy to procesy współzależne. Zarówno model fizyczny określał kształt realizacji technicznej, jak i zagadnienia techniczne wpływały na model symulowanych zjawisk. Zastosowano cykliczne podejście, typowe dla programowania odkrywczego (ang. *exploratory programming*) [4]. Cykl składał się z następujących etapów: 1) tworzenie i modyfikacja modelu fizycznego; 2) przekształcenie modelu fizycznego w model oprogramowania; 3) implementacja; 4) walidacja – obserwacja i ocena wyników symulacji.

Modelowanie komórki piramidowej hipokampa

Najistotniejszą cechą proponowanego modelu jest uwzględnienie hipotez dotyczących mechanizmów plastyczności komórki nerwowej, opartych na alternatywnym składaniu genów (ang. *alternative splicing*) [5]. Z uwagi na wysoki poziom złożoności modelu przyjęto następującą kolejność realizacji poszczególnych składowych modelu: 1) aktywność elektryczna pojedynczego elementu symulacji; 2) aktywność receptorów glutamatergicznych w pojedynczym elemencie symulacji; 3) aktywność jądra komórkowego oraz transportu receptorów do błony komórkowej neuronu; 4) uwzględnienie wielu elementów symulacji; 5) rozproszenie aplikacji.



Rys. 1. Komórka piramidowa

Komórka piramidowa hipokampa

Komórka piramidowa hipokampa zbudowana jest z ciała komórki oraz wypustek nerwowych: jednego neurytu, który przewodzi sygnał wyjściowy i około 150000 dendrytów, przez które (za pośrednictwem synaps) docierają sygnały wejściowe.

Impuls przekazywany jest przez synapsy poprzez receptory gultamatergiczne wbudowane w błony zakończeń pre- i postsynaptycznych. Receptory aktywowane są przez kwas glutaminowy. Receptory danego typu mogą występować w wariantach różniących się składem podjednostkowym. Poszczególne warianty danego typu receptora różnią się parametrami fizyko-chemicznymi. Właściwości receptorów (np. ich przewodność dla jonów) mogą zmieniać się w wyniku ich modyfikacji, np. fosforylacji receptora [5].

Liczba receptorów w synapsie jest zmienna i zależy od rodzaju oraz historii synapsy. Siła połączenia synaptycznego zależy między innymi od przewodności, jaką posiadają aktywowane przez neuroprzekaźnik receptory. Przewodność synapsy zależy m.in. od liczby receptorów, ich składu podjednostkowego, stopnia fosforyzacji [5].

Receptory AMPA i NMDA

Receptory AMPA przewodzą jony sodu i potasu, a niektóre odmiany podjednostkowe w niewielkim stopniu przewodzą także jony wapnia. Do aktywacji receptora AMPA konieczne jest przyłączenie cząsteczki glutaminianu.

Receptory NMDA przewodzą jony sodu, potasu i wapnia. Do aktywacji receptora konieczne jest przyłączenie dwóch cząsteczek glutaminianu. W porównaniu do receptora AMPA, NMDA trudniej jest pobudzić co ma znaczenie dla przetwarzania sygnałów.

Gdy polaryzacja elektryczna błony komórki nerwowej wzrasta, wówczas kanał receptora NMDA jest blokowany przez jon magnezu i wówczas receptor nie przewodzi jonów, co ma miejsce podczas potencjału spoczynkowego, kiedy to receptory NMDA prawie wcale nie przewodzą.

Do przewodzenia receptora NMDA potrzebne są dwa warunki: 1) receptor musi zostać aktywowany neuroprzekaźnikiem; 2) Błona musi zostać depolaryzowana.

Spełnienie pierwszego warunku zależy od błony presynaptycznej, a drugiego od błony postsynaptycznej [5].

Rola wapnia i wtórnych przekaźników

Stężenie wapnia jest istotnym parametrem regulującym procesy plastyczności. Wzrost stężenia wapnia aktywuje za pośrednictwem tzw. wtórnych przekaźników szereg mechanizmów odpowiedzialnych za uczenie się komórki. Stężenia wtórnych przekaźników regulują nie tylko fosforylacje, ale także wpływają na mechanizmy związane z umieszczeniem nowych receptorów w błonie (dokowanie), usuwaniem receptorów z błony i ich rozkład. Przede wszystkim jednak stężenie wtórnych przekaźników reguluje procesy syntezy receptorów w jądrze komórkowym poprzez odblokowanie informacji genetycznej [5].

Synteza i transport receptorów

Transport receptorów nie jest dobrze zbadany. Prawdopodobnie odbywa się dzięki czynności cytoszkieletu. Ruchy cytoszkieletu wpływają na mieszanie się cytoplazmy i szybkość dyfuzji.

Fosforylacja receptorów

Reakcja przyłączenia reszty kwasu fosforowego ma wpływ na właściwości receptorów. Przewodność sfosforylowanego receptora może się zmienić.

Plastyczność komórki piramidowej

Najbardziej istotnymi procesami dla plastyczności komórki są zjawiska długotrwałego wzmocnienia synaptycznego (ang. *long-term potentiation*, LTP) oraz długotrwałego osłabienia synaptycznego (ang. *long-term depression*, LTD). Długotrwałe wzmocnienie synaptyczne zachodzi pod wpływem krótkiego, wysokoczęstotliwościowego pobudzenia błony postsynaptycznej. Wzmocnienie siły sprzężenia synaptycznego utrzymuje się długo po ustąpieniu bodźca, nawet przez wiele dni. Długotrwałe osłabienie synaptyczne pojawia się w wyniku długotrwałego pobudzania komórki niską częstotliwością [5].

Wykrywanie jednoczesności

Do aktywacji receptorów NMDA spełnione muszą być jednocześnie dwa warunki: 1) Receptor musi zostać aktywowany neuroprzekaźnikiem; 2) Błona musi zostać depolaryzowana.

Oznacza to, że jednoczesna aktywność błony presynaptycznej oraz błony postsynaptycznej powoduje otwieranie się receptorów NMDA i napływ wapnia do wnętrza komórki. Receptory NMDA pełnią role mechanizmu wykrywania jednoczesności (ang. *coincidence detection*), która sygnalizowana jest wzrostem stężenia wapnia wewnątrz komórki. Wzrost stężenia wapnia uruchamia szereg procesów prowadzących do wzmocnienia siły sprzężenia synaptycznego. Receptory NMDA stanowią realizację reguły Hebba na poziomie molekularnym.

Wzmocnienie siły sprzężenia synaptycznego zachodzi przy dużej korelacji pomiędzy pobudzeniem błony presynaptycznej a pobudzeniem błony postsynaptycznej. Reguła ta jest spełniona także dla większej liczby synaps. Rozważmy komórkę o dwóch synapsach: a i b. Synapsa a ma silne sprzężenie synaptyczne, wystarczające do depolaryzacji całej komórki. Synapsa b ma słabe sprzężenie synaptyczne, niewystarczające do depolaryzacji komórki. Pobudzanie synapsy b o słabym sprzężeniu synaptycznym mogłoby doprowadzić do depolaryzacji jedynie przypadkowo. Wzmocnienie

siły synapsy b może jednak łatwo nastąpić, jeśli będzie ona pobudzana równocześnie z synapsą a. Wówczas, dla synapsy b spełnione są oba warunki niezbędne do długotrwałej zmiany w jej sprzężeniu synaptycznym. Pobudzanie synapsy b neuroprzekaźnikiem umożliwi jej receptorom NMDA aktywację, a fala depolaryzacji pochodząca od synapsy a zwolni blokadę magnezową receptorów NMDA synapsy b. Zajdzie jednocześnie pobudzenie błony presynaptycznej i postsynaptycznej (warunek konieczny dla zwiększenia sprzężenia synaptycznego). Czynność receptorów NDMA, wykrywanie jednoczesności i zmiany plastyczne jako wynik ekspresji genów pozwalają wyjaśnić na poziomie molekularnym działanie reguły Hebba zarówno homosynaptyczne jak i heterosynaptyczne. Opis ten wyjaśnia zarówno wzmacnianie pobudzanej synapsy, jak i wytwarzanie skojarzenia pomiędzy wieloma synapsami [5].

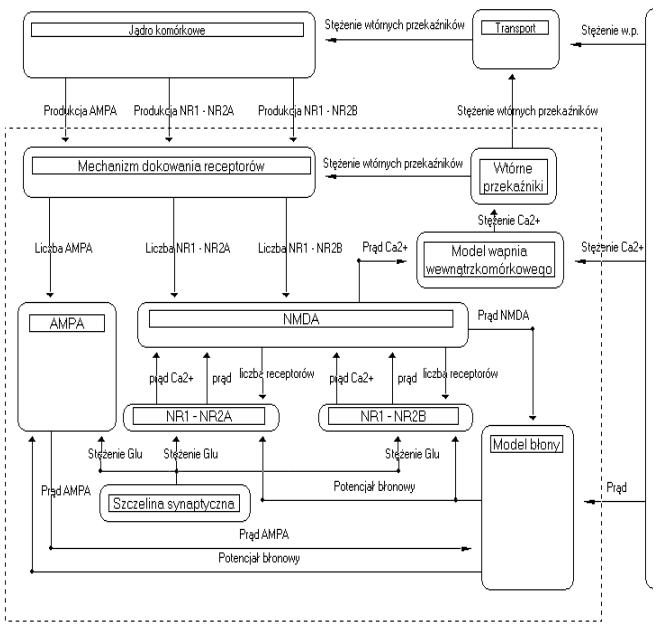
Składniki modelu fizycznego

Konstruowanie modelu fizycznego komórki piramidowej wymagało połączenia wiedzy na temat aktywności elektrycznej komórki, fizjologii receptorów postsynaptycznych, a także mechanizmów transportu substancji wewnątrz komórki. Niektóre założenia oparte zostały na hipotezach wymagających weryfikacji i uściślenia. Model zawiera szereg uproszczeń służących usprawnieniu obliczeń i ułatwieniu ich rozproszenia.

Zagadnienia techniczne wywarły więc wpływ na ostateczny kształt modelu fizycznego.

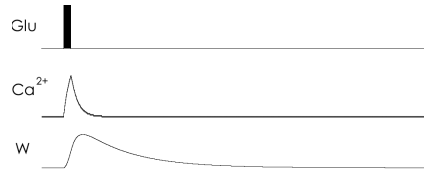
W modelu fizycznym zostały uwzględnione: aktywność elektryczna błony komórkowej, receptory błonowe (AMPA i NMDA), wzajemny wpływ sąsiadujących elementów symulacji, rola wapnia i wtórnych przekaźników, fosforylacja receptorów, powstawanie i zanikanie wtórnych przekaźników, transport wtórnych przekaźników do jadra komórkowego, mechanizm dokowania receptorów, odmiany podjednostkowe receptorów NMDA, mechanizm syntezy i transportu receptorów.

Przekształcenie modelu fizycznego w model przepływu danych

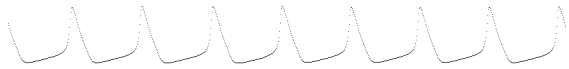


Rys. 2. Diagram przepływu danych

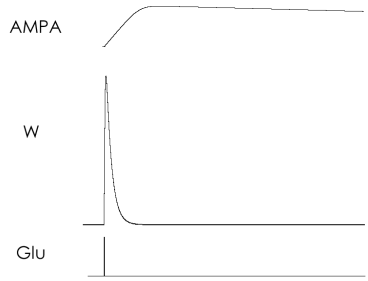
Wybrane wyniki symulacji



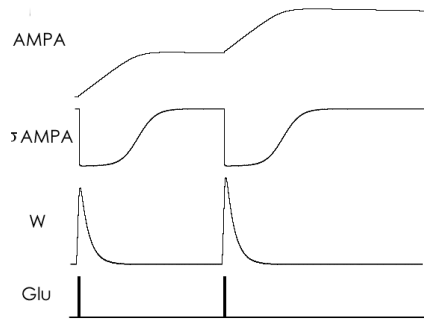
Rys. 3. Napływ wapnia i synteza wtórnych przekaźników; Glu – stężenie glutaminianu w szczelinie synaptycznej, Ca^{2+} – stężenie jonów wapnia w cytoplazmie, W – stężenie wtórnych przekaźników



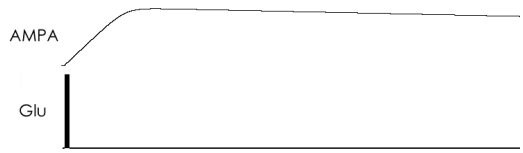
Rys. 4. Potencjały czynnościowe wywołane impulsami glutaminianu o częstotliwości 50 Hz



Rys. 5. Zjawisko LTP; AMPA – liczba zadokowanych receptorów; W – stężenie wtórnych przekaźników, Glu – stężenie glutaminianu w szczelinie synaptycznej



Rys. 6. Efekt powtórki; AMPA – liczba zadokowanych receptorów; W – stężenie wtórnych przekaźników, Glu – stężenie glutaminianu w szczelinie synaptycznej, ςAMPA – stężenie recept. w cytoplazmie



Rys. 7. Efekt powtórki. AMPA – liczba zadokowanych receptorów, Glu – stężenie glutaminianu w szczelinie synaptycznej. Początkowa liczba zadokowanych receptorów wynosi 50, a po wystąpieniu LTP rośnie tylko do 62

Wnioski neurobiologiczne

Zastąpienie złożonego i nie dość dobrze poznanego mechanizmu wtórnych przekaźników jedną wielkością W pozwoliło abstrahować od nieznanych szczegółów, zachowując najistotniejsze cechy mechanizmu przekazywania informacji do jądra komórkowego oraz regulacje dokowania receptorów

w synapsie. Już tak prosty model, oparty o jedną zmienną, pozwolił na wyjaśnienie kolejno opisanych zjawisk.

Translacja informacji

Przekształcanie informacji o wykryciu jednoczesności pobudzeń pre- i postsynaptycznych odbywa się dzięki wzrostowi stężenia jonów wapnia i innych substancji. Zmienna *W* zastępuje tu całą kaskadę reakcji chemicznych.

Przekazywanie informacji do jądra komórkowego

Przekazywanie informacji o jednoczesności do jądra komórkowego. Wtórne przekaźniki W mogą być transportowane do jądra komórkowego przez cytoszkielet i na drodze dyfuzji.

Regulacja syntezy receptorów

Wtórne przekaźniki W regulują szybkość syntezy receptorów w jądrze komórkowym, która odbywa się na drodze ekspresji genów. W odpowiedzi na zwiększone stężenie W jądro komórkowe wytwarza więcej receptorów.

Regulacja dokowania receptorów

Model ze zmienną W wyjaśnia także regulacje dokowania receptorów w synapsie, a zwłaszcza selektywne nasilenie procesu dokowania receptorów w tych synapsach, w których wzrasta stężenie wtórnych przekaźników. W ten sposób wzmocnienie następuje w tej synapsie, która jest bardziej pobudzona (receptory dokowane są w tych synapsach, z których pochodziło pobudzenie). Wyjaśnia to, dlaczego jądro nie adresuje receptorów do poszczególnych synaps.

Odstępy pomiędzy pobudzeniami a szybkość uczenia się

W zależności od stałej czasowej procesu usuwania wtórnych przekaźników w synapsie silniej lub słabiej może zaznaczać się wpływ odstępów pomiędzy pobudzeniami na szybkość uczenia się. Jeśli pod wpływem pobudzenia komórka wytworzyła receptory, to zostaną one zadokowane w obrębie tych synaps, w których występują wtórne przekaźniki. Stężenie wtórnych przekaźników może się zwiększyć w synapsie pod wpływem kolejnego pobudzenia. Jeśli takie pobudzenie, które spowoduje wzrost stężenia w obrębie synapsy, wystąpi równocześnie ze zwiększeniem się stężenia receptorów gotowych do dokowania, to nastąpi wzmożenie procesu dokowania. W ten sposób, jeśli pobudzi się synapsę po raz drugi w czasie, gdy w postaci receptorów przyszła odpowiedź z jądra na pierwsze pobudzenie, to wystąpi ułatwione dokowanie. Można by nazwać to zjawisko homosynaptycznym efektem powtórki. Możliwy jest także następujący scenariusz: Synapsa A zostaje pobudzona, a wtórne przekaźniki wysyłają sygnał dla jądra komórkowego, co zwiększa produkcję receptorów. W cytoplazmie komórki pojawia się więcej receptorów. Synapsa A podlega wzmocnieniu, ponieważ przekaźniki W nadal występują w jej obrębie. W momencie, gdy w cytoplazmie pojawiło się więcej receptorów, pobudzono synapsę B. Rosnące stężenie wtórnych przekaźników w obrębie synapsy B umożliwia jej dokowanie receptorów. Wzmocnienie synapsy B zwiększy się bardziej niż wzmocnienie synapsy A, ponieważ w chwili, gdy było dużo wolnych receptorów, synapsa B miała wyższe stężenie W niż synapsa A. Takie zjawisko można by nazwać heterosynaptycznym efektem powtórki. Jak wcześniej wspomniano, stopień, w jakim występują te efekty, zależy od stałej czasowej zanikania wtórnych przekaźników W. Jeżeli stężenie przekaźników będzie szybko spadać, to docierające z jądra do synaps receptory nie zdążą być zadokowane, chyba że synapsa zostanie pobudzona ponownie. W takim przypadku efekt powtórki będzie skrajnie silny. Jeżeli natomiast wtórne przekaźniki będą usuwane bardzo wolno, to dokowanie receptorów będzie najczęściej łatwe, bo wysokie stężenie W będzie się długo utrzymywać, a efekt powtórki prawie nie będzie występował. Ustalenia wymaga, jaka jest stała czasowa tego procesu w żywej komórce. Należy się spodziewać, że do obniżenia stężenia W dochodzi szybko. Długi czas zanikania W wydaje się niekorzystny, ponieważ doprowadzałby w krótkim czasie do nasycenia się synaps receptorami.

Zjawiska homo- i heterosynaptycznego efektu powtórki mogą mieć istotne znaczenie w układach złożonych z wielu komórek nerwowych. Wiadomo, że w procesach poznawczych odstępy pomiędzy powtórkami mają wpływ na skuteczność zapamiętywania [5], [6], [7], [8]. Być może przedstawione tu zjawiska stanowią molekularną podstawę efektów występujących w wyższych procesach poznawczych.

Literatura cytowana

1. Bernabeu R., Bevilaqua L., Ardenghi P. i wsp.: Involvement of hippocampal cAMP/cAMP-dependent protein kinase signaling pathways in a late memory consolidation phase of aversively motivated learning in rats. Proceedings of the National Academy of Sciences of the United States of America, 94(13): 7041-6, 1997.
2. Dayan P., Abbott L. F.: Theoretical Neuroscience. Computational and Mathematical Modeling of Neural Systems., The MIT Press, Cambridge, Massachusetts, London, England, 2001.
3. Destexhe A., Mainen Z. F., Sejnowski T. J.: Methods in Neuronal Modeling, 2nd Edition, The MIT Press, Cambridge, MA, 1998.
4. Hedin G., Magnusson B.: Supporting Exploratory Programming, chapter 27, In Knudsen et al., 1993.
5. Gorzelańczyk E. J., Kwieciński Ł., Laskowska I., Lenart-Jankowska D., Woźniak P.: Alternatywne składanie genów jako wyjaśnienie procesów plastycznych zachodzących w synapsie zapamiętującej, Aktualności Neurologiczne, 4(2): 89-97, 2004.
6. Sternberg R. J.: Cognitive Psychology, Harcourt, 1999 (2nd ed.)
7. Woźniak P. A., Gorzelańczyk E. J.: Optimization of repetition spacing in the practice of learning, Acta Neurobiologiae Experimentalis, 54, 59-62, 1994.
8. Woźniak P. A., Gorzelańczyk E. J., Murakowski J. A.: Two components of long-term memory, Acta Neurobiologiae Experimentalis, 55, 301-305, 1995.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 221-226.

MODELE REGRESYJNE UPRAWIANIA TURYSTYKI AKTYWNEJ PRZEZ OSOBY NIEPEŁNOSPRAWNE WZROKOWO I RUCHOWO

REGRESSION MODELS OF ACTIVE TOURISM PRACTICED BY PERSONS WITH VISUAL AND MOTORIC DISABILITIES

KRZYSZTOF KAGANEK*, HELENA STANUCH**

* *Instytut Turystyki i Rekreacji, Akademia Wychowania Fizycznego w Krakowie*
** *Zakład Bioinformatyki i Telemedycyny, Colleqium Medicum Uniwersytetu Jagiellońskiego, mystanuc@cyf-kr.edu.pl*

Streszczenie. Jednym ze sposobów pomocy osobom niepełnosprawnym jest włączenia ich w życie społeczne. Jednym z elementów tego życia jest wypoczynek i turystyka. Najbardziej terapeutyczną z punktu widzenia szeroko pojętej rehabilitacji jest turystyka aktywna, do uprawiania której są potrzebne określone kwalifikacje, umiejętności oraz sprzęt. Niniejsze opracowanie jest próbą poszerzenia wiedzy w tym zakresie, wypełnienia istniejącej luki, jak również próbą zastosowania rozwiązań modelowych w omawianym zagadnieniu.
Praca ta zawiera dwa modele regresyjne czynników sprzyjających uprawianiu turystyki aktywnej przez osoby niepełnosprawne wzrokowo i ruchowo. Podstawę do opracowania tych modeli stanowiły dane z wywiadu przeprowadzonego wśród osób niepełnosprawnych z dysfunkcjami, a ukierunkowanych na cel badań. Badania przeprowadzono w 2003 i 2004 r. wśród osób niepełnosprawnych z terenu całej Polski. Do wyodrębnienia czynników istotnie wpływających na uprawianie turystyki aktywnej w badanych rodzajach niepełnosprawności zastosowano model regresji logistycznej.

Słowa kluczowe: model regresji logistycznej, niepełnosprawność, turystyka aktywna

Abstract. One of the ways to help disabled people is to incorporate them into social life. One such way is recreation and tourism. Active tourism has the best therapeutic value as a rehabilitation process, but requires specific qualifications, skills and equipment. The aim of this paper is to try to broaden the knowledge in this respect, fill the existing information gap, and find suitable model solutions.
Two regression models of factors favorable to active tourism of disabled persons with visual and motoric disabilities are presented. These models were based on interviews with the persons with disabilities pertaining to the scope of the research. Experiments were conducted in the years 2003 and 2004 among disabled from all over Poland. To distinguish factors having a significant influence on active tourism in the analyzed types of disability, the logistics regression model has been applied.

Key words: logistics regression model, disabled, active tourism

1. Wstęp

Powszechność uprawiania w dzisiejszych czasach turystyki oraz jej dostępność dotyczy również osób niepełnosprawnych. Turystykę można bowiem uprawiać w każdych warunkach terenowych, w każdym okresie życia, jak również będąc niepełnosprawnym.

„Osoba niepełnosprawna to osoba, której stan fizyczny, psychiczny lub umysłowy trwale lub okresowo utrudnia, ogranicza bądź uniemożliwia wypełnianie ról społecznych, a w szczególności zdolność do wykonywania pracy zawodowej, jeżeli uzyskała orzeczenie o:

- zakwalifikowaniu przez organy orzekające do jednego z trzech stopni niepełnosprawności (znaczny, umiarkowany lub lekki),
- całkowitej lub częściowej niezdolności do pracy,
- rodzaju i stopniu niepełnosprawności (dotyczy osób, które nie ukończyły 16 roku życia)".[1]

„Kryteria orzekania o stopniach niepełnosprawności opierają się na *Międzynarodowej klasyfikacji uszkodzeń, niepełnosprawności i ograniczeń w rolach*, sformułowanej w 1980 r. przez WHO. Zgodnie z tą klasyfikacją zdefiniowano:

- uszkodzenie *(ang. impairment)* jako wszelką stratę lub wadę psychicznej, filozoficznej lub anatomicznej struktury albo czynności;
- niepełnosprawność *(ang. disability)* jako wszelkie ograniczenie lub brak wynikający z uszkodzenia – możliwości wykonywania czynności na poziomie uważanym za normalny dla człowieka;

[1] Ustawa o rehabilitacji zawodowej, społecznej oraz zatrudnianiu osób niepełnosprawnych z 1997 r.

- **upośledzenie** (ang. *handicap)* jako niekorzystną (gorszą) sytuację danej osoby, będącą wynikiem uszkodzenia lub niepełnosprawności, polegającą na ograniczeniu lub uniemożliwieniu wypełniania ról, które są uważane za normalne, biorąc pod uwagę jej wiek, płeć, czynniki kulturowe i społeczne"[6] .

Wydaje się, że dopiero te trzy określenia rozważane łącznie pozwalają w pełni zdefiniować sytuację osób niepełnosprawnych, uwzględniają bowiem zarówno stan psychofizyczny, jak i niemożność pełnienia funkcji uznanych w społeczeństwie za właściwe oraz wynikające z tego utrudnienia. Te trzy pojęcia pozwalają określić relacje osób z ograniczoną sprawnością z resztą społeczeństwa oraz ustalić fizyczne, kulturowe czy społeczne bariery, które ograniczają korzystanie przez nie z różnych struktur społecznych, w pełni dostępnych dla innych obywateli. Pozwalają również na uniknięcie schematycznego unifikowania, sprowadzania do wspólnego mianownika sytuacji życiowej wszystkich osób niepełnosprawnych.

„Cechy najczęściej przypisywane ludziom niepełnosprawnym charakteryzują ich w większości jako ludzi słabych, lękliwych, nerwowych, samotnych, niepewnych siebie, niezadowolonych z życia, ubogich. Występowanie tych cech pozostaje ze sobą w wyraźnym korelacyjnym związku, możemy zatem mówić, że stanowią one elementy stereotypu, uogólnionego obrazu osób niepełnosprawnych w społeczeństwie. Sposób ich widzenia kojarzy się więc prawie wyłącznie z niepowodzeniami życiowymi, a nie z jakimikolwiek sukcesami" [9].

Mówiąc o osobie niepełnosprawnej, należy pamiętać, że posiada ona prawa ludzkie, potrzeby, pragnienia, marzenia i aspiracje. Dąży do możliwie dużej samodzielności i oczekuje szacunku ze strony innych. W rzeczywistości ludzie niepełnosprawni często żyją poza głównym nurtem społecznym. Niejednokrotnie są zepchnięci na jego margines, a ich podstawowe potrzeby życiowe są lekceważone, w tym potrzeby turystyczne.

Badania naukowe wykazują, iż co siódma osoba jest niepełnosprawna. Co czternasta charakteryzuje się poważnym w skutkach ograniczeniem sprawności psychofizycznej. Liczby te dotyczą osób, których stan fizyczny, psychiczny lub umysłowy powoduje trwałe lub okresowe utrudnienia, ograniczenie lub uniemożliwienie wypełniania zadań, ról społecznych według przyjętych kryteriów i obowiązujących norm [8].

„Do grupy osób niepełnosprawnych zalicza się ludzi o naruszonej sprawności psychofizycznej, powodującej ograniczenie funkcjonalne sprawności lub aktywności życiowej w stopniu utrudniającym pełnienie właściwych dla nich ról społecznych. Inwalidzi, więc osoby, u których uszkodzona została w sposób trwały lub okresowy (odwracalny) sprawność fizyczna lub psychiczna w stopniu uniemożliwiającym lub znacznie utrudniającym pełnienie właściwych ról społecznych, stanowią cześć populacji osób niepełnosprawnych – około 4% ogółu, tj. 1,7 tys. osób" [7]. W tej sytuacji trudno traktować społeczność ludzi niepełnosprawnych jako marginalną. Jedną z cennych wartości XX wieku jest racjonalnie pojęta troska o niepełnosprawnych, których prawo do pełnego życia, kształcenia i pracy oraz udziału w życiu społecznym zostało określone w art. 25 Powszechnej Deklaracji Praw Człowieka. Jednym z elementów pełnego życia, w tym i społecznego, jest możliwość uprawiania turystyki. Możliwości i prawa są tym, o co muszą walczyć wszyscy niepełnosprawni. Możliwość wyjazdów i podróżowania powinna być prawem każdej osoby niepełnosprawnej.

Turystykę rozumie się jako sposób bycia każdego współczesnego człowieka, który opuszcza czasowo miejsce stałego zamieszkania w różnych celach. Szczególnie ważną funkcją turystyki jest funkcja wychowawcza w stosunku do turysty, funkcja kształtowania jego osobowości, w której to funkcji można wyodrębnić cztery elementy:

a) potęgowanie zdrowia (profilaktyka i rehabilitacja),
b) poznawanie świata (przyrody, kultury i ludzi),
c) kształtowanie postaw,
d) integracja ze środowiskiem społecznym.

Turystyka osób niepełnosprawnych wg M. Weissa, to: „turystyka osób w różnym wieku, wykazujących dysfunkcje ustrojowe" [11].

Turystyka dawno już przestała być domeną ludzi zdrowych. W wielu krajach świata, w tym i w Polsce, zaczęto się interesować turystyką jako jedną z metod usprawniania ludzi dotkniętych różnego rodzaju kalectwem. Ludzie „sprawni inaczej" uprawiają wiele dyscyplin turystycznych, np. wędrują pieszo, jeżdżą rowerami, pływają kajakami, łodziami żaglowymi, uprawiają turystykę motorową, narciarstwo itp. O formach turystyki uprawianej przez osoby „sprawne inaczej" będzie mowa w dalszej części pracy.

J. Wolski, jeden z głównych propagatorów uprawiania turystyki przez osoby niepełnosprawne, uważa, iż jest ona istotnym elementem rehabilitacji tych osób, którego celem jest przywrócenie możliwie maksymalnej sprawności fizycznej, psychicznej, społecznej i zawodowej oraz umożliwienie stopniowego przystosowania się do normalnego życia. Jest wzbogaceniem i przedłużeniem terapii ruchowej, rozbudza też zainteresowanie działalnością poznawczą" [12].

Turystyka jako forma rehabilitacji jest ukierunkowana nie tylko na usuwanie lub ograniczanie zasięgu dysfunkcji, włącza też ludzi chorych w krąg kultury fizycznej i jej dobroczynnego wpływu na ciało i psychikę.

Aby w szerszym zakresie umożliwić osobom o specjalnych potrzebach korzystanie z turystyki, należy uwzględnić preferowane rodzaje turystyki oraz potrzeby wynikające z sytuacji danej kategorii osób. Z pewnością funkcje profilaktyczne i rehabilitacyjne będą szczególnie ważne. Jednak zawsze turystyka umożliwia rozszerzenie indywidualnych możliwości człowieka, co jest istotnym czynnikiem rozwoju osobniczego. Turystyka osób o specjalnych potrzebach pozwala sprawdzić samego siebie, sprzyja przekraczaniu swoich możliwości. Szczęście człowieka polega na pokonywaniu coraz to większych trudności, a uprawianie turystyki stwarza znakomitą ku temu okazję.

Każdy człowiek potrzebuje wypoczynku i odnowy sił po pracy. Inwalidzi oczekują dodatkowo ćwiczeń usprawniających i zajęć rehabilitacyjnych poprawiających nie tylko ich sprawność i wydolność fizyczną, ale i psychiczną. Turystyka może być i jest ważnym czynnikiem rehabilitacji osób niepełnosprawnych. Wydaje się ona niezbędnym elementem ich życia, potrzebnym do osiągnięcia sprawności umożliwiającej normalną egzystencję.

Turystyka jest dziedziną rekreacyjno-poznawczą działalności ludzkiej, która wykorzystywana w rehabilitacji pełni nie tylko doniosłą rolę w zaspokajaniu potrzeb ruchowych, ale

również psychicznych i intelektualnych. Jest ona bowiem formą aktywności związaną ściśle z krajoznawstwem, staje się pasją intelektualną ukierunkowaną na poznawanie i odkrywanie coraz to nowych rzeczy i zjawisk, miejsc, regionów i krajów. Spacer, wycieczka czy też uprawianie jakiejś z form turystyki kwalifikowanej, poza uaktywnieniem fizycznym dostarczają wrażeń płynących z otoczenia, motywujących do dalszej aktywności fizycznej, jednocześnie otwierając na nowo horyzont życiowy, zamknięty czterema ścianami pokoju. Ponadto uprawianie turystyki podnosi sprawność i zaradność życiową, stanowi „furtkę" do kontaktów społecznych, do uczestnictwa w naturalnych życiowych sytuacjach, umożliwia sprawdzenie się w zakresie umiejętności praktycznego działania i społecznego zachowania się w środowisku.

Turystyka dysponuje tak wielką ilością form, że właściwie nie może się zdarzyć, aby osoba niepełnosprawna nie mogła wybrać dostępnej dla siebie dziedziny działalności. Powodem, dla którego środowisko inwalidów zaakceptowało turystykę jako formę zajęć terapeutycznych, jest jej atrakcyjność, możliwość wyrwania się z izolacji oraz walory poznawcze.

„Turystyka jest dziedziną działalności ludzi niepełnosprawnych dającą duże pozytywne możliwości rehabilitacji i jest u nas wciąż za mało stosowana" – stwierdza W. Dega [3]. Podobnie jak sport, turystyka powinna pozostawać w zgodzie z programem rehabilitacji osób niepełnosprawnych. Jeżeli ma służyć interesom tych osób, musi być również odpowiednio dobrana do ich możliwości fizycznych i psychicznych, musi być bezpieczna dla uczestników i ich otoczenia oraz powinna być powszechna.

Nie wszystkie osoby z problemami zdrowotnymi wymagają usprawnienia leczniczego, ale wszyscy wymagają podnoszenia ogólnej sprawności oraz kształcenia nawyku wykorzystywania czasu wolnego. W procesie tym ważną rolę spełnia turystyka dobrana formą do różnego wieku inwalidów, rodzaju niesprawności i możliwości psychofizycznych. Turystyka i krajoznawstwo, uprawiane przez osoby niepełnosprawne, pozwalają wypoczywać w sposób aktywny, łączyć proces poznawania przez bezpośredni kontakt z odwiedzanymi terenami i obiektami. Dla osób z dysfunkcją wzroku oraz dysfunkcją narządu ruchu istotne jest wyrabianie współuczestniczącej postawy w procesie leczenia i rehabilitacji, czynnej rekonwalescencji i reaktywacji, ogólne podniesienie sprawności ruchowej, uniezależnianie się od pomocy innych osób, kształtowanie przekonania o własnej sprawności i przydatności życiowej.

Niezależnie od usprawnienia psychicznego, najważniejsze zadania turystyki to: wydobycie ludzi niepełnosprawnych z izolacji społecznej, jaką często sami sobie narzucają, przerwanie monotonii, dostarczenie przeżyć psychicznych, pobudzenie chęci do życia, do podejmowania ćwiczeń specjalistycznych w procesie rehabilitacji, dostarczenie rozrywki i relaksu.

Najbardziej terapeutyczną z punktu widzenia szeroko pojętej rehabilitacji jest turystyka aktywna, która rozwija fizycznie, zdrowotnie, etycznie, estetycznie, moralnie, politechnicznie. Definicję turystyki aktywnej można najprościej przedstawić w następującej formie: jest to forma turystyki do uprawiania, w której są potrzebne określone kwalifikacje, umiejętności oraz sprzęt, taki jak np. narty, kajak, rower itp. Istnieje wiele czynników, które warunkują uprawianie turystyki aktywnej przez osoby niepełnosprawne, jednak brak jest opracowań, które w sposób szczegółowy zajmują się tym tematem i określaniem, które z nich są mniej lub bardziej ważne. Dlatego też niniejsze opracowanie jest próbą poszerzenia wiedzy w tym zakresie, wypełnienia istniejącej luki, jak również próbą zastosowania rozwiązań modelowych w omawianym temacie. Nie są do końca znane czynniki warunkujące uprawianie turystyki aktywnej przez osoby niepełnosprawne. Przedstawione rozwiązanie modelowe może przyczynić się do szerszego rozpoznania tych czynników, wskazania ich roli w podejmowaniu aktywności turystycznej przez osoby niepełnosprawne.

2. Materiał

W celu zdiagnozowania czynników warunkujących uprawianie turystyki aktywnej przez osoby niepełnosprawne przeprowadzono wywiad wśród 322 osób niepełnosprawnych wzrokowo i 411 osób niepełnosprawnych ruchowo. Badania przeprowadzono w 2003 i 2004 roku w organizacjach inwalidzkich zrzeszających osoby z wymienionymi dysfunkcjami. Zakres terytorialny badań obejmował województwa: małopolskie, podkarpackie, lubelskie, świętokrzyskie, środkowopomorskie, zachodniopomorskie.

Dla celów analizy i tworzenia modelu wyodrębniono następujące cechy: wiek, posiadanie samochodu, tradycje turystyczne w rodzinie, pomoc innych osób, pomoc organizacji, ocena sytuacji materialnej. Z analizy literatury przedmiotu oraz własnych badań pilotażowych wynikało, że wymienione cechy mają wpływ na poziom aktywności turystycznej osób niepełnosprawnych w zakresie aktywnych form turystyki. Pod uwagę wzięto również dane dotyczące znaczenia dla badanego poszczególnych funkcji turystyki, tj.: zdrowotnej, rehabilitacyjnej, integracyjnej, wypoczynkowej, kształceniowej, wychowawczej i poznawczej. Funkcje te oceniane były jako: bardzo ważna, ważna, mało ważna.

3. Metoda analizy

Modele regresyjne umożliwiają obliczenie wartości teoretycznej zmiennej zależnej, jeżeli znane są wartości zmiennych niezależnych występujących w danym modelu. Rozróżniane są różne typy modeli regresyjnych w zależności od charakteru zmiennych, ich rozkładów czy typu zależności funkcyjnych. Jednym z takich modeli, jest model liniowej regresji zapisanej macierzowo:

$$Y = X\beta + e, \tag{1}$$

gdzie: $Y = \begin{bmatrix} Y_1 \\ \cdot \\ \cdot \\ \cdot \\ Y_m \end{bmatrix}$; $\beta = \begin{bmatrix} \beta_0 \\ \cdot \\ \cdot \\ \cdot \\ \beta_p \end{bmatrix}$;

$$X = \begin{bmatrix} 1, x_{11}, \ldots, x_{1p} \\ \cdot \\ \cdot \\ \cdot \\ 1, x_{m1}, \ldots, x_{mp} \end{bmatrix}; \; e = \begin{bmatrix} e_1 \\ \cdot \\ \cdot \\ \cdot \\ e_m \end{bmatrix}$$

m – liczba obiektów (badanych osób),
p – liczba zmiennych niezależnych,
x_{ij} – wartość j-*tej* zmiennej niezależnej dla i-*tego* obiektu

e_i – zmienne losowe nieskorelowane i $E(e_i)=0$; $Var(e_i)=\sigma^2$
Y_i – wartość zmiennej zależnej dla i-*tego* obiektu; przy czym $E(Y) = X\boldsymbol{\beta}$; $Var(Y)=\sigma^2 I$.

W przypadku, kiedy zmienna zależna *Y* jest typu dychotomicznego, takiego jak uprawianie aktywnej turystyki *(1)* lub nie *(0)*, to do opisu zależności jej od zmiennych niezależnych można zastosować model regresji logistycznej [4], [10].

W modelu tym zakłada się, iż logarytm stosunku prawdopodobieństwa warunkowego wystąpienia danego zjawiska, czyli prawdopodobieństwa uprawiania turystyki aktywnej przy konkretnych wartościach zmiennych niezależnych $x_1, \ldots, x_p$, a opisanego symbolem $P(Y=1/ x_1, \ldots, x_p)$, a w formie skróconej $P(Y=1/x)$, do prawdopodobieństwa warunkowego niewystąpienia tego zjawiska, opisanego przez $1-P(Y=1/x)$, jest funkcją liniową zmiennych niezależnych. Co można zapisać:

$$u(x) = \ln \frac{P(\mathrm{Y}=1/\mathbf{x})}{1-P(\mathrm{Y}=1/\mathbf{x})} = \mathbf{x}'\boldsymbol{\beta}, \qquad (2)$$

przy czym $x=[1, x_1, \ldots, x_p]'$,
gdzie *p* – ilość zmiennych niezależnych,
$\boldsymbol{\beta}$ – wektor *p+1* współczynników regresji.

Po przekształceniu (2) otrzymano model regresji logistycznej, określający zależność zmiennej zależnej *Y* od zmiennych niezależnych x:

$$P(Y=1/x) = \frac{e^{u(\mathbf{x})}}{1+e^{u(\mathbf{x})}}. \qquad (3)$$

Estymatory $\tilde{\boldsymbol{\beta}}$ parametrów $\boldsymbol{\beta}$ tego modelu szukane są metodą największej wiarogodności. W tym przypadku funkcja wiarogodności jest postaci:

$$l(\boldsymbol{\beta}) = \prod_{i=1}^{m} P(Y=1/\mathbf{x_i})^{y_i}[1-P(Y=1/\mathbf{x_i})]^{1-y_i}, \qquad (4)$$

gdzie $x_i=[1,x_{i1}, \ldots, x_{ip}]'$.

Maksymalizując logarytm tej funkcji ze względu na $\beta_0, \ldots \beta_p$, otrzymuje się do rozwiązania układ równań wiarogodności, które po rozwiązaniu dają wzory na estymatory $\tilde{\boldsymbol{\beta}}$. Estymatory te umożliwiają obliczenie estymatora prawdopodobieństwa warunkowego zajścia zdarzenia $\{Y=1\}$ przy odpowiednich wartościach $x=[1, x_1, \ldots, x_p]$, zapisanego jako $\tilde{P}(Y=1/\mathbf{x})$.

Weryfikację hipotezy zerowej $H_0: \beta_j=0$, kontra alternatywnej $H_1: \beta_j \neq 0$ dla j=0, p przeprowadzono testem Wald'a:

$$W = \tilde{\boldsymbol{\beta}}'(\mathbf{X}'\mathbf{V}\mathbf{X})\tilde{\boldsymbol{\beta}}, \qquad (5)$$

dla którego funkcja testowa W ma przy prawdziwości H_0 rozkład zgodny z rozkładem χ^2 dla ν=p+1 stopni swobody, przy czym:

$$\tilde{\boldsymbol{\beta}} = \begin{bmatrix} \tilde{\beta}_0 \\ . \\ . \\ . \\ \tilde{\beta}_p \end{bmatrix}, \quad X = \begin{bmatrix} 1, x_{11}, \ldots, x_{1p} \\ . \\ . \\ . \\ 1, x_{m1}, \ldots, x_{mp} \end{bmatrix},$$

$$V = \begin{bmatrix} \tilde{P}(Y=1/\mathbf{x_1})(1-\tilde{P}(Y=1/\mathbf{x_1})), . & . & 0 \\ . & & \\ . & & \\ 0, & .. & \tilde{P}(Y=1/\mathbf{x_m})(1-\tilde{P}(Y=1/\mathbf{x_m})) \end{bmatrix},$$

Dla estymatorów $\beta_0, \ldots \beta_p$ można wyznaczyć przedziały ufności. Wartość e^{β_i} określana jest ilorazem szans (*OR* – odd ratio). Wartość ilorazu szans (*OR*) dla i-*tej* cechy odpowiada wartości e^{β_i} i określa udział tej cechy w wystąpieniu badanego zjawiska w analizowanym modelu. Jeżeli *OR*>0, czyli β_i,>0, to wystąpienie i-*tej* cechy sprzyja wystąpieniu badanego zjawiska, czyli ma pozytywny wpływ w modelu. I na odwrót, jeżeli *OR*<0 (β_i,<0), to wystąpienie i-*tej* cechy nie sprzyja wystąpieniu badanego zjawiska, czyli ma negatywny wpływ w modelu. Obliczany może być także 95% przedział ufności dla *OR* i-*tej* cechy w modelu. Jeżeli przedział ten zawiera cyfrę *1*, to oznacza, że i-*ta* cecha nie ma istotnego wpływu na wystąpienie badanego zjawiska w analizowanym modelu ($e^0=1$) i wniosek ten pokrywa się ze stwierdzeniem, iż współczynnik β_i nie różni się istotnie od zera.

Adekwatność danego modelu do danych empirycznych oceniano testem chi-kwadrat zgodności [5], [1]. Dodatkowo oceniano dopasowanie modelu do danych empirycznych przez obliczenie czułości i swoistości danego modelu. Parametry te wyliczane były w oparciu o dane z tab. 1 wg wzorów:

$$\text{Czułość} = \frac{a}{a+c}100\%, \qquad (6)$$

$$\text{Swoistość} = \frac{d}{b+d}100\%.$$

Tab. 1. Parametry do obliczania czułości i swoistości modelu

Kwalifikacja na podstawie modelu	Uprawiający turystykę aktywną	
	Tak	Nie
$\tilde{P}(Y=1/\mathbf{x}) \geq 0{,}5$	a	b
$\tilde{P}(Y=1/\mathbf{x}) < 0{,}5$	c	d
Razem	a+c	b+d

4. Wyniki obliczeń

Do obliczeń parametrów regresji logistycznej wzięto 13 cech ($x_1,\ldots,x_{13}$). Wartości obliczonych dla tych cech współczynników regresji $\tilde{\beta}_j$, dla $j = 0, \ldots, 13$ przedstawiono w tab. 2. Gwiazdkami oznaczono te współczynniki $\tilde{\beta}_j$, które różniły się istotnie od zera. Weryfikację przeprowadzono na poziomie istotności α=0,05.

Widać więc, że u osób niepełnosprawnych wzrokowo jedynie funkcja poznawcza odgrywa istotne znaczenie w uprawianiu turystyki aktywnej (p<0,01).

Tab. 2. Współczynniki regresji logistycznej w badanych rodzajach niepełnosprawności

cechy Symbol	Cechy	Niepełnosprawność	
		Wzrokowa	Ruchowa
		$\tilde{\beta}_j$	
x_1	Wiek	0,002	0,018*
x_2	Ocena sytuacji materialnej	-0,207	0,127
x_3	Pomoc innych osób	0,424	0,707**
x_4	Pomoc organizacji	0,249	0,612
x_5	Tradycje turyst. w rodzinie	0,212	0,795***
x_6	Posiadanie samochodu	-0,179	0,622**
x_7	Funkcja zdrowotna	0,389	0,154
x_8	Funkcja rehabilitacyjna	-2,99	-0,273
x_9	Funkcja integracyjna	0,245	0,175
x_{10}	Funkcja wypoczynkowa	-0,353	-0,430**
x_{11}	Funkcja kształceniowa	0,031	-0,090
x_{12}	Funkcja wychowawcza	0,016	-0,174
x_{13}	Funkcja poznawcza	0,387**	0,105
	Stała	-2,008	-4,730***

*** p<0,001 ** p<0,01; *p<0,05;

U niepełnosprawnych ruchowo istotne znaczenie ma wiek (p<0,05), pomoc osób innych (p<0,01), a przede wszystkim tradycje turystyczne w rodzinie (p<0,001). Istotne znaczenie w uprawianiu turystyki aktywnej ma również funkcja wypoczynkowa (p<0,01).

Ostatecznie w obu grupach niepełnosprawności przeprowadzono ponownie obliczenia parametrów regresji tylko dla tych cech, które miały istotny wpływ na uprawianie turystyki aktywnej. Obliczono też parametry oceniające adekwatność otrzymanych modeli.

Tab. 3. Parametry modelu regresji logistycznej w niepełnosprawności wzroku

Symbol	Cechy	$\tilde{\beta}_j$	95% przedział dla ilorazu szans
x_{13}	Funkcja poznawcza	0,373 **	1,135 – 1,857
	Stała	-0,830	
Czułość modelu =72,3%, Swoistość modelu=40,3%			
Wartość p =0,003 (dla modelu)			

** p<0,01

Wyniki ponownych obliczeń dla niepełnosprawnych wzrokowo przedstawiono w tab. 3. Widać, że 95% przedział ufności dla funkcji poznawczej obejmuje wartości powyżej 1 (*OR*=1,5), czyli czynnik poznawczy dla tych osób sprzyja uprawianiu turystyki aktywnej. Pozostałe cechy u tych osób nie mają istotnego wpływu na wybór turystyki aktywnej. Wartość p oceny istotności modelu właściwie pokrywa się z oceną istotności współczynnika $\tilde{\beta}_j$ (*dla j=13).* Zatem czynnikiem warunkującym uprawianie turystyki aktywnej wśród osób z dysfunkcją wzroku jest jedynie funkcja poznawcza, co można zapisać funkcyjnie:

$$u(x)=0,37x_{13}-0,83 \qquad (7)$$

Tab. 4. Parametry modelu regresji logistycznej w niepełnosprawności ruchowej

Symbol	Cechy	$\tilde{\beta}_j$	95% przedział dla ilorazu szans
x_1	Wiek	0,019*	1,005-1,034
x_3	Pomoc innych osób	0,670**	1,277-2,992
x_5	Tradycje turyst. w rodzinie	0,776***	1,382-3,416
x_6	Posiadanie samochodu	0,639**	1,250-2,874
x_{10}	Funkcja wypoczynkowa	-0,359**	0,547-0,893
Stała		-3,745***	
Czułość modelu = 68,7%, Swoistość modelu = 56,7%			
Wartość p =0,001 (dla modelu)			

*** p<0,001 ** p<0,01; *p<0,05;

Wyniki obliczeń dla modelu zawierającego tylko cechy istotnie wpływające na uprawianie turystyki aktywnej dla niepełnosprawnych ruchowo przedstawiono w tab. 4. Model dla osób niepełnosprawnych ruchowo jest:

$$u(x)=0,02x_1+0,67x_3+0,78x_5.+0,64x_6-0,36x_{10}-3,75 \qquad (8)$$

Czynnikami warunkującymi uprawianie turystyki aktywnej wśród osób z dysfunkcją ruchu są: wiek, pomoc innych osób, tradycje turystyczne w rodzinie, posiadanie samochodu. Należy zwrócić uwagę na to, iż dla funkcji poznawczej 95% przedział ufności dla ilorazu szans obejmuje wartości poniżej 1 (*OR*=0,7), co świadczy o tym, iż osoby z dysfunkcją ruchową uprawiające turystykę aktywną oceniają funkcję wypoczynkową jako mało ważną, czyli mającą negatywny wpływ na uprawianie turystyki aktywnej przez te osoby (można więc przypuszczać, że przyczyną tego może być to, iż turystyka aktywna wymaga przygotowania psychofizycznego oraz odpowiedniej kondycji).

5. Wnioski

Opracowane modele pozwalają określić konkretne cechy implikujące uprawianie turystyki aktywnej w badanych grupach.

6. Literatura cytowana

1. Akaike H. (1973): Information Theory and an Extension of the Maximum Likelihood Principle. In second International on Information Theory, 267-281, Budapest, Akademia Akaido.
2. Cambel M. J., Machin D.: Medical Statistics, Wiley,1992.
3. Dega W.: Sport i Turystyka jako czynnik rehabilitacji osób poszkodowanych na zdrowiu, w: Możliwości uprawiania sportu przez inwalidów, Materiały z konferencji, Warszawa 27 maja 1970, PZWL, 1971.
4. Hosmer D. W., Lemeshow S.: Applied Logistic Regression, Wiley 1989.
5. Jones R., Ford P., Hamman R.: Seasonality Comparisons among Groups Using Incidence Data, Biometrics, 1988, 44, 1131-1144.
6. Kinalski R.: Kompendium rehabilitacji i fizjoterapii, WM Urban & Partner, Wrocław 2002.
7. Łobożewicz T.: Turystyka i rekreacja ludzi niepełnosprawnych, WSE w Warszawie 2000.

8. Ogólnopolski informator osoby niepełnosprawnej, artykuł mgr. M. Prokopiuk: Kultura fizyczna w procesie rehabilitacji osób niepełnosprawnych, Wydawnictwo Infopress s.c.
9. Ratyński W.: Problemy i dylematy polityki społecznej w Polsce, T.2, Difin, Warszawa 2003.
10. Stanuch H., Urbanik A.: Zastosowanie regresji logistycznej do oceny wyników spektroskopii rezonansu magnetycznego w diagnostyce otępienia o typie Alzheimera, Colloqium Biometryczne, 34a: 171-180, 2004.
11. Weiss M.: Możliwości uprawiania turystyki przez osoby niepełnosprawne, Instytut Turystyki, Warszawa 1979.
12. Wolski J.: Profilaktyczne leczenie i rehabilitacyjne funkcje turystyki. Cz. I Aspekty fizjologiczne i psychologiczne, Warszawa 1979.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 227-232.

MODELOWANIE POZIOMU EKSPRESJI GENU TNFα, TNF RI, TNF RII W KOMÓRKACH RAKA JELITA GRUBEGO

MODELING OF TNFα, TNF RI, TNF RII GENE EXPRESSION LEVEL IN CELLS OF COLON CANCER

TERESA KOKOT[*], EWA NOWAKOWSKA-ZAJDEL[*], JANUSZ WIERZGOŃ[**], ANDRZEJ KOZOWICZ[*], MAŁGORZATA MUC-WIERZGOŃ[*]

[*]*Katedra i Oddział Kliniczny Chorób Wewnętrznych Śląskiej Akademii Medycznej, Bytom*
[**] *Klinika Chirurgii Onkologicznej, Centrum Onkologii im. M. Skłodowskiej-Curie, Oddział Gliwice*

Streszczenie. W pracy zaproponowano wykorzystanie elementów dynamiki nieliniowej i symulacji komputerowej jako zupełnie nowego spojrzenia na biologiczną rolę czynnika martwicy nowotworu i jego receptorów w raku jelita grubego.
Celem pracy była:
1. ocena dynamiki zmian w czasie poziomu ekspresji genów TNF, TNF RI, zmienionych węzłach chłonnych,
2. poszukiwanie praw rządzących zachowaniem (z reguły dynamicznym i nieliniowym) tych parametrów w chorobie nowotworowej,
3. przedstawienie wzajemnych relacji TNF/TNF R za pomocą układu trzech nieliniowych równań różniczkowych,
4. opracowanie modelu matematycznego przebiegów czasowych zmian poziomów ekspresji genów tej cytokiny i jej receptorów.

Badaniami objęto 54 chorych na raka jelita grubego w III stopniu zaawansowania klinicznego nowotworu wg klasyfikacji TNM leczonych operacyjnie. U wszystkich pacjentów na podstawie pooperacyjnego badania histopatologicznego rozpoznano gruczolakoraka (Adenocarcinoma) jelita grubego. Model matematyczny opracowano na podstawie wyników badań klinicznych dotyczących analizy jakościowej i ilościowej genów TNFα, TNF RI, TNF RII oraz produktów ich ekspresji. Miarą ekspresji badanych genów TNFα, TNF RI oraz TNF RII była liczba kopii mRNA w przeliczeniu na 1μg całkowitego RNA. Na podstawie uzyskanych wyników stwierdzono, że w komórkach raka jelita grubego, w przerzutowo zmienionych węzłach chłonnych oraz okolicznej tkance niezmienionej chorobowo obserwuje się zwiększony, ale zmienny poziom ekspresji genu TNFα, TNF RI oraz TNF RII. Największa liczba kopii mRNA TNFα obecna jest w komórkach zmienionych przerzutowo węzłów chłonnych, co sugeruje największą (najsilniejszą) ekspresję genu TNF w badanym materiale.
W opracowanym modelu matematycznym każdy analizowany układ zachowuje się inaczej. Pobudzenie komórek niezmienionych chorobowo nie wytrąca ich ze stanu równowagi koniecznego do utrzymania homeostazy organizmu. Komórki guza w zależności od poziomu ekspresji genów TNF, TNF RI, TNF RII zachowują się jak oscylatory drgające w sposób chaotyczny. W przerzutowo zmienionych węzłach chłonnych obniżaniu poziomu ekspresji genu TNF towarzyszy wzrost poziomu ekspresji jego receptorów, co może świadczyć o zachowanym ujemnym sprzężeniu zwrotnym w tym układzie.
Uzyskane wyniki pozwalają na sformułowanie następujących wniosków:
1. W raku jelita grubego występują podwyższone poziomy ekspresji genów TNF, TNF RI, TNF RII na komórkach guza i tkankach sąsiednich,
2. Zastosowanie układu trzech równań różniczkowych do analizy prognostycznej zmian poziomu ekspresji tych genów pozwala na opracowanie modelu matematycznego ilustrującego dynamiczne i nieliniowe zachowanie parametrów w czasie,
3. Charakter uzyskanych rozwiązań jest determinowany przez współczynniki równań, rząd równań oraz nieliniowości i warunków początkowych,
4. Analiza prognostyczna uzyskanych rozwiązań pozwala na weryfikację kliniczną i teoretyczną danych, a także może stanowić element prognozowania choroby.

Słowa kluczowe: rak jelita grubego, ekspresja genów TNF, TNF RI, TNF RII

Abstract. Genetic disregulation of gene coding TNF receptors was observed in neoplastic disease by increased number of receptors on tumour cells and ligand – receptor activity. It might cause tumour proliferation and metastatic potential.
Structure of TNF receptors influences TNF activity in vivo and structure of TNF RI and TNF RII genes may suggest post-transcription modification based on alternative splicing.
In this study elements of non-linear dynamic and computer simulation as a new proposal of biological role of tumour necrosis factor and its receptors of colorectal cancer were proposed.
The aim of this study was
1. the analysis of dynamics in time of gene expression TNF, TNF RI, TNF RII in tumour cells, unchanged tissue and in metastatic lymphnodes,
2. search for the rights governing the behaviour of these parameters (dynamic and non-linear)
3. presentation of the relation of TNF/TNF R using three non-linear differential equations
4. proposal of a mathematic model of time changes in level expression of cytokine and its receptors

The study included 54 patients with histopathological confirmed adenocarcinoma (Stage III according to the AJC TNM Classification). Mathematical model was analysed on the basis of clinical examination (TNF α, TNF RI, TNF RII). Examinations were conducted in biologic molecular laboratory of Dept. of Pharmacy in Sosnowiec. The measurement of gene expression TNFα, TNF RII was the number of copies mRNA in correlation to 1μg total RNA.
Based on the results of this analysis it was confirmed that in colorectal cancer cells, metastatic lymphodes and surrounding unchanged tissues increased but changeable expression level of TNFα, and TNF RII is observed. The biggest number of mRNA copies TNFα was observed in metastatic lymphnodes cells, which suggests the strongest TNF gene expression in analysed material.
In this mathematical model every analysis is different. The stimulation of unchanged cells seems not to change the balance of organic homeostasis. Tumor cells in correlation to gene expression level (TNF, TNF RII) tend to reach a stable state or behave as oscillators moving in chaotic way.
In metastatic lumphnodes decreasing the level of TNF is correlated with increasing of receptors expression level, which could be the evidence of negative feedback in that system.
The results of this analysis allow to conclude:
1. In colorectal cancer increased expression gene levels of TNF, TNF RII in tumour cells and in surrounding tissues.
2. Application of a system of three differential equations to the prognostic analysis of expression level changes of these genes allows working out mathematic model illustrating dynamic and non-linear behaviour of parameters in time.
3. Character of obtained solutions is determined by the coefficient and order of equations as well as nonlinearity and initial conditions.
4. Prognostic analysis of this results allows to make clinical and theoretical data

Key words: colon cancer, TNFα, TNF RI, TNF RII genes expressiongnostic analysis of this results allows to make clinical and theoretical data

1. Wstęp

Złożoność procesów immunologicznych współuczestniczących w powstawaniu, rozwoju i funkcjonowaniu organizmu w stanie zdrowia i choroby to niezwykle skomplikowany proces, z siecią wzajemnych interakcji amplifikujących lub działających negatywnie w jego obrębie. Towarzyszą mu (jak większości zjawisk występujących w przyrodzie) pewne znamiona nieokreśloności, które mogą pochodzić albo z losowej natury tych zjawisk, lub ich nieprecyzyjnego opisu [2]. Nieokreśloność w układzie immunologicznym obecna jest zarówno na poziomie makro-, jak i mikromolekularnym, obejmując również swoim zasięgiem niskocząsteczkowe białka lub glikoproteiny zwane cytokinami [2], [3].

Badania nad cytokinami, prowadzone w ciągu ostatnich lat, nie tylko pozwoliły na zrozumienie ich szczególnej roli w koordynacji wielu procesów zachodzących w organizmie ludzkim, ale stworzyły także nowe możliwości poznania dróg informacji międzykomórkowej oraz przenoszenia sygnałów wewnątrz komórki.

Wiemy, że w układzie immunologicznym oraz pomiędzy nim a innymi układami oddziaływania cytokiny zachodzą na siebie wzajemnie, tworząc dynamiczny, nieliniowy układ powiązań czynnościowych, czyli *sieć cytokin* [3].

W genetycznie zaprogramowanej sieci nie istnieje hierarchia ważności, a funkcjonujące w jej obrębie zależności mają charakter nielinearny [3].

Czynnik martwicy nowotworu (TNFα) jest jednym z istotnych elementów dynamicznego, nieliniowego układu, jakim jest sieć cytokin. Jak stwierdził Lloyd Old (1985) *„razem z innymi cytokinami tworzy on skomplikowaną sieć sygnałów składających się na reakcje immunologiczne organizmu”* [8].

W stanach fizjologicznych może pełnić rolę czynnika wzrostu, cytotoksyny, stymulatora różnicowania. Bierze udział między innymi w regulacji procesów immunologicznych, embriogenezy, hematopoezy, przemian metabolicznych, regulacji snu.

Działanie czynnika martwicy nowotworu udowodniono w patogenezie wstrząsu septycznego, schorzeń o charakterze zapalnym, chorób autoimmunologicznych, reakcjach odrzucenia przeszczepu, ostrej i przewlekłej niewydolności oddechowej, zastoinowej niewydolności krążenia, przewlekłej niewydolności krążenia i innych. Jego aktywność zależy od rodzaju komórek docelowych i obecności innych czynników regulacyjnych, natomiast efekt działania biologicznego – od specyficznych receptorów, określanych jako p55 (typ I, TNF R55) i p75 (typ II, TNF R75) [4].

Receptory TNF obecne są praktycznie na wszystkich komórkach ssaków.

Ze względu na szczególne znaczenie fizjopatologiczne TNFα w procesach odpornościowych oraz w procesach nadzoru immunologicznego, istotne miejsce wśród stanów chorobowych wykazujących zmienne wartości tej cytokiny i jej receptorów zajmują *choroby nowotworowe*. *Dysregulacja genetyczna* charakterystyczna dla choroby nowotworowej dotyczy również genów receptorów TNF [7], w następstwie czego obserwuje się zwiększenie liczby receptorów na powierzchni komórki nowotworowej, często sięgające kilkuset tysięcy na jedną komórkę. Występuje także wysoka aktywność przylegania kompleksu ligand/receptor, co nawet przy prawidłowej produkcji cytokiny może być czynnikiem przyspieszającym proliferację. Liczne badania doświadczalne i kliniczne wskazują na udział tego kompleksu zarówno w hamowaniu rozwoju komórek guza, jak i w niektórych przypadkach we wzroście nowotworu i tworzeniu przerzutów (np. czerniak złośliwy) [7].

Obecnie nie wiadomo jednoznacznie, w jakim stopniu wahania stężeń tej cytokiny korelują z miejscowymi zmianami liczby i złuszczania jej receptorów oraz sekrecji tego białka w obrębie guza i najbliższym otoczeniu.

Stąd też bardziej dokładnymi badaniem ilustrującym jej zmienność w określonym mikrośrodowisku jest analiza poziomu ekspresji genu TNF oraz TNF RI i TNF RII w komórkach nowotworowych i ich otoczeniu.

Olbrzymia wiedza dotycząca znaczenia TNFα we wzroście, różnicowaniu komórek układu odpornościowego, możliwościach ich funkcjonowania w procesach fizjologicznych i patologicznych pozwala na wykorzystanie tych informacji do tworzenia numerycznych symulacji złożonych, nieliniowych układów biologicznych jako ewentualnej metody wspomagającej diagnostykę i postępowanie lecznicze.

Spośród chorób nowotworowych rak jelita grubego jest typowym przykładem nowotworu człowieka, w którego patogenezie i progresji wykazano zmiany genetyczne, warunkujące poszczególne etapy transformacji. Zachorowalność na ten nowotwór systematycznie wzrasta, stąd też celowe wydaje się poszukiwanie kolejnych metod, które mogłyby wspomagać diagnostykę i leczenie.

Celem pracy jest próba uzyskania odpowiedzi na następujące pytania:

- Czy można zastosować elementy dynamiki nieliniowej do oceny poziomu ekspresji genów TNF, TNF RI, TNF RII w komórkach guza, tkance niezmienionej chorobowo oraz przerzutowo zmienionych węzłach chłonnych?
- Czy przewidywana teoretycznie zmienność analizowanej cytokiny i jej receptorów w komórkach raka jelita grubego i ich najbliższym środowisku znajduje uzasadnienie w opracowanym modelu matematycznym?

2. Materiał

Badaniami objęto 52 chorych na raka jelita grubego w III stopniu zaawansowania klinicznego nowotworu wg klasyfikacji TNM *(każdy T, każdy N+, M0*) leczonych operacyjnie. U wszystkich pacjentów na podstawie pooperacyjnego badania histopatologicznego rozpoznano gruczolakoraka (*Adenocarcinoma)* jelita grubego. Wycinki do badań pobierano z następujących okolic:

a) z guza nowotworowego, z linii cięcia chirurgicznego od strony usuwanych tkanek,

b) z tkanki ocenionej makroskopowo i histopatologicznie jako niezmieniona patologicznie (oznaczana umownie w pracy jako tkanka zdrowa),

c) z okolicznych przerzutowych węzłów chłonnych.

Uzyskany materiał dzielono na dwie części. Jedna była przesyłana do pracowni histopatologicznej, druga – do pracowni biologii molekularnej.

3. Metoda

Zasadnicza koncepcja proponowanego modelu obejmuje:

Etap pierwszy

Do opracowania matematycznego wykorzystano wyniki badań klinicznych dotyczące analizy jakościowej i ilościowej genów TNFα, TNF-RI, TNF-RII oraz produktów ich ekspresji w komórkach pochodzących z: guza pierwotnego, tkanek określonych makroskopowo jako niezmienione chorobowo oraz z okolicznych, przerzutowych węzłów chłonnych.

W ekstrakcie całkowitego RNA (total RNA) wyznaczono liczbę kopii mRNA-TNF, mRNA-TNF RI oraz mRNA-TNF RII. Liczbę kopii mRNA badanych genów wyznaczano techniką QRT-PCR z zastosowaniem analizatora sekwencji ABI PRISM ™ 7700, a otrzymany wynik przeliczano na 1 μg całkowitego RNA.

Etap drugi

Celem etapu było opracowanie na podstawie uzyskanych wyników badań molekularnych matematycznego modelu zmienności poziomu ekspresji badanych genów w komórkach raka jelita grubego oraz tkankach otaczających guz, usuniętych podczas zabiegu operacyjnego. Należy podkreślić, że zgodnie z teorią chaosu, uzyskane wartości poziomu ekspresji genów TNFα, TNF-RI oraz TNF-RII traktowano jako *warunki początkowe* każdego układu (p), przewidując na ich podstawie zachowanie się tych układów w czasie.

Korzystając z podstawowych opracowań dotyczących teorii modelowania układów biologicznych oraz analizy danych układu TNF-receptor, w ostatecznej wersji do opracowania modelu matematycznego zmienności poziomu ekspresji badanych genów wykorzystano układ trzech nieliniowych równań różniczkowych :

$$dz/dt = k_z [v(t-\tau)g (t-\tau)-z(t)] \quad (1)$$

$$dv/dt = \beta\nu(t) - [v(t)f(t)]^n \quad (2)$$

$$df/dt = k_a \{\rho f(t) - [\nu(t)f(t)]^n - f(t)\}, \quad (3)$$

gdzie:

k_a, k_z – parametry sterujące szybkością przebiegających reakcji.

Punktem wyjścia do opracowania modelu odpowiedzi immunologicznej była zachodząca stale w organizmach żywych reakcja enzym-substrat. Kinetyka tej reakcji, a także charakterystyka opisujących ją równań była prezentowana między innymi przez Berga i Purcell'a, de Lisi, Potanina i wsp., Wiegela [9]. Widoczne jest to również w przeglądzie klasycznych modeli reakcji, prezentowanych przez Orlika [7], gdzie wyjaśniono typowe przejścia i zachowania relacji autokatalitycznych.

Tak więc zmiennymi dynamicznymi modelu są:

- całkowita liczba komórek układu immunologicznego i innych docelowych uczestniczących w rozwoju i ewentualnie utrzymywaniu się choroby nowotworowej,
- poziom ekspresji genu czynnika martwicy nowotworu (TNFα),
- poziom ekspresji genów i jego receptorów TNF RI, TNF RII,

czyli:

Z(t) – to liczba komórek wydzielających TNF w czasie t, gdzie t może równać się zero (jest to moment pobrania materiału operacyjnego – początek badania), czyli $Z(t_1)$ – to całkowita liczba komórek w 30 dniu od pobrania materiału do badań. W warunkach *in vivo* nie ma możliwości rozróżnienia i modelowania działania poszczególnych aktywowanych komórek. Stąd przyjęto, że jest to pewna stała, określona wartość.

V(t) – to poziom ekspresji genu TNFα w badanych komórkach w czasie t,

F(t) – to poziom ekspresji genów TNF RI i TNF RII (suma wartości) w badanych komórkach w czasie t,

t – czas; w pracy zaproponowano t jako dni, biorąc pod uwagę równocześnie trzy analizowane układy (czyli, tkanka zdrowa, guz, przerzutowe węzły chłonne) i ich przewidywane zachowanie po leczeniu operacyjnym,

t – τ czyli opóźnienie w czasie spowodowane odpowiedzią pobudzonych komórek (nowotworowych lub z ich najbliższego otoczenia), tzn. zadziałanie konkretnego antygenu nie wywołuje efektu w danej chwili czasu t, ale z opóźnieniem t– τ.

Matematyczne modelowanie ekspresji omawianych genów w raku jelita grubego oprócz trzech podstawowych zmiennych wymaga uwzględnienia szeregu pewnych określonych stałych danych, które warunkują prawidłowość przebiegu określonego zjawiska biologicznego, opisują jego zachowanie oraz ilustrują wzajemne zależności pomiędzy trzema podstawowymi parametrami. W równaniach występują one jako ustalone współczynniki równań, odzwierciedlające istniejącą biologiczną zmienność w analizowanym układzie. Proponując określone współczynniki równań, będące pewnymi wartościami średnimi, korzystano z wiedzy doświadczalnej wielu badaczy i ich prac dotyczących matematycznego modelowania układu TNF/receptor oraz charakteryzujących zachowanie cytokin w warunkach *in vitro*.

Są to:

n – wykładnik potęgowy ilustrujący potęgową zmianę (wzrost lub obniżenie) czynnika martwicy nowotworu, zależną od szybkości jego neutralizacji,

α – prawdopodobieństwo aktywacji w komórkach analizowanych genów,

β – współczynnik zmienności poziomu ekspresji genu TNFα,

γ – prawdopodobieństwo połączenia i ewentualnie neutralizacji TNFα przez receptory,

ε – współczynnik szacujący liczbę receptorów potrzebnych do związania jednej cząsteczki TNF,

σ – współczynnik zmienności poziomu ekspresji genów TNF-RI oraz TNF-RII w komórkach, określany w przeliczeniu na jedną komórkę/minutę.

Pamiętając o różnych wprowadzonych do badań wartościach liczbowych, podobnie jak w pracy [7] w celu uproszczenia, a jednocześnie uogólnienia wyników wprowadzono wielkości bezwymiarowe (*oznaczane małymi literami*). Ułatwia to matematyczne opracowanie. Wprowadzone w powyższych równaniach wielkości bezwymiarowe zostały zdefiniowane następująco:

$$z(t) = Z(t)/Z_0, \quad v(t)=V(t)/V_0, \quad g(t) = G(t)/G_0,$$
$$f(t)=F(t)/F_0$$
$$k_z = \alpha V_0 G_0/Z_0,$$
$$\gamma = (F_0V_0)^n/V_0,$$
$$\rho=\sigma/k_a Z_0/F_0, \quad k_a= \varepsilon\gamma (V_0F_0)^n /F_0$$

Informacje o zachowaniu układu przedstawiono nie w postaci kolejnych etapów rozwiązań trzech równań różnicz-

kowych i dokładnych wyników matematycznych, ale za pomocą wykresów ilustrujących przebiegi czasowe i fazowe analizowanych parametrów, co wydaje się bardziej zrozumiałe dla lekarza i klinicysty.

4. Wyniki

W komórkach guza średnie wartości liczby kopii mRNA TNFα wynosiły *43 750*, natomiast mRNA TNF RI – *2604*, a RII – *16 637*.

W tkance zdrowej średnie wartości liczby kopii mRNA TNF wynosiły *31 516*, mRNA TNF RI i mRNA TNF RII odpowiednio *5498* oraz *18 811.*

W komórkach węzła chłonnego średnie wartości liczby kopii mRNA TNF, mRNA TNF RI, mRNA TNF RII wynosiły kolejno: *123 618*; *4161*; *27 690*.

W komórkach tych obecna jest największa liczba kopii mRNA TNF, co świadczy o najsilniejszej ekspresji genu receptora TNF.

W odniesieniu do analizowanych układów: *guz, przerzutowy węzeł chłonny, tkanka zdrowa* modelowaną matematycznie zmienność poziomu ekspresji genów TNF α, TNF RI, TNF RII przedstawiono na kolejnych wykresach, uznając uzyskane laboratoryjne wartości liczby kopii mRNA TNF oraz mRNA TNF RI, mRNA TNF RII jako początkowe *(wyjściowe)* do przeprowadzenia analizy czasowej zmian poszczególnych układów.

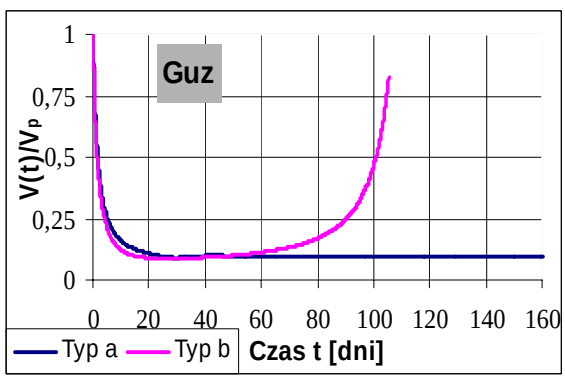
Aplikacja systemu TNF wiąże się z koniecznością zaproponowania jednolitej prezentacji uzyskanych wyników. W każdym przypadku uwzględniono:

- rozkłady czasowe v(t), f(t),
- dwuwymiarowe trajektorie fazowe ilustrujące zmiany poziomu ekspresji mRNA TNF RI + mRNA TNF RII w funkcji poziomu ekspresji mRNA TNF.

Komórki guza (rys. 1a,b)

Typ a. Poziom ekspresji genu TNF obniża się gwałtownie po ok. 20-30 dniach, ulegając następnie stabilizacji. Podobne zachowanie cechuje rozkłady czasowe poziomu ekspresji genów TNF RI oraz TNF RII z uwzględnieniem retardacji (opóźnienia).

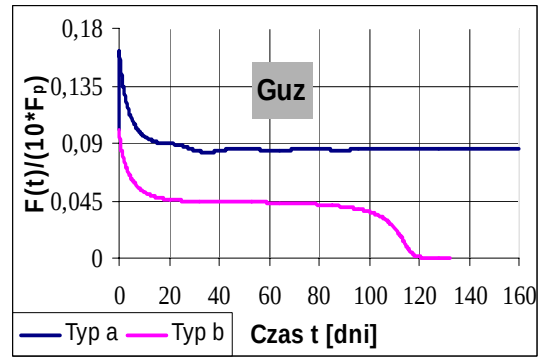
Stabilizacja procesu i zachowanie prawidłowej odpowiedzi immunologicznej – to praktycznie uzyskanie remisji procesu nowotworowego.



Rys. 1a. Trajektoria ilustrująca zmiany w czasie v(t) w komórkach raka jelita grubego

Typ b. Poziom ekspresji genu TNF obniża się gwałtownie po ok. 20-30 dniach, ulegając stabilizacji, a następnie (po ok. 80 dniach) gwałtownie wzrasta, dążąc do nieskończoności z równoczesnym obniżeniem poziomu ekspresji genów jego receptorów (wartości bliskie 0) – taki przebieg krzywych może świadczyć o stopniowym wyczerpywaniu się prawidłowej odpowiedzi immunologicznej.

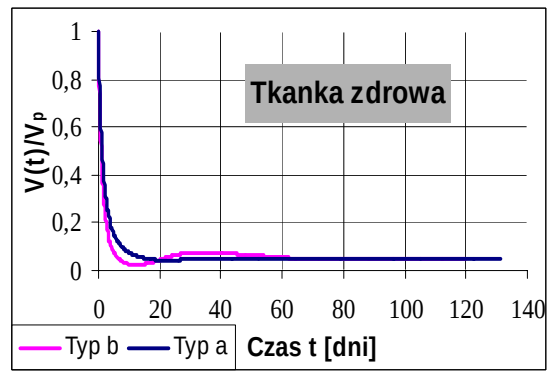
W praktyce klinicznej taki typ odpowiedzi może świadczyć o szybkiej progresji procesu chorobowego.



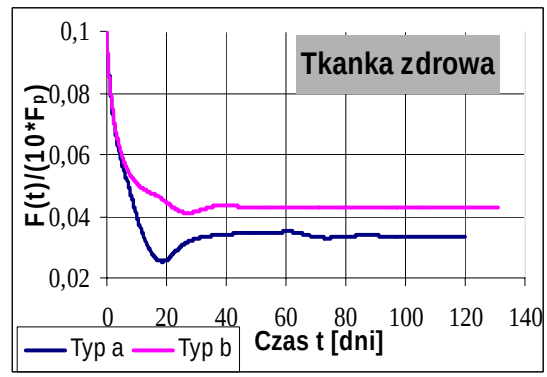
Rys. 1b. Charakterystyki czasowe poziomu ekspresji genów TNF RI oraz TNF RII w komórkach guza

Tkanka zdrowa (rys. 2a,b)

Wykresy przedstawiają gasnące oscylacje poziomu ekspresji genów w czasie o różnej amplitudzie i różnej szybkości tłumienia wahań. Po początkowym pobudzeniu wartości v(t) oraz f(t) stabilizują się (po około 40 dniach), układając w linię zbliżoną do prostej – czyli przyjmują pewien stan równowagi.



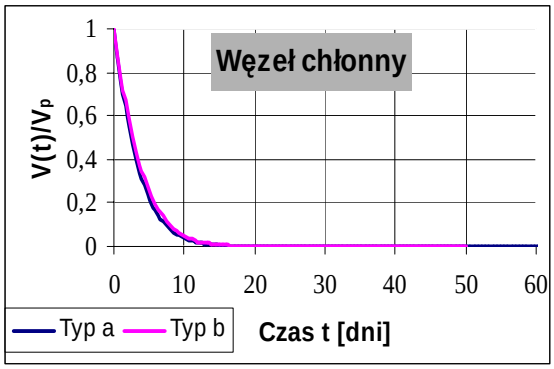
Rys. 2a. Wykres czasowy wartości v(t) w tkance zdrowej



Rys. 2b. Trajektoria czasowa zmian poziomu ekspresji genów TNF RI oraz TNF RII w tkance zdrowej

Zachowania układu są podobne zarówno w *typie a*, jak i *typie b* rozwiązań.

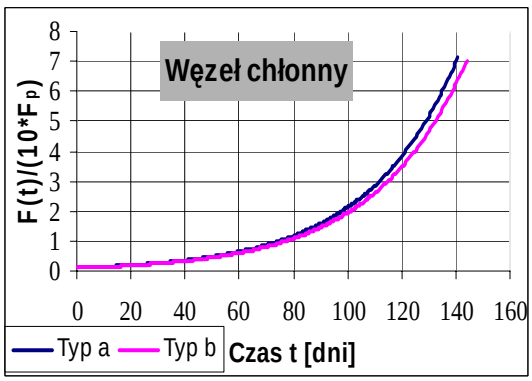
Przerzutowy węzeł chłonny (rys. 3 a,b)



Rys. 3a. Trajektoria czasowa poziomu ekspresji genów TNF w komórkach węzła chłonnego

Poziom ekspresji genu TNF od wartości wysokich, gwałtownie obniża się w czasie, osiągając wartości bliskie zero po ok. 20-30 dniach. W tym czasie wzrasta poziom ekspresji genów receptorów od wartości niskich do nieskończenie wysokich, po ok. 140 dniach – zachowane jest prawdopodobnie ujemne sprzężenie zwrotne pomiędzy poziomem ekspresji genu TNF a poziomami ekspresji genów jego receptorów.

Charakterystyczny jest fakt, że przebieg i zachowanie się krzywych w czasie nie zależy od wartości wyjściowych liczby kopii mRNA TNF, mRNA TNF RI, mRNA TNF RII – bardzo zbliżone w typie a i typie b.



Rys. 3b. Charakterystyka czasowa wartości f w komórkach węzła chłonnego

5. Podsumowanie

Biologia czynnika martwicy nowotworu (*via* analiza poziomu ekspresji genu tej cytokiny oraz genów jej receptorów) to jedna z klinicznych metod pozwalających na zestawienie danych o aktualnej aktywności układu immunologicznego zarówno w stanie zdrowia, jak i choroby. Nigdy nie istnieje, oczywiście, prosta zależność, a uzyskane wartości to wypadkowe wielu niezależnie działających stymulatorów i inhibitorów modyfikujących ekspresję [7]. Piscitelli S. C. [9], wracając uwagę na dynamiczną i nieliniową naturę cytokin, podkreślał sensowność opracowywania modeli matematycznych ilustrujących ich oddziaływanie w tak skomplikowanym układzie, jakim jest organizm ludzki. Pellegrini, Contasta i wsp. [9] stwierdzają nawet, że model matematyczny jest jedyną skuteczną metodą pozwalającą na poznanie układu immunologicznego w całości.

Modelowanie zachowań układu TNF/receptor w raku jelita grubego, poprzez pomiar ekspresji genów jest zupełnie nową, dotychczas nie omawianą metodą.

W literaturze obecne są pojedyncze opracowania dotyczące wykorzystania wyników badań klinicznych jako elementów modelu matematycznego w konkretnych jednostkach chorobowych – np. kolagenozy, *myasthenia gravis*, reumatoidalne zapalenie stawów [6].

W przypadku prezentowanego w pracy modelu, bezpośrednim czynnikiem stymulującym wydzielanie TNF α jest rozwijający się guz nowotworowy. Interesujące wydaje się poznanie wzajemnych relacji pomiędzy guzem a sąsiednimi tkankami.

Pierwszy etap daje możliwość uzyskania odpowiedzi na pytanie, czy i w jakim stopniu dochodzi do pobudzenia ekspresji genu znanej cytokiny o działaniu przeciwzapalnym, przeciwnowotworowym itd. oraz czy ma to wpływ na ekspresję genów jej receptorów.

Drugi etap to przewidywanie zmienności w czasie zachowań komórek nowotworowych oraz komórek sąsiadujących z nowotworem na podstawie uzyskanych wartości liczby kopii mRNA TNF, mRNA TNF RI oraz mRNA TNF RII. Warunki początkowe układu i szczególna wrażliwość na nie pozwalają symulacyjnie wyznaczyć w czasie trajektorie poziomu ekspresji genów i ocenić na podstawie tych wykresów wzajemne zależności pomiędzy komórką nowotworową, przerzutowym węzłem chłonnym i komórką niezmienioną chorobowo.

Opracowując model matematyczny, ze względu na uzyskane wyniki badań molekularnych (liczby kopii mRNA od kilku tysięcy do ponad stu tysięcy) przeprowadzono hipotetyczną linię rozdzielającą uzyskane wartości. Zaproponowano hipotetyczny podział na typ a odpowiedzi (do 30 000 liczby kopii mRNA) oraz typ b (powyżej 30 000). W ten sposób starano się nie popełnić błędu związanego z matematyczną analizą danych w tak szerokim zakresie wartości. Ta hipotetyczna linia zapewnia przede wszystkim prawidłową analizę wyników, ukazując rzeczywiste zachowanie konkretnego układu w czasie. Jednocześnie zaproponowanie hipotetycznego podziału na dwa typy odpowiedzi, w zależności od uzyskanych wartości pozwala na bardziej wiarygodną prezentację modelu i możliwych rozwiązań oraz prawdopodobieństwo uzyskiwania podobnych trajektorii mimo zwiększania liczby badanych pacjentów.

Czyli nawet w przypadku kolejnych bardzo niskich lub bardzo wysokich wartości liczby kopii mRNA w komórkach zdrowych, guzie nowotworowym lub przerzutowych węzłach chłonnych zaproponowane dwa typy odpowiedzi powinny okazać się wystarczające.

Praca składa się z dwóch jakby niezależnych etapów, z których każdy prowadzi do pewnych wniosków i ogólnych sformułowań.

W zaproponowanym modelu każdy z analizowanych układów zachowuje się inaczej, jakby nie miał żadnego związku z dwoma pozostałymi.

Jak zauważył Callard i wsp. [3] – siła sygnału allogenicznego modeluje zmienność stężeń czynnika martwicy nowotworu (*wrażliwość na warunki początkowe*

układu), ale tylko do pewnego momentu w miarę nasilania się poziomu ekspresji genu TNF stwierdzono proporcjonalny do niego wzrost stężeń, potem układ zachowywał się w sposób nieprzewidywalny.

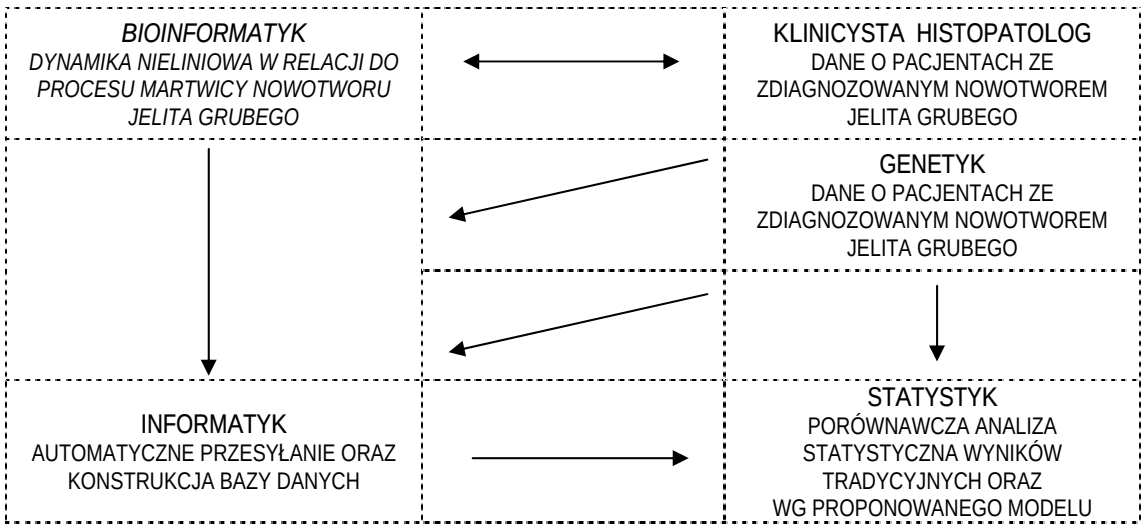
W ostatnich trzech latach w literaturze naukowej ukazuje się coraz więcej doniesień dotyczących wykorzystania możliwości, jakie stwarza dynamika nieliniowa do opisania i przewidywania *zachowań* guza nowotworowego. Daniel Rew [10] sugeruje, że nieliniowość jest charakterystyczna w fazie wzrostowej guza, jego inwazji miejscowej i przerzutach odległych. Czy i w jaki stopniu ma na to wpływ ekspresja genu TNF, dotychczas nie wiadomo.

Matematyczne modele nie zastępują badań laboratoryjnych czy diagnostyki klinicznej, pełnią natomiast istotną rolę w zrozumieniu złożoności procesów zachodzących w komórce i między tymi komórkami, tkankami, narządami.

Każdy leczony pacjent jest indywidualnością, u każdego też przebieg choroby może odbiegać od innych opisanych tego typu przypadków. Ustalając plan postępowania, lekarz zawsze kieruje się własną wiedzą i doświadczeniem, a proponowany model matematyczny ze swoimi rozwiązaniami i sugestiami ma stanowić jedynie uzupełnienie jego możliwości praktycznych z ewentualnym przewidywaniem korzystnych lub niekorzystnych wydarzeń.

Literatura cytowana

1. Arai K., Lee F., Mijajima A., Mijatake S., Arai N.: Cytokine: coordinators of immune and inflamatory responses, Annu. Rev. Biochem., 59:783- 836, 1990.
2. Amiot F., Boussadia O., Cases S., Fitting C., Lebastard M., Cavaillon J. M., Milon G., Dautry F.: Mice heterozygous for a deletion in the tumour necrosis factor alpha and lymphotoxin alpha genes: biological importance of a non-linear response of tumour necrosis factor alpha gene dosage, Eur. J. Immunol. 27, 1035-1042, 1997.
3. Callard R., George A. J. T., Stark J.: Cytokines, Chaos, and Complexity, Immunity, 11, 507-13, 1999.
4. Campbell D. A., Field M., Mc Ardle C. S., Cooke T. G., Gallagher G.: Polymorphism at the tumour necrosis factor locus: a marker of genetic predisposition to colorectal cancer? *Lancet*. 343:293,1994.
5. Diez-Ruiz A, Tilz G. O., Zangerele R., Baier-Bitterlich G., Wachter H., Fuchs D.: Soluble receptors for TNF in clinical laboratory diagnosis, Eur. J. Haematol, 54,1-8,1997
6. Fu – ZQ, Harrison R. W., Reed C., Wu J.: Model complexes of TNF alpha with receptors RI and RII, Protein Eng. Dec. 8(12):1233-1241, 1995.
7. Muc-Wierzgoń M.: Modelowanie aktywności biologicznej czynnika martwicy nowotworu (TNF) i jego rozpuszczalnych receptorów w wybranych schorzeniach ostrych i przewlekłych. Rozprawa habilitacyjna, ŚAM, Zabrze 1999.
8. Old L. J.: Tumor necrosis factor. Science, 230: 630-632, 1985.
9. Piscitelli S. C., Reiss W. G., Figg W. D., Petros W. P.: Pharmacokinetic Studies with recombinant cytokines. Clin. Pharmacokinet. May, 32(5):368,1997.
10. Rew D.A.: Tumour biology, chaos and non-linear dynamics. Eur J Surg Oncol., 25(1):86-89, Feb. 1999.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 233-236.

MODEL SYMULACYJNY WYMIANY GAZOWEJ TLENU

THE SIMULATION MODEL OF OXYGEN GAS EXCHANGE

WOJCIECH LASOŃ*, PIOTR BOGDAŁ, PIOTR WALECKI, WIESŁAW PYRCZAK

**Zakład Bioinformatyki i Telemedycyny, Collegium Medicum, Uniwersytet Jagielloński, Kraków, *wojtekls@poczta.fm*

Streszczenie. W referacie przedstawiono model matematyczny regulacji wymiany gazowej tlenu w procesie oddychania człowieka. Doprowadzenie tlenu do komórek odbywa się zgodnie z gradientem ciśnienia cząstkowego, a jego regulacja następuje poprzez zmiany wentylacji. Wymiana gazowa opisana została za pomocą układu równań różniczkowych (równań dyfuzji tlenu). Ze względu na złożoność badanego systemu model pomija wpływ innych procesów fizjologicznych w organizmie na oddychanie. Przyjęto również szereg założeń, co pozwoliło uprościć równania modelu. Wprowadzone ograniczenia nie spowodowały znacznych odchyleń wielkości wyjściowych od stanów fizjologicznych. Przeprowadzone eksperymenty symulacyjne potwierdzają poprawność działania w określonych przedziałach zmienności parametrów. Poprzez analogię można zbudować modele wymiany gazowej dwutlenku węgla czy też np. stężenia jonów wodorowych pH.

Słowa kluczowe: modelowanie, symulacja, system oddychania

Abstract. This paper presents a mathematical model of an oxygen gas exchange regulation in the process of human respiration. Leading the oxygen to cells is held in accordance with a partial pressure gradient and its regulation follows across the changes of ventilation. The gas exchange was described thanks to arrangement of differential comparisons (oxygen diffusion comparisons). Because of the complexity of the investigated system the model does not account for influence of other physiological processes in organism respiration. Moreover, a range of assumptions were taken to permit the model arrangements simplifying. Such applied limitations did not cause any considerable deviations on the initial quantities from physiological states. The correctness of operations in definite sections of parameters variability was confirmed by undertaken simulation experiments. There may be other gas exchange models analogically created as e.g. carbon dioxide or hydrogen pH ion concentration.

Key words: modeling, simulation, system of respiration

1. Fizjologia oddychania

Istotą procesu oddychania jest wyzwolenie energii zgromadzonej w organizmie. Do jej wyzwolenia ze związków chemicznych niezbędny jest tlen atmosferyczny. Proces ten, nazywany oddychaniem, dzieli się na oddychanie zewnętrzne (doprowadzenie cząsteczek tlenu atmosferycznego do wnętrza komórek) i oddychanie wewnętrzne (wchodzenie cząsteczek tlenu w reakcje chemiczne). Oddychanie zewnętrzne jest procesem złożonym, w którym biorą udział: układ oddechowy (drogi oddechowe i płuca), mięśnie poprzecznie prążkowane szkieletowe, krew i układ sercowo-naczyniowy oraz ośrodki nerwowe sterujące oddychaniem.

Doprowadzenie tlenu atmosferycznego do komórek i usuwanie dwutlenku węgla odbywa się zgodnie z gradientem ciśnienia cząstkowego. Oddychania zewnętrzne można podzielić na procesy:

- wentylacji płuc,
- dyfuzji gazów pomiędzy powietrzem pęcherzykowym a krwią,
- transportu gazów za pośrednictwem krwi,
- dyfuzji gazów pomiędzy krwią a tkanka.

W czasie wentylacji płuc powietrze atmosferyczne zawierające azot i tlen oraz niewielka ilość dwutlenku węgla wciągane są do pęcherzyków płucnych. W powietrzu wydychanym ilość tlenu jest znacznie mniejsza, natomiast stężenie dwutlenku węgla znacznie większe. W pęcherzykach płucnych zachodzi wymiana gazów pomiędzy powietrzem pęcherzykowym i krwią płynącą przez sieć naczyń włosowatych krwionośnych otaczających pęcherzyki. Cząsteczki tlenu dyfundują ze światła pęcherzyków do krwi, ponieważ w powietrzu pęcherzykowym prężność tlenu jest większa (pO_2 = 13 kPa) niż w dopływającej krwi (pO_2 = 5 kPa). W przeciwnym kierunku dyfundują cząsteczki dwutlenku węgla (dla krwi pCO_2 = 6 kPa, dla powietrza pCO_2 = 5 kPa). Cząsteczki tlenu, dyfundując do krwi, muszą przenikać przez ścianę pęcherzyka płucnego i ścianę naczynia włosowatego, a po jej pokonaniu, rozpuszczają się w osoczu wypełniającym naczynia włosowate na zasadzie rozpuszczalności fizycznej. Rozpuszczone w osoczu cząsteczki tlenu natychmiast dyfundują do krwinek czerwonych, gdzie wiążą się podobnie jak cząsteczki dwutlenku węgla na drodze chemicznej.

Krew tętnicza dopływająca do wszystkich tkanek ma wyższą prężność tlenu i niższą prężność dwutlenku węgla

w porównaniu z odpływającą krwią żylną. Zgodnie z gradientem koncentracji uwolniony z hemoglobiny tlen dyfunduje do komórek, dwutlenek węgla zaś dyfunduje z komórek do osocza. W zależności od intensywności metabolizmu wewnątrzkomórkowego występują dość znaczne różnice w prężności tlenu w poszczególnych tkankach.

Badania w obrębie fizjologii pozwoliły na sformułowanie chemicznej zasady regulacji oddychania. Zgodnie z tą zasadą ciśnienia parcjalne dwutlenku węgla, tlenu bądź stężenie jonów wodorowych w organizmie są regulowane poprzez zmiany wentylacji. Mechanizm ten ma charakter chemicznego odruchu i można wyjaśnić go na podstawie morfologicznej i funkcjonalnej odrębności chemoreceptorów (czujniki), ośrodka oddechowego (wzmocnienie) oraz pomocniczego aparatu oddechowego (zmiany wentylacji – element wykonawczy). Pobudzenie powstające samoistnie w ośrodku wdechu jest modulowane, a więc oddechy są przyspieszane i pogłębiane lub zwalniane i spłycane.

Zasadniczym modulatorem aktywności ośrodka wdechu są impulsy biegnące od chemoreceptorów kłębków szyjnych i kłębków aortalnych. Bodźcem drażniącym chemoreceptory jest nieznaczny wzrost ciśnienia parcjalnego dwutlenku węgla i koncentracja jonów wodorowych lub znaczny spadek ciśnienia parcjalnego tlenu we krwi tętniczej [3].

2. Model matematyczny

Prezentowany model opisuje zależności pomiędzy podstawowymi wielkościami gazometrii tlenu. Przyjętymi zmiennymi są ciśnienia parcjalne gazów w pęcherzykach płucnych, krwi tętniczej, żylnej oraz tkankach (płynie tkankowym). Modelując układ oddechowy, przyjęto parametry funkcji przejścia: płuca – tętnice, tętnice – tkanki (współczynniki zlinearyzowanych krzywych absorpcji). Posługując się nazewnictwem technicznym, wprowadzono określenia:

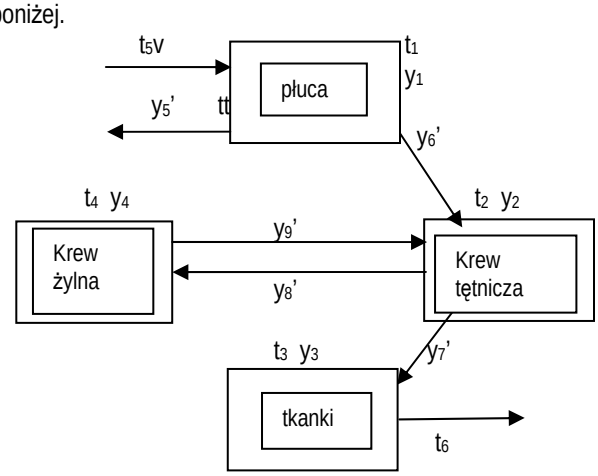
- obiekt - układ oddechowy,
- wielkości wyjściowe – ciśnienia parcjalne tlenu we krwi, tkankach i płucach,
- zakłócenia – zaburzenia w procesach tkankowej przemiany materii (patologia) oraz zmiany zawartości O_2 we wdychanym powietrzu.

W celu matematycznego opisu złożonego procesu oddychania konieczne jest przyjęcie szeregu założeń oraz ograniczeń. Upraszcza to opis matematyczny oraz znacznie zmniejsza nakłady obliczeniowe. Równocześnie przyjęte założenia nie powodują powstania znacznych odchyleń wielkości wyjściowych od stanów fizjologicznych.

Koncepcje budowy modelu oparto częściowo o model Grodinsa, dotyczący układu regulacji stężenia dwutlenku węgla w organizmie. Podczas tworzenia modelu gazometrii tlenu przyjęto, że:

- Oddychanie odbywa się rytmicznie, objętość powietrza w pęcherzykach płucnych zmienia się w zależności od częstości i głębokości oddechów. Ponieważ interesującą nas wielkością jest czasowa wentylacja płuc (iloczyn częstości i objętości oddechowej), przyjmujemy, że zamiast okresowej wentylacji zachodzi ciągły przypływ powietrza przez płuca. Pozwala to traktować objętość powietrza pęcherzykowego jako stałą. Podejście takie wymaga pominięcia różnicy pomiędzy ilością powietrza wdychanego i wydychanego.
- W obrębie rozważanego obiektu fizjologicznego istnieją opóźnienia wynikające z czasów dyfuzji gazów w tkankach, płucach między krwią i tkankami, między krwią a płucami, a także z czasu wymiany gazów między atmosferą i płucami oraz płucami i tkankami. Opóźnienia te, rzędu sekund, mogą być pominięte w równaniach z czasami ustalania się wielkości wyjściowych, które są rzędu minut.
- Tkanki traktujemy jako obszar homogenny, tzn. przyjmujemy, że we wszystkich tkankach panuje jednakowe ciśnienie parcjalne O_2 równe ciśnieniu parcjalnemu uśrednionemu po całej objętości płynu tkankowego. Z punktu widzenia przemiany materii tkanki należy traktować jako obszar heterogenny.
- Przepływ krwi ma charakter ciągły, a jego stała w czasie i niezależna od ciśnienia parcjalnego tlenu wielkość odpowiada maksymalnej pojemności minutowej serca.
- Zarówno dla krwi tętniczej, jak i krwi żylnej odpowiednie krzywe absorpcji mają charakter liniowy.
- Dyfuzja tlenu odbywa się zawsze zgodnie z gradientem panującego ciśnienia parcjalnego tlenu.
- Ciśnienie cząstkowe tlenu w pęcherzykach płucnych w górnych, środkowych i dolnych partiach płuc jest równe ciśnieniu panującemu w partiach środkowych.
- Ciśnienie parcjalne tlenu nie zależy od ciśnień innych gazów oddechowych (szczególnie dwutlenku węgla).

Schemat przepływu tlenu przyjęty w modelu przedstawiono poniżej.



Rys. 1. Schemat funkcjonalny wymiany gazowej tlenu (strzałkami oznaczono kierunek przepływu tlenu)

W modelu przyjęto następujące zmienne i stałe:

Zmienne:

y_1 – ilość O_2 w pęcherzykach płucnych [dm^3],
y_2 – ilość O_2 we krwi tętniczej [dm^3],
y_3 – ilość O_2 w tkankach (płynie tkankowym) [dm^3],
y_4 – ilość O_2 we krwi żylnej [dm^3],
y_5 – ilość O_2 wydychana w powietrzu wydechowym [dm^3],
y_6 – ilość O_2 przyjmowana przez krew tętniczą z pęcherzyków płucnych [dm^3],
y_7 – ilość O_2 przyjmowana przez tkanki [dm^3],
y_8 – ilość O_2 przyjmowana przez krew żylną [dm^3],
y_9 – ilość O_2 przyjmowana przez krew tętniczą z krwi żylnej [dm^3],
v – wentylacja pęcherzykowa [dm^3/s],
p_p – ciśnienie O_2 w pęcherzykach płucnych [kPa],

p_{kt} – ciśnienie O_2 we krwi tętniczej [kPa],
p_t – ciśnienie O_2 w tkankach [kPa],
p_{kz} – ciśnienie O_2 we krwi żylnej [kPa],

Stałe:

t_1 – objętość powietrza pęcherzykowego [dm^3],
t_2 – objętość krwi tętniczej [dm^3],
t_3 – objętość płynu tkankowego [dm^3],
t_4– objętość krwi żylnej [dm^3],
t_5 – stężenie tlenu w gazie wdechowym,
t_6 – tkankowa konsumpcja tlenu [dm^3/s],
t_7 – maksymalna pojemność minutowa serca [dm^3/s],
h_1 – ciśnienie w pęcherzykach płucnych [kPa],
h_2 – ciśnienie krwi tętniczej [kPa],
h_3 – ciśnienie płynu tkankowego [kPa],
h_4 – ciśnienie krwi żylnej [kPa],
h_5 – nachylenie zlinearyzowanej krzywej absorpcji (pęcherzyki płucne – krew tętnicza) [kPa^{-1}],
h_6 – nachylenie zlinearyzowanej krzywej absorpcji (krew tętnicza – tkanki) [kPa^{-1}],
v_0 – wentylacja pęcherzykowa odpowiadająca warunkom normalnym [dm^3/s],
α – stała regulatora (względna zmiana wentylacji na jednostkową zmianę ciśnienia) [kPa^{-1}],

Równania opisujące przepływy tlenu:

dla płuc: $$y_1' = t_5 v - y_5' - y_6', \tag{1}$$

dla krwi tętniczej: $$y_2' = y_6' - y_7' - y_8' + y_9', \tag{2}$$

dla tkanek: $$y_3' = y_7' - t_6, \tag{3}$$

dla krwi żylnej: $$y_4' = y_8' - y_9', \tag{4}$$

przepływ tlenu dla powietrza wydychanego związany jest zależnością:

$$y_5' = v\, y_1 / t_1 \tag{5}$$

Po uwzględnieniu założeń upraszczających, otrzymujemy odpowiednio dla wiązań: pęcherzyki płucne – krew tętnicza oraz wiązania krew tętnicza – tkanki równania:

$$y_6' = t_7 h_1 h_5 y_1 / t_1 \tag{6}$$

$$y_7' = t_7 h_2 h_6 y_2 / t_2 \tag{7}$$

Z warunków zachowania przepływów dla krwi tętniczej i żylnej otrzymujemy odpowiednio:

$$t_7 y_2 / t_2 = y_7' + y_8' \tag{8}$$

$$t_7 y_4 / t_4 = y_9' \tag{9}$$

Po uporządkowaniu równań (5), (6), (7), (8), (9) i podstawieniu do (1), (2), (3), (4) otrzymujemy układ czterech liniowych równań różniczkowych:

$$y_1' = -\frac{\left(v + t_7 h_1 h_5\right) y_1}{t_1} + t_5 v \tag{10}$$

$$y_2' = \frac{t_7 h_1 h_5 y_1}{t_1} - \frac{t_7 y_2}{t_2} + \frac{t_7 y_4}{t_4} \tag{11}$$

$$y_3' = \frac{t_7 h_2 h_6 y_2}{t_2} - t_6 \tag{12}$$

$$y_4' = \frac{t_7\left(1 - h_2 h_6\right) y_2}{t_2} - \frac{t_7 y_4}{t_4} \tag{13}$$

Jako wielkości wyjściowe modelu przyjęto ciśnienia parcjalne tlenu. Uproszczona funkcje przejścia stężenie–ciśnienie ma postać:

dla płuc: $$p_p = h_1 y_1 / t_1, \tag{14}$$

dla krwi tętniczej: $$p_{kt} = h_2 y_2 / t_{21}, \tag{15}$$

dla tkanek: $$p_t = h_3 y_3 / t_3, \tag{16}$$

dla krwi żylnej: $$p_{pz} = h_4 y_4 / t_4, \tag{17}$$

Układ równań w układzie otwartym przyjmuje ostatecznie postać:

$$p_p' = -\frac{\left(v + t_7 h_1 h_5\right) p_p}{t_1} + \frac{t_5 v h_1}{t_1} \tag{18}$$

$$p_{kt}' = \frac{t_7 h_2 h_5 p_p}{t_2} + \frac{t_7 h_2 p_p}{t_2 h_2} + \frac{t_7 h_2 p_{kz}}{t_2 h_4} \tag{19}$$

$$p_t' = \frac{t_7 h_3 h_6 p_{kt}}{t_3} - \frac{t_6 h_3}{t_3} \tag{20}$$

$$p_{kz}' = \frac{t_7\left(1 - h_2 h_6\right) h_4 p_{kt}}{h_2 t_4} - \frac{t_7 h_4 p_{kz}}{h_4 t_4} \tag{21}$$

Symulacje wykonywane na modelu w układzie otwartym mogą być wykorzystywane do badania właściwości systemu oddechowego w warunkach sztucznej wentylacji, gdzie wentylacja wymuszana jest zewnętrznie, a mechanizm fizjologicznej regulacji jest blokowany. Można również prognozować zachowanie się systemu w warunkach patologicznych, przy różnym stopniu degeneracji układu regulacji. Układ otwarty, pomijający aparat regulacji, nie obrazuje jednak właściwości systemu oddechowego w warunkach fizjologii.

Zmiany ciśnień parcjalnych gazów w organizmie znajdują swoje odzwierciedlenie w zmianach wentylacji. Konieczne jest zatem wprowadzenie regulatora działającego bezpośrednio na wentylację pęcherzykową, czyli wydajność źródła zasilającego organizm w tlen. W niniejszym modelu przyjęto, że zmiany wentylacji reagują w sposób bezzwłoczny na zmiany ciśnienia parcjalnego w tkankach. Równanie regulatora przyjmuje postać:

$$v = v_0\left[1 + \alpha\left(p_t - p_{t_0}\right)\right], \tag{22}$$

gdzie: α – stała regulatora,
p_{t_0} – zadany poziom stabilizacji ciśnienia tlenu, odpowiadający ciśnieniu parcjalnemu tlenu w tkankach w warunkach fizjologicznych przy wentylacji pęcherzykowej v_0. Współczynnik α przyjmuje wartość ujemną, ponieważ zwiększenie ciśnienia tlenu w tkankach ponad zadany poziom fizjologiczny

powoduje zmniejszenie wentylacji. Przyjęta w modelu liniowość opisu regulatora jest założeniem upraszczającym.

Wstawiając równanie regulatora (22) do układu równań (18)... (21), otrzymujemy układ równań różniczkowych opisujących wymianę tlenu i położenie równowagi obiektu uwzględniający mechanizm regulacji. Regulator oddziałuje bezpośrednio tylko na zmienną p_p, wobec czego równanie (18) przybiera postać:

$$p_p' = -\frac{\left\{v_0\left[1+\alpha\left(p_t - p_{t_0}\right)\right] + t_7 h_1 h_5\right\} p_p}{t_1} + \frac{t_5 h_1 v_0\left[1+\alpha\left(p_t - p_{t_0}\right)\right]}{t_1} \tag{23}$$

Po przekształceniach otrzymujemy układ równań:

$$p_p' = \frac{\left(v_0 \alpha p_{t_0} - v_0 - t_7 h_1 h_5\right) p_p}{t_1} + \frac{t_5 h_1 v_0 \alpha p_t}{t_1} + \frac{v_0 \alpha p_p p_t}{t_1} + \frac{t_5 h_1 v_0\left(1-\alpha p_{t_0}\right)}{t_1} \tag{24}$$

$$p_{kt}' = \frac{t_7 h_2 h_5 p_p}{t_2} - \frac{t_7 p_{kt}}{t_2} + \frac{t_7 h_2 p_{kz}}{t_2 h_4} \tag{25}$$

$$p_t' = \frac{t_7 h_3 h_6 p_{kt}}{t_3} - \frac{t_6 h_3}{t_3} \tag{26}$$

$$p_{kz}' = \frac{t_7 h_4\left(1 - h_2 h_6\right) p_{kt}}{h_2 t_4} - \frac{t_7 p_{kz}}{t_4} \tag{27}$$

$$v = v_0 \alpha p_t + v_0\left(1-\alpha p_{t_0}\right) \tag{28}$$

W celu przeprowadzenia eksperymentów symulacyjnych konieczne jest sprowadzenie równań opisujących gazometrię do postaci układu równań różniczkowych dla narzędzia symulacyjnego. Należy również wyliczyć parametry równań, wstawiając wartości odpowiednich stałych fizjologicznych.

Różniczkując równanie regulatora i przyjmując:

$$x_1 = p_p, x_2 = p_{kt}, x_3 = p_t, x_4 = p_{kz}, x_5 = v$$

$$x_1(0) = p_{p_0}, x_2(0) = p_{kt_0}, x_3(0) = p_{t_0},$$
$$x_4(0) = p_{kz_0}, x_5(0) = v_0$$

otrzymujemy układ pięciu równań różniczkowych:

$$x_1' = \frac{\left(v_o \alpha p_{t_o} - v_o - t_7 h_1 h_5\right) x_1}{t_1} + \frac{t_5 h_1 v_o \alpha x_3}{t_1} - \frac{v_o \alpha x_1 x_3}{t_1} + \frac{t_5 h_1 v_o\left(1-\alpha x_3(0)\right)}{t_1} \tag{29}$$

$$x_2' = \frac{t_7 h_2 h_5 x_1}{t_2} - \frac{t_7 x_2}{t_2} + \frac{t_7 h_2 x_4}{t_2 h_4} \tag{30}$$

$$x_3' = \frac{t_7 h_3 h_6 x_2}{t_3} - \frac{t_6 h_3}{t_3} \tag{31}$$

$$x_4' = \frac{t_7 h_4\left(1 - h_2 h_6\right) x_2}{h_2 t_4} - \frac{t_7 x_4}{t_4} \tag{32}$$

$$x_5' = \frac{v_0 \alpha t_7 h_3 h_6 x_2}{t_3} - \frac{v_0 \alpha t_6 h_3}{t_3} \tag{33}$$

z warunkami początkowymi:

$$x_1(0), x_2(0), x_3(0), x_4(0), x_5(0)$$

3. Podsumowanie

Przeprowadzone eksperymenty dotyczyły różnych warunków fizjologicznych i patologicznych, w tym: zmniejszenia stężenie tlenu we wdychanym powietrzu, zaburzeń konsumpcji tkankowej, zmian pojemności minutowej serca, zaburzeń wentylacji początkowej oraz symulowanej ostrej niewydolności oddechowej. Prezentowany model odzwierciedla poprawnie zależności między podstawowymi wielkościami gazometrycznymi w aspekcie funkcjonalnym i jakościowym. W określonych przedziałach zmienności parametrów wykazuje także poprawność ilościową. Konsekwencją uproszczenia opisu i przyjęcia szeregu założeń są w niektórych przypadkach błędne odpowiedzi modelu dla tych zaburzeń stanów fizjologicznych, które wykraczają poza przyjęte przedziały. Stąd też w celu poprawy działania modelu postuluje się weryfikację ograniczeń i rozbudowanie modelu z uwzględnieniem wpływu procesów fizjologicznych obecnie pominiętych. Zasadne wydaje się także przyjęcie innej postaci funkcji wymuszającej. Jednostajne wymuszenie liniowe, wygodne ze względów obliczeniowych, powinno zostać zastąpione poprzez funkcję cykliczną, zbliżoną kształtem do sinusoidy. Przyjęcie takiego opisu wymuszenia pozwala na śledzenie dynamiki zmian ciśnienia parcjalnego tlenu w procesie oddychania człowieka.

Poprzez analogię można zbudować model wymiany gazowej dwutlenku węgla lub stężenia jonów wodorowych.

Literatura cytowana

1. Fichtenholz G. M.: Rachunek różniczkowy i całkowy, Państwowe Wydawnictwo Naukowe, Warszawa 1985.
2. Tadeusiewicz R., Lazarewicz M. T.: Nowy paradygmat neurobiologii obliczeniowej: *„experiment in computo”*, Materiały V KK MCSB – 2000, s. 13-40, Kraków 2000.
3. Traczyk W. Z.: Fizjologia człowieka w zarysie, Wydawnictwo Lekarskie PZWL, Kraków 1997.
4. Trąbka J.: W uniwersalnym kręgu cybernetyki, Oficyna Wydawnicza ABRYS, Kraków 2001.
5. Zeigler B. P.: Teoria modelowania i symulacji, Państwowe Wydawnictwo Naukowe, Warszawa 1984.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 237-240.

TERAPIA FOTODYNAMICZNA NOWOTWORÓW UPIGMENTOWANYCH. MODELOWANIE WŁASNOŚCI OPTYCZNYCH MELANINY METODAMI MONTE CARLO

PHOTODYNAMIC THERAPY OF PIGMENTED TUMORS. COMPUTER MODELLING OF OPTICAL PROPERTIES OF SYNTHETIC MELANIN WITH THE USE OF MONTE CARLO METHOD

ZENON MATUSZAK[*], ANETA SAWOW[**], MARTA WASILEWSKA-RADWAŃSKA[**]

[*] *Zakład Fizyki Medycznej, Wydział Fizyki i Informatyki Stosowanej Akademii Górniczo-Hutniczej, Kraków, Zakład Biofizyki, Wydział Biotechnologii, Uniwersytet Jagielloński, ul.Gronostajowa 7, 30-387 Kraków, matuszak@novell.ftj.agh.edu.pl*
[**] *Zakład Fizyki Medycznej, Wydział Fizyki i Informatyki Stosowanej Akademii Górniczo-Hutniczej, Kraków, sawow@novell.ftj.agh.edu.pl, radwanska@novell.ftj.agh.edu.pl*

Streszczenie. Z użyciem metody Monte Carlo wykonano symulacje krzywej absorpcji wodnych roztworów DOPA-melaniny i porównano z krzywymi otrzymanymi doświadczalnie. Własności optyczne pojedynczych cząsteczek melaniny traktowanych jako sfery o określonym promieniu zostały obliczone na podstawie teorii Mie,go. Zbadano wpływ rozmiarów cząsteczek melaniny oraz rodzaju generatorów liczb przypadkowych na jakość symulacji krzywych absorpcji metodą Monte Carlo. Stwierdzono dobrą odtwarzalność charakteru zaniku krzywej absorpcji, obie krzywe doświadczalna i wysymulowana miały podobny eksponencjalny charakter zmian intensywności w funkcji długości fali, jednak wartość nachylenia przebiegu dla krzywych wsymulowanych była systematycznie mniejsza.

Słowa kluczowe: Monte Carlo, DOPA-melanina, absorpcja

Abstract. Monte Carlo (MC) simulations of the absorption curves of the aqueous solution of DOPA-melanin were performed and compared with experiment results. The simulations of the optical properties of individual melanin particles treated as spheres were performed according to the Mie theory. The influence of the melanin particles diameters and the type of the random numbers generators on quality of Monte Carlo simulations was investigated and briefly discussed. The MC simulation gave good reproducibility of the exponential character of absorbance dependence on wavelength, but the inclination of the simulated absorption curve was systematically smaller comparing to experimental one.

Key words: Monte Carlo, DOPA-melanin, absorption

1. Wprowadzenie

Studia nad transferem promieniowania optycznego przez tkanki ludzkie, w szczególności przez skórę jest przedmiotem zainteresowania wielu dziedzin biomedycyny od kosmetologii poprzez diagnostykę fotodynamiczną (Photodynamic Diagnostic of Tumours – PDD) i terapię fotodynamiczną nowotworów (Photodynamic Therapy of Tumours – PDT) aż po tomografię optyczną [9], [18].

Z punktu widzenia optyki skóra ludzka jest zbudowana z kilku warstw o różnej grubości i różnych własnościach optycznych; warstwy zawierają zarówno struktury rozpraszające, jak też absorbujące światło [6], [18]. Jest zatem skóra środowiskiem optycznie nieprzezroczystym, czego konsekwencją jest brak prostej metody określenia rozkładu intensywności światła wewnątrz tkanki.

Wśród absorberów obecnych w tkankach zasadniczą rolę odgrywa melanina, polimer o poliindolowej strukturze, którego własności optyczne decydują między innymi o kolorze skóry, czy barwie tęczówki [5], [7], [18]. Pomimo podstawowego znaczenia tego polimeru w optyce skóry jego własności optyczne są słabo poznane i istnieje niewiele prac, które własności te analizują [5], [7], [12], [13], [21]. Widmem absorpcji polimeru jest monotonicznie malejąca funkcja długości fali w obszarze od ultrafioletu do bliskiej podczerwieni [7], [14], [15], [21].

Dokładniejsza charakterystyka optycznych własności melaniny ma także fundamentalne znaczenie dla lepszego planowania terapii fotodynamicznej (PDT – Photodynamic Therapy) czerniaków, nowotworów upigmentowanych zawierających melaninę. Skuteczność takiej terapii melanom jest ostatnio przedmiotem badań na hodowlach komórkowych i zwierzętach i wiąże się z nimi spore nadzieje na poprawę skuteczności leczenia tego opornego na terapie typu nowotworu [12], [13], [17].

Terapia PDT polega na aktywacji światłem leków – fotouczulaczy, które, wzbudzone, oddziaływują z tlenem obecnym w tkance. Powstający w tym procesie tlen singletowy powoduje utlenienie ważnych biologicznie związków, przede wszystkim lipidów w błonach biologicznych, co w konsekwencji powoduje ich destrukcję i śmierć komórki [9]. Uzyskanie odpowiedniej ilości wzbudzonych molekuł fotouczulacza związane jest bezpośrednio z intensywnością i składem spektralnym światła docierającego w miejsce jego lokalizacji w tkance. W przypadku nowotworów upigmentowanych intensywność docierającego do fotouczulacza światła zależy przede wszystkim od ilości i rozmiarów cząsteczek melaniny na drodze wiązki światła aktywującej lek.

Ilościowy opis rozkładu intensywności światła w środowiskach optycznie nieprzezroczystych, czyli także w tkankach, może być uzyskany poprzez rozwiązanie równania transportu promieniowania [1], [4]

$$\frac{dI(\vec{r},\vec{s})}{ds} = -(\mu_a + \mu_s)I(\vec{r},\vec{s}) + \frac{\mu_a + \mu_s}{4\pi} \times \int_{4\pi} p(\vec{s},\vec{s}')I(\vec{r},\vec{s}')d\Omega' , \qquad (1)$$

gdzie I(r,s) jest jednostkową intensywnością w jednostkowy kąt bryłowy, r jest lokalizacją cząstki rozpraszającej w kierunku s (s – wektor jednostkowy), μ_a,μ_s są odpowiednio współczynnikami absorpcji i rozproszenia środowiska, natomiast p(s,s) jest funkcją fazową reprezentującą charakter rozproszenia, wkład udziału rozproszenia z kierunku s do s, Ω jest kątem bryłowym. Równanie opisuje zatem zmianę intensywności wiązki światła jako funkcje optycznych parametrów środowiska, w którym zachodzi propagacja wiązki. Pierwsza cześć wyrażenia opisuje stratę intensywności wiązki światła I(r,s) na jednostkę długości w kierunku s z powodu absorpcji i rozproszenia, druga część składowa równania opisuje zmiany w I(r,s) na jednostkę długości w kierunku s spowodowane rozproszeniem z innego rozproszonego kierunku światła I(r,s') dΩ' (tj. światła w graniczny elementarny kąt bryłowy z kierunku s'). Funkcja fazowa dla tkanek jest nieznana, przyjmuje się najczęściej, że ma postać funkcji Heyeyey-Gerstein [1], [10], [19], która zależy od współczynnika anizotropii.

Ogólnie, rozwiązanie równania transportu promieniowania jest skrajnie trudne i rozwiązania analityczne nie są znane. Do większości problemów stosuje się metody przybliżone i procedury symulacyjne oparte najczęściej o metodę Monte Carlo.

Rozwiązanie równania wymaga znajomości współczynników absorpcji i rozproszenia jako funkcji długości fali. Dla melaniny współczynniki te nie są znane, wyznaczono je dla nielicznych długości fali, a prezentowana praca kontynuuje badania zapoczątkowane wcześniej w pracach [14], [15] i jest ich kontynuacją i rozszerzeniem.

2. Materiały i metody

2.1. Materiały

Melaninę użytą w części doświadczalnej pracy zsyntetyzowano metodą autoksydacyjną z DOPA (dwuhydroksyfenyloalanina [8].

2.2. Metody eksperymentalne

Pomiary spektrofotometryczne roztworów wodnych melaniny o różnych stężeniach przeprowadzono na spektrofotometrze Jasco V-550 (Japonia) w standardowych kuwetach o długości drogi optycznej 0,5 cm (rys. 1). Ilość cząsteczek melaniny w roztworze utrzymywano na stosunkowo niskim poziomie w celu uniknięcia efektów ekranowania optycznego, samoprzesłaniania.

Parametry optyczne użyte w symulacjach zostały uzyskane z pomiarów absorbancji roztworów melaniny. Metoda obliczenia parametrów optycznych, niezbędnych w symulacji Monte Carlo widm absorpcji, została opisana w [10], [18], [19].

Ilość cząsteczek DOPA-melaniny w roztworze określano, przyjmując gęstość melaniny ρ = 1.2 g/cm^3, metodą pokrywająca się z użytą w pracach [14], [15].

2.3. Metody symulacyjne

Istnieje wiele metod rozwiązywania równania transportu (1). W pracy użyto metody Monte Carlo, która posiada wiele zalet. Metoda ta zakłada znajomość parametrów optycznych środowiska, współczynników absorpcji oraz rozproszenia. Wielkości te uzyskiwane są z pomiarów, fotony w symulacjach MC traktowane są jako cząstki klasyczne. Symulacji poddaje się trajektorie indywidualnych fotonów, kierunki propagacji są wybierane przez wybór długości kroku oraz kierunku rozchodzenia się fotonu po każdym rozproszeniu. Dla dostatecznie wielu fotonów rozkład dróg dla wszystkich fotonów odtwarza całkowite natężenie pola świetlnego w próbce.

W pracy użyto zmodyfikowanych i adaptowanych kodów oryginalnych programów, w których zastosowano metody Monte Carlo: MCML, CONV oraz ich uproszczonej wersji MC231. Programy zostały opisane w szeregu publikacji [11], [19], [20], a ich oryginalne kody (w ANSI C) są dostępne i mogą być modyfikowane. Część oprogramowania realizującego metodę MC w oparciu o ww. kody zmodyfikowano i zrealizowano w systemie Windows, z użyciem systemu Borland C++ 5.0 [3]. W większości obliczeń użyto oryginalnych kodów, programy skompilowano z użyciem kompilatora Borland C++ 4.52.

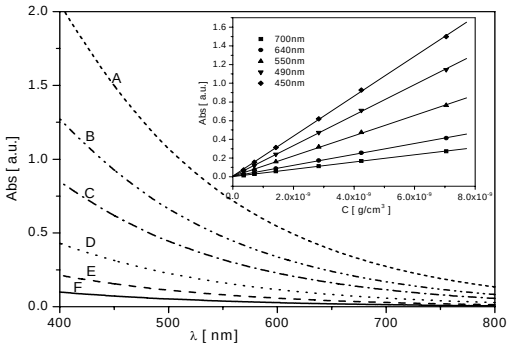
Własności rozproszeniowe cząsteczki melaniny, traktowanej jako kula o określonym promieniu, symulowano przy użyciu teorii Mie'go [1]. Jest wiele dostępnych kodów realizujących oryginalny program fortranowski, zamieszczony w tej książce. Program użyty w pracy jest modyfikacją oryginalnego kodu i jest zrealizowany w języku C [13]. Program umożliwia wyliczenie przekrojów czynnych na rozpraszanie i absorpcję, a zatem także współczynników absorpcji i rozproszenia oraz anizotropii rozproszenia dla pojedynczych cząsteczek o kształcie kulistym (promień cząsteczki zmieniano) i zanurzonych w środowisku o określonym współczynniku załamania. Współczynnik załamania dla cząsteczek melaniny przyjęto jako n=1.5 [14].

3. Wyniki

3.1. Własności optyczne roztworów DOPA-melaniny

Typowa zależność absorpcji DOPA-melaniny od długości fali jest przedstawiona na rys. 1. Obserwujemy monotoniczny

zanik poziomu absorpcji ze wzrostem długości fali. Zależność pomiędzy intensywnością absorpcji przy ustalonej długości fali od stężenia polimeru jest w badanym zakresie stężeń liniowa, roztwory spełniają prawo Lamberta-Beera (rys. 1, wklejka). Intensywność absorbancji melaniny maleje praktycznie eksponencjalnie ze wzrostem długości fali.



Rys. 1. Widma absorpcji melaniny dla różnych stężeń: A – 7,04*10^{-9} g/cm^3, B – 5,83*10^{-9} g/cm^3, C – 2,81*10^{-9} g/cm^3, D – 1,42*10^{-9} g/cm^3, E – 7,11*10^{-10} g/cm^3, F – 3,55*10^{-10} g/cm^3. Wklejka. Zależność absorbancji melaniny od stężenia dla wybranych długości fal. Roztwory polimeru spełniają prawo Lamberta-Beera

3.2. Wyliczenie parametrów optycznych używanych w symulacji

Dla określonego stężenia melaniny wyliczano własności absorpcyjne i rozproszeniowe warstwy zawierającej określoną ilość cząsteczek melaniny, traktując cząsteczki melaniny jako obiekty kuliste o określonym promieniu.

Przykładowe wyliczone parametry warstwy melaniny dla określonych wybranych długości fali przedstawia tabela 1.

Tabela 1. Parametry symulacyjne roztworów DOPA-melaniny używane w metodzie MC

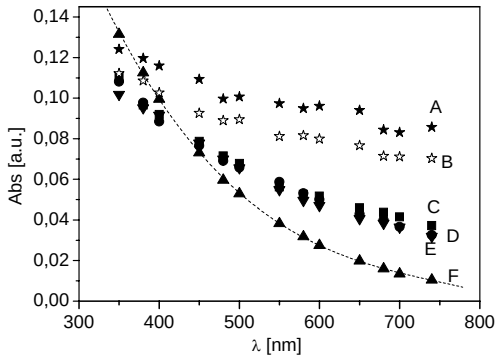
λ[nm]	μ_a [cm^{-1}]	μ_s [cm^{-1}]	g
400	0,958	1,26	0,802
450	0,697	0,97	0,765
500	0,498	0,747	0,744
550	0,355	0,582	0,719
600	0,2528	0,506	0,695
650	0,1719	0,464	0,675

3.3. Wpływ wielkości cząsteczek na przebieg symulacji krzywej absorpcji melaniny

Rozmiary cząsteczek melaniny w roztworach nie są znane dokładnie, rozmiary te zależą od warunków środowiskowych (rozpuszczalnika, pH roztworu) i szacuje się, że zmieniają się od około 45 nm do około 1 μm [10], [16]. Dla danego stężenia polimeru w systemie dominują cząsteczki o określonym promieniu.

W symulacjach dla określonego stężenia melaniny badano zgodność eksperymentalnej krzywej absorpcji z krzywymi wysymulowanymi, zmieniając rozmiary cząsteczki i badając wpływ tego parametru na przebieg symulacji.

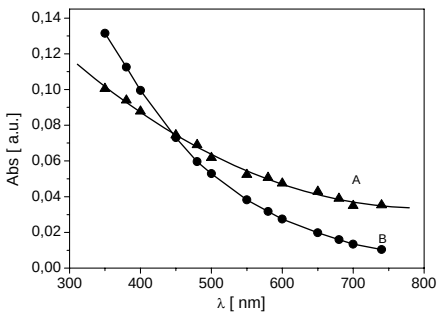
Jak się okazało, na zgodność symulacji z wynikami eksperymentalnymi ma także wpływ rodzaj generatora liczb losowych stosowanego w symulacjach Monte Carlo. Lepszą zgodność uzyskano, zmieniając rodzaj generatora liczb losowych z oryginalnego [11], [19], [20] na generator borlandowski [3]. Podsumowanie wyników tego etapu pracy zostało pokazane na rys. 2.



Rys. 2. Wpływ rozmiarów cząsteczek oraz rodzaju generatora liczb przypadkowych na symulacje metodą MC widm absorpcji DOPA-melaniny w wodzie. Stężenie melaniny c= 5,83*10^{-9} g/cm^3. Cząsteczki melaniny są traktowane jako sfery o zadanym promieniu r. A. r=0,15 μm, B. r=0,2, C. r=0,2 μm, D. r=0,3 μm, E. r=0,4 μm, F. Dane doświadczalne.

Symulacje dla punktów na krzywych A i B wykonano, używając metody MC z oryginalnym generatorem MC programów [11], [19], [20]. Dla punktów na krzywych C, D, E zmieniono rodzaj generatora na standardowy generator liczb losowych wbudowany w C++ funkcja random() w połączeniu z randomize() [3]

Rys. 2 pokazuje, że zgodność krzywych absorpcji DOPA-melaniny doświadczalnej i symulowanej posiada charakter półilościowy. Zarówno przebieg (charakter spadku), jak też wartości absorbancji są odtworzone z niezłą dokładnością, niemniej jednak wartości doświadczalne i symulacyjne różnią się zarówno szybkością zaniku, jak też wartościami intensywności. Fakt ten zilustrowano na rys. 3.



Rys. 3. A – Widmo absorpcji DOPA-melaniny (c=1,17*10^{-8} g/cm^3). Symulacja widma absorpcji melaniny metodą MC, generator borlandowski. Promień cząsteczki melaniny r=0,4 μm; B – doświadczalny pomiar absorpcji melanimy. Obie krzywe są dobrze odtwarzane przez funkcje eksponencjalne (współczynnik korelacji powyżej 0.99), natomiast szybkości zmian funkcji α (zmiana absorbancji na jednostkę długości fali) różnią się wyraźnie. Krzywej A, α_A=-180, krzywej B, α_B=-256

4. Podsumowanie

Przeprowadzone symulacje MC widm absorpcji melaniny pozwoliły – po zastosowaniu wbudowanego generatora liczb przypadkowych – na półilościowe odtworzenie charakteru zależności absorpcji melaniny od długości fali. Otrzymywane widma symulacyjne nie oddają jednak ilościowo obserwowanych w doświadczeniu wartości. Wysymulowane wartości absorbancji są systematycznie zaniżane dla krótszych

długości fal i zawyżane dla fal o większych długościach (rys. 2). Obserwacja ta dotyczy symulacji dla obydwu rodzajów użytych generatorów liczb losowych, z tym że symulacje z użyciem generatora borlandowskiego lepiej odtwarzają szybkość zaniku absorbancji w zakresie długofalowym widma. Poprawa wyników symulacji po zastosowaniu generatora borlandowskiego nie została jak na razie wyjaśniona.

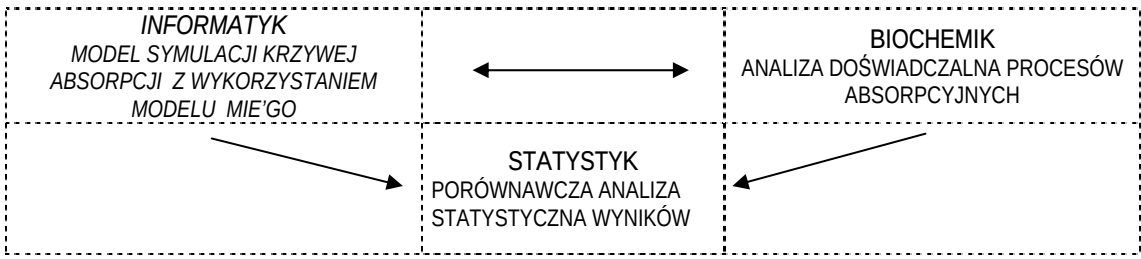
Fizyczny model cząsteczki melaniny (sfera) używany w symulacjach krzywych absorpcji metodą MC jest modelem uproszczonym. Istnieją dane eksperymentalne, że melanina może tworzyć – szczególnie przy większych stężeniach polimeru – bardziej złożone agregaty o kształcie elipsoid, a nawet cząstek o dość nieregularnym kształcie, których zachowanie w procesach rozpraszania jest o wiele bardziej skomplikowane [10], [16], [21]. Fakt ten wpływałby na niedowartościowanie współczynnika rozproszenia w obszarze krótszych długości fal, czego konsekwencją byłoby zaniżenie wartości symulowanej absorbancji.

Najprawdopodobniej także wzrost stężenia polimeru może się objawiać zwiększeniem ilości cząsteczek o mniejszych rozmiarach (rozpad agregatów melaninowych), co wpływa na jednorodność populacji cząsteczek melaniny, układ staje się systemem o różnym rozkładzie wielkości obiektów, czego konsekwencją jest nakładanie się krzywych absorbancji od cząsteczek o różnych wymiarach i rozpraszanie w całym zakresie długości fal jest złożeniem rozpraszania Rayleigha i Mie'go [1], [2], [10], czego zastosowany model na tym etapie rozwoju nie uwzględnia. Uzyskane symulacje wskazują, że w tym kierunku powinno podążać jego doskonalenie; wskazanie tego kierunku było jednym z celów prezentowanej pracy.

Wyniki badań wskazują, że problem wyznaczenia dokładnego rozkładu światła wewnątrz warstwy melaniny, pozostaje nadal nierozwiązany i wymaga dalszych badań zarówno doświadczalnych jak tez symulacyjnych.

Literatura cytowana

1. Bohren C. F., Huffman D. R.: Absorption and Scattering of Light by Small Particles, Wiley- Interscience Publ., Wyd. Ros. Mir., Moscow 1986.
2. Borland C++5.0: Opis programu. Borland Software Corporation 2003.
3. Bridelli M. G.: Self-assembly of melnin studied by laser ligt scattering, Biophys. Chem., 73: 227-239, 1998.
4. Chandrasekhar S.: Radiative Transfer: Dover, New York 1960.
5. Chedekel M. L.: Photophysics & Photochemistry of Melanin, in Melanin: Its Role in Human Photoprotection, ed. Zeise L., Chedekel M. R., Fitzpatrick T. H., Valdenmar Publ. Comp. Overland Park, Kansas,11-22, 1995.
6. Cheong W. F., Prahl S. A., and. Welch A. J.: Review of the Optical Properties of Biological Tissues, IEEE J. Quantum Electronics, 26: 2166-2185, 1990.
7. Crippa R., Horak V., Prota G., Svoronos P., Wolfram L.: Chemistry of Melanins, in. The Alkaloids, Acad. Press, Inc., 96(6): 254-312, 1989.
8. Felix C. C., Hyde, J. S., Sarna T., & Sealy R.C.: Interactions of Melanin J.S. with Metal Ions. Electron Spin Resonance Evidence for Chelate Complexes of Metal Ions with Free Radicals, J. Am. Chem. Soc. 100: 3922-3926, 1978.
9. Graczyk A. (red.): Fotodynamiczna metoda rozpoznawania i leczenia nowotworów, Bellona, Warszawa 1999.
10. Huang J. S., Sung J., Eisner M., Moss S. C., Gallas J.: The fractal nature and the dynamics of aggregation of synthetic melanin in low pH aqueous solutions, J. Chem. Phys. 90(1): 25-29, 1989.
11. Jacques S. L.: Light Distributions from Point, Line and Plane Sources handicap Photochemical reactions and Fluorescence in Turbid Biological Tissues, Photochem. Photobiol. 67(1): 23-32, 1998.
12. Kukielczak B., Cieszka K. A., Matuszak Z. T., Subczynski W. K.: Experimental Photodynamic Therapy (PDT) of Bomirski Hamster Melanoma using Merocyjanine 540 and Visible Light Part I. Pigmented (ma) line., Curr. Topics in Biophys., 19: 66-70, 1995.
13. Lim D. S., Ko S. H., Lee W. Y.: Silkworm-phophorbe a mediated photodynamic theraphy against B16F10 pigmenetd melanoma, J. Photochem. Photobiol. B., 74(1): 1-6, 2004.
14. Matuszak Z., Wasilewska-Radwańska M.: Measurement and Computer Modeling of Optical Properties of Synthetic Melanins in Solution, Polish J. Med. Phys & Engn., 6(4): 211-223, 2000.
15. Matuszak Z., Wasilewska-Radwańska M.: Melanin as skin and eye photoprotection factor. Theoretical and experimental investigations, Prace Mineralogiczne (Mineralogical Transactions), 89: 47-55, 2000.
16. Nofsinger J. B., Forest S. E., Eibest L. M., Gold K. A. and Simon J. D.: Probing the Buiding Blocks of Eumelenin Using Scanning Electron Microscopy, Lgment cell res., 13: 179-184, 2000.
17. Urbańska K., Romanowska-Dixon B., Matuszak Z., Oszajca J., Nowak-Śliwińska P., Stochel G.: Indocyanine green as a prospective sensitizer for photodynamic therapy of melanomas, Acta Biochim.Polon. 49: 387-391, 2002
18. Young A. R.: Chromophores in human skin, Phys. Med. Biol., 42: 789-802, 1997.
19. Wang L. H., Jacques S. L., Zheng L. Q.: MCML-Monte Carlo modeling of light transportation in multi-layered tissues, Comp Methods. Prog. Biomed., 47: 131-146, 1995.
20. Wang L. H., Jacques S. L., Zheng L. Q.: CONV- Convolution for responses to a finite diameter photon beam incident on multilayered tissues, Computer Methods and Programs in Biomedicine, 54: 141-150, 1997.
21. Wohlbarst M. L., Walsh A. W. and George G.: Melanin, a unique biological absorber, Appl. Opt. 20: 2184-2186,1981.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 241-246.

PRZYSTOSOWANIE MODELU SYSTEMU WZROKOWEGO MUCHY DO ANALIZY RUCHU DROGOWEGO

ADAPTATION OF THE FLY VISUAL SYSTEM MODEL TO TRAFFIC ANALYSIS TASKS

ZBIGNIEW MIKRUT, MICHAŁ ZAJĄCZKOWSKI

Katedra Automatyki Akademii Górniczo-Hutniczej, Kraków, zibi@agh.edu.pl

Streszczenie. System wzrokowy muchy jest jednym z prostszych i lepiej poznanych systemów wizyjnych. W niniejszym artykule wprowadzono rozszerzenia mające na celu dostosowanie podstawowego modelu do zadań analizy ruchu drogowego: detekcji i lokalizacji obiektów oraz estymacji kierunku ruchu i chwilowej prędkości. Zmodyfikowany model przetestowano na sekwencjach rzeczywistych obrazów otrzymanych z krakowskiego wideodetektora.

Słowa kluczowe: system wzrokowy muchy, analiza ruchu drogowego, detektor ruchu

Abstract. The fly visual system is one of the simplest and better known biological systems. In the present paper several modifications have been introduced to adapt the base model to visual traffic analysis: object detection and localization and estimation of movement direction and velocity. The developed model was tested by using real image sequences acquired from a Krakow videodetector.

Key words: fly visual system, traffic analysis, movement detector

1. Wstęp

Zagadnienie detekcji ruchu jest jednym z kluczowych problemów w analizie ruchu drogowego. Prawidłowe i szybkie wykrycie ruchu w sekwencji obrazów warunkuje dalsze etapy analizy, takie jak detekcja pojazdów czy estymacja ich prędkości. Do realizacji tych celów coraz częściej używa się modeli prostszych niż skomplikowane systemy wzrokowe zwierząt wyższych. Takimi względnie prostymi modelami są systemy wzrokowe owadów: szarańczy [1] czy też muchy [4], [3]. Pomimo tej prostoty są to systemy wyjątkowo sprawne. Umożliwiają szybką reakcję organizmu, niezbędną do kontroli trasy lotu, omijania przeszkód i bezpiecznego lądowania.

Celem prac opisanych w artykule było sprawdzenie, czy model układu wzrokowego muchy – po wprowadzeniu odpowiednich rozszerzeń – może zostać użyty do realizacji podstawowych zadań analizy ruchu drogowego, takich jak wykrywanie ruchu, lokalizacja obiektów (pojazdów lub osób pieszych) oraz wyznaczenie kierunku ich przemieszczania oraz prędkości.

2. Model systemu wzrokowego muchy i jego modyfikacja

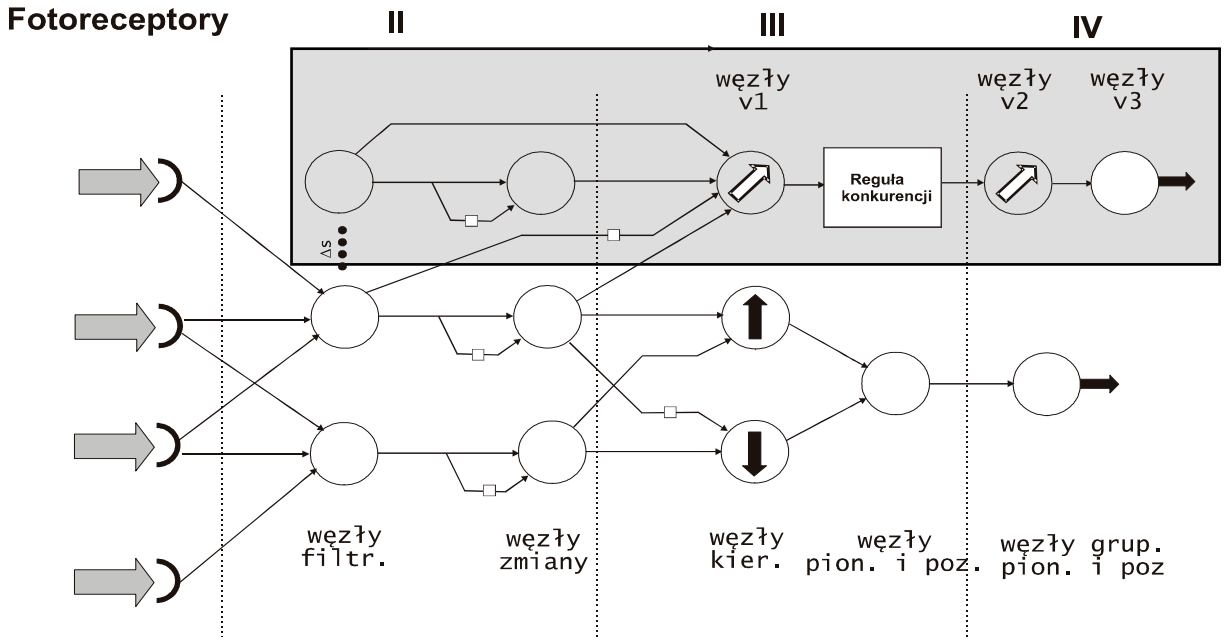
Opisany poniżej model opiera się na pracy [4]. Ogólny schemat zmodyfikowanego modelu został przedstawiony na rys. 1. Został on rozszerzony o elementy umożliwiające lokalizację obiektów i wyznaczanie chwilowej prędkości. Dodatkowe elementy ograniczono prostokątem.

Model składa się z czterech podstawowych warstw. W pierwszej warstwie znajdują się receptory rejestrujące natężenie padającego światła. Źródłem danych dla tej warstwy (w modelu) może być obraz pobierany bezpośrednio z kamery i podawany na wejście *frame grabbera* lub sekwencja kolejnych klatek zapisana w jednym z formatów graficznych.

Sygnały z fotoreceptorów odbierane są w warstwie drugiej przez węzły filtrujące, które wygładzają obraz poprzez uśrednienie aktywności ośmiu sąsiadujących z nimi fotoreceptorów. Kolejnymi komórkami w tej warstwie są węzły opóźniające oraz węzły zmiany. Węzły zmiany porównują aktywność węzłów opóźniających z aktualną aktywnością węzłów filtrujących (rys 3 i 4). Odbywa się to według wzoru:

$$x_z = x_f - x_o, \qquad (1)$$

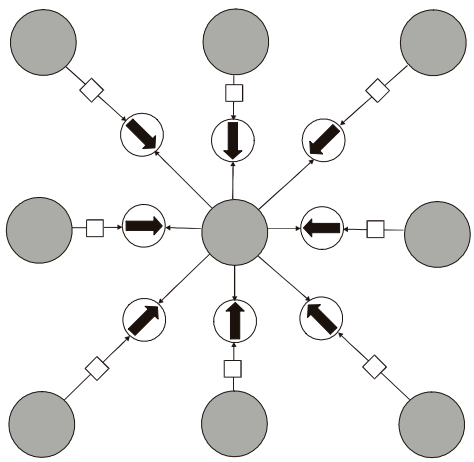
gdzie: x_z – aktywność węzła zmiany, x_f – aktywność węzła filtrującego, x_o – aktywność węzła opóźniającego. Węzły opóźniające oznaczono na schemacie małymi kwadratami. Dla uproszczenia obliczeń zastosowano tutaj zwykłą różnicę aktywności komórek w miejsce równania różniczkowego z pracy [4]. Zaletą tego rozwiązania jest zmniejszenie czasu obliczeń, wadą natomiast brak możliwości wykrycia ruchu z prędkością poniżej jednego piksela na klatkę. Węzły zmiany są ułożone w dwuwymiarowej tablicy, zachowując sąsiedztwo takie samo, jak odpowiadające im fotoreceptory.



Rys. 1. Zmodyfikowany model systemu wzrokowego muchy. Modyfikacje zaznaczono prostokątną ramką. Oprócz elementów ograniczonych ramką zmieniono sposób grupowania w ostatniej warstwie oraz dodano lokalizację obiektów. Opis w tekście

Trzecia warstwa została rozbudowana o neurony służące do detekcji prędkości. W tej warstwie znajdują się węzły opóźniające, węzły kierunkowe oraz węzły kierunkowe poziome i pionowe. Do detekcji prędkości w tej warstwie służą węzły v1. Rysunek 1 ilustruje architekturę trzeciej warstwy. Zostają tutaj wyznaczone kierunki i prędkości ruchu dla małych wycinków pola widzenia.

Węzły opóźniające tej warstwy przechowują aktywność komórek zmiany z poprzedniego kroku czasowego. Aktywność węzłów kierunkowych wskazuje na ruch w preferowanym przez nie kierunku. Istnieje osiem rodzajów tych węzłów – po jednym dla każdego z podstawowych kierunków i zwrotów ruchu (por. rys. 2).



Rys. 2. Schemat połączeń pomiędzy węzłami zmiany (szare koła) a węzłami kierunkowymi. Małe kwadraciki oznaczają elementy opóźniające

Aktywność tych węzłów jest proporcjonalna do aktywności węzła zmiany i odpowiedniego węzła opóźniającego:

$$x_k(t) = \max\left\{ 0, \left[\left(x_{zp}(t-\Delta t) x_{zs}(t) \right) - \Gamma \right] \right\} , \qquad (2)$$

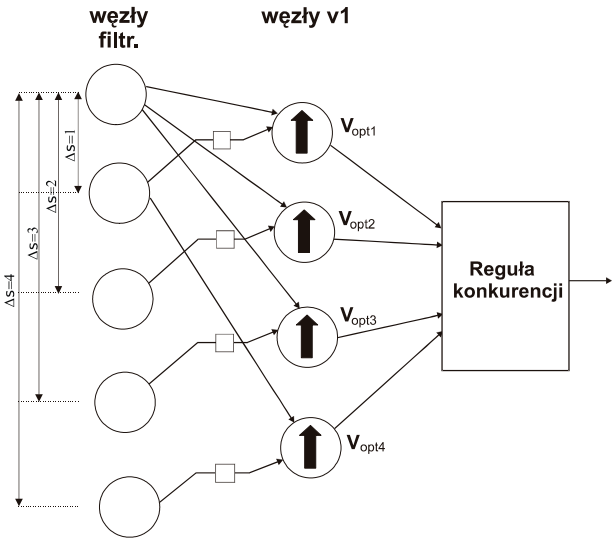
gdzie x_{zp} – aktywność przyległego węzła zmiany, x_{zs} – aktywność środkowego węzła zmiany, Γ – próg eliminujący szum.

Węzły kierunkowe łączą się z komórkami pionowymi i poziomymi. Aktywność węzłów pionowych i poziomych jest wynikiem zrzutowania na odpowiadające im kierunki aktywności węzłów kierunkowych. Wyznaczają one kierunek i zwrot ruchu na małym wycinku pola widzenia.

Węzły v1 połączone są z węzłami zmiany oraz węzłami opóźniającymi v1, które przechowują aktywność węzłów v1 z poprzedniego kroku czasowego (rys. 3). Sposób tych połączeń wynika z zastosowania do detekcji prędkości zmodyfikowanego detektora Reichardta (pół-detektora), którego model był wszechstronnie testowany przez autorów pracy [5]. W prezentowanym modelu systemu wzrokowego muchy zastosowano kilka pół-detektorów dla każdego pola percepcyjnego. Aktywność komórki v1 obliczana jest jako iloczyn aktywności węzła filtrującego oraz węzła opóźniającego v1. Aktywność ta jest większa od zera tylko wtedy, gdy wartości obydwu węzłów zmiany odpowiadających wejściom komórki v1 są niezerowe.

W systemie zamodelowano cztery rodzaje komórek v1 dla czterech podstawowych kierunków ruchu. Oprócz tego podziału każdy z węzłów v1 jest dostrojony do jednej z kilku wartości prędkości. Ilość tych wartości zależna jest od zakresu prędkości, jaki chcemy mierzyć. Dostrojenie węzłów v1 odbywa się poprzez łączenie ich z odpowiednio oddalonymi od siebie węzłem zmiany oraz węzłem opóźniającym. Ta różnica odległości powoduje, że każda z tych komórek wykazuje różną aktywność dla ruchu z tą samą prędkością. Przy użyciu reguły konkurencji dla węzłów v1 w obrębie jednego kierunku wybierany jest węzeł o najwyższej aktywności. Prefe-

rowana prędkość wygrywającej komórki jest wartością najbardziej zbliżoną do rzeczywistej prędkości w polu percepcyjnym.



Rys. 3. Węzły v1. Rozpoznają prędkość w czterech podstawowych kierunkach. Nie zostały zaznaczone połączenia z węzłami zmiany, aby nie zaciemniać rysunku.

Ostatnia (czwarta) warstwa zawiera węzły grupujące, wyznaczające wektory ruchu dla większych obszarów pola widzenia. Węzły te, podobnie jak w poprzedniej warstwie, reprezentują składową poziomą i pionową wektora ruchu. Niezwykle istotny jest sposób połączeń pomiędzy węzłami kierunkowymi pionowymi i poziomymi a węzłami grupującymi. W opisywanym rozwiązaniu zostały przyjęte i zaimplementowane dwa wzajemnie uzupełniające się rodzaje grupowania.

Pierwszy jest związany z wyznaczaniem prędkości. Obszar widzenia jest podzielony siatką przylegających do siebie obszarów oraz drugą, zachodzącą na nią siatką (rys. 4). Każdy z neuronów grupujących połączony jest z neuronami kierunkowymi z obszaru podlegającego mu elementu siatki. Aby uniknąć zakłóceń spowodowanych niewielkimi zmianami obrazu, przyjęto dla tych neuronów pewien próg aktywacji.

Detekcją prędkości w czwartej warstwie zajmują się węzły v2 oraz v3 (por. rys. 1). Komórki v2, podobnie jak v1, podzielone są na cztery grupy – zależnie od odpowiadającego im kierunku ruchu. Obsługują one takie same obszary pola widzenia jak odpowiadające im w tej warstwie węzły grupujące. Aktywność węzła v2 jest wynikiem zsumowania wzbudzeń pochodzących od połączonych z nim węzłów v1. Pobudzająco działają tylko połączenia z tymi neuronami v1, które wygrały rywalizację i reprezentują najbardziej prawdopodobną prędkość ruchu w ich polu percepcyjnym. Aktywność węzłów v2 opisana jest wzorem:

$$x_{v2_k} = \frac{1}{n}\sum_{i=0}^{n} x_{v1_{k,i}} \quad , \tag{3}$$

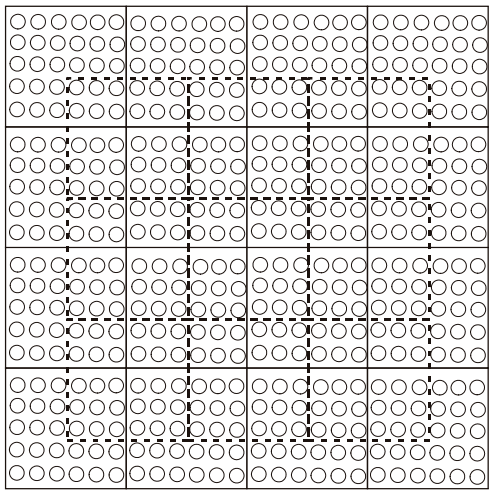
gdzie:

x_{v2_k} – aktywność węzła v2 mierzącego prędkość w k-*tym* kierunku (k⊂[1,4]) w pikselach na klatkę,

$x_{v1_{k,i}}$ – aktywność i-*tego* wygrywającego węzła v1, zawierającego się w obszarze grupowania oraz mierzącego prędkość w k-*tym* kierunku (wartość wyrażona w pikselach na klatkę),

n – ilość wygrywających węzłów v1 w danym obszarze grupowania.

W ten sposób zostają wyznaczone dla obszaru grupowania prędkości w czterech podstawowych kierunkach. Z każdymi czterema węzłami v2 połączony jest węzeł v3. Wyznacza on wypadkową prędkość na danym obszarze. Łącząc informację o kierunku ruchu przechowywaną przez węzły grupujące pionowe i poziome z wartością prędkości wyznaczoną przez węzły v3, otrzymujemy pełny wektor ruchu w danym obszarze pola widzenia.



Rys. 4. Sposób grupowania w czwartej warstwie. Na węzły kierunkowe pionowe i poziome zostają nałożone dwie zachodzące na siebie siatki. Węzły z każdej kratki siatki są podłączone do jednej pary węzłów grupujących

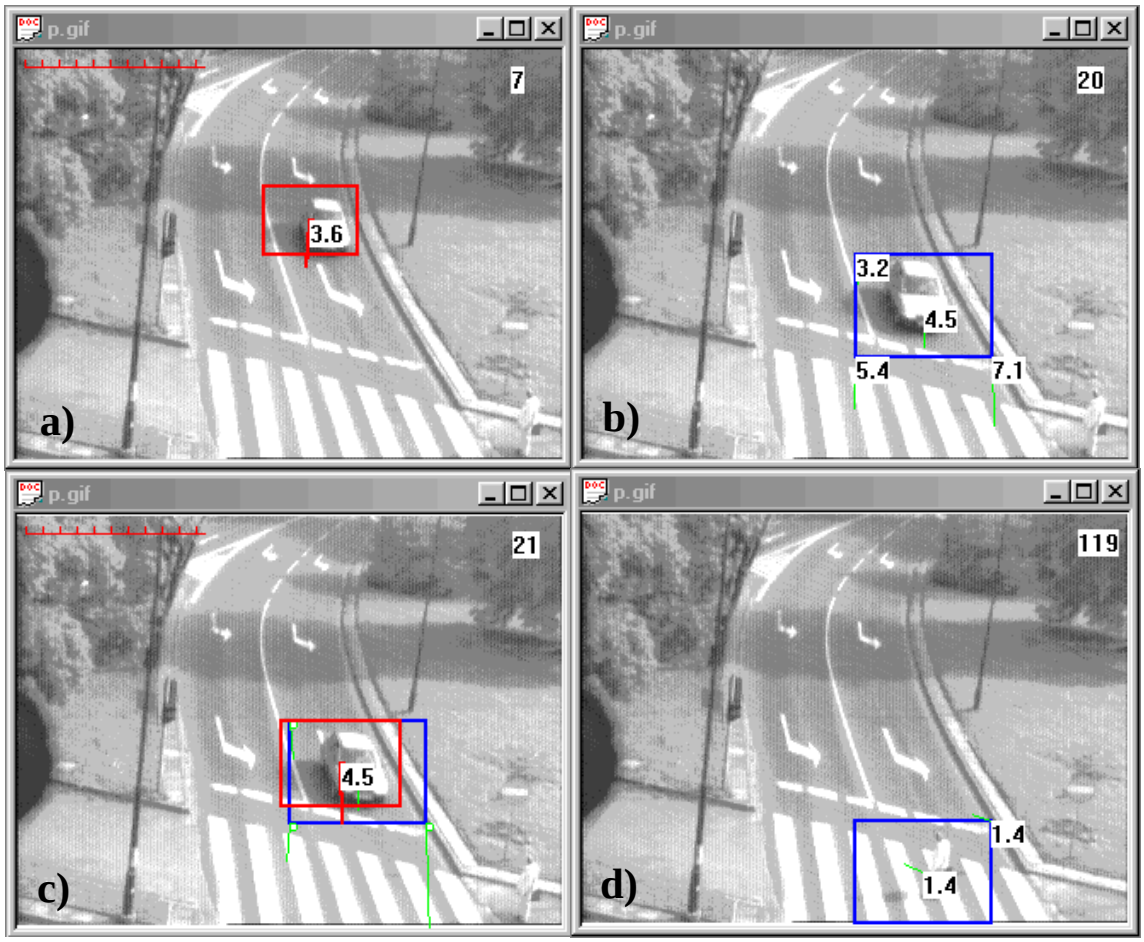
Drugim sposobem grupowania jest lokalizacja poruszających się obiektów. Lokalizacja w opisywanym rozwiązaniu realizowana jest poprzez algorytm sekwencyjny. Został on wybrany ze względu na prostotę, skuteczność i względnie niską złożoność obliczeniową.

Algorytm znajduje najpierw małą grupę sąsiadujących ze sobą węzłów kierunkowych pionowych i poziomych, posiadającą aktywność wyższą od pewnej zadanej wartości progowej. W następnym kroku prostokątny obszar obejmujący grupę aktywnych neuronów jest, rozszerzany w pionie i poziomie. Operacja ta wykonywana jest dopóki ilość węzłów o aktywności powyżej wartości progowej nie stanie się mniejsza niż pewien zadany odsetek wszystkich węzłów w grupie. Następnie znajdowana jest kolejna grupa aktywnych neuronów, nie zawierająca się w żadnym z wcześniej znalezionych obszarów. Powyższe kroki są powtarzane, dopóki nie będzie można znaleźć najmniejszej grupy neuronów spełniającej zadane warunki. W ten sposób otrzymuje się jeden lub kilka prostokątnych obszarów obejmujących najbardziej aktywne skupiska neuronów grupujących pionowych i poziomych. W większości przypadków jest to tożsame ze zlokalizowaniem i określeniem wielkości poruszających się obiektów. Mając dane obszary lokalizujące ruchome obiekty, wyznaczamy ich kierunek oraz prędkość. Odbywa się to w sposób analogiczny jak w pierwszej metodzie grupowania.

3. Wyniki testowania modelu

Model systemu wzrokowego muchy opisany w punkcie 3 został zaimplementowany i przetestowany. Do testów użyto kilkanastu przykładów sekwencji obrazów, zawierających poruszające się obiekty. Były wśród nich obrazy stworzone sztucznie oraz sekwencje rzeczywiste, uzyskane ze stanowiska wideodetekcji. Poniżej zostanie przedstawiony przykład sekwencji rzeczywistej, na której porusza się samochód oraz pieszy. Cztery klatki, wybrane z ponad 100, przedstawiono na rys. 5.

Jak można zauważyć na etykietach wyświetlanych w środku ograniczających prostokątów, prędkość samochodu wzrasta w miarę zbliżania się do przejścia dla pieszych. Kierunek ruchu wskazywany przez odcinek zaczepiony w środku ograniczającego prostokąta jest także zgodny z rzeczywistością. Samochód jest prawidłowo lokalizowany zarówno przy włączonej opcji lokalizacji (rys. 5a i c) jak i przy użyciu siatki wskazującej najaktywniejszy element (rys. 5b). Po przejechaniu samochodu, na pasy wchodzi pieszy i porusza się ze stałą prędkością, w przybliżeniu równą 2 piksele/klatkę. Podobnie jak poprzednio samochód, także pieszy został prawidłowo zlokalizowany. Jego kierunek ruchu oraz prędkość także zgadzają się z odpowiednimi wielkościami rzeczywistymi (por. rys. 5d).

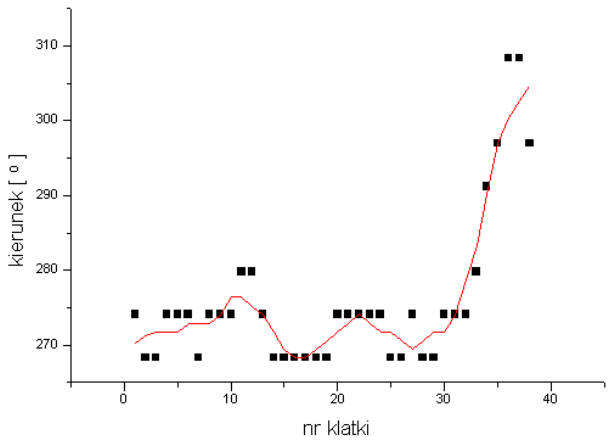


Rys. 5. Wizualizacja wyników: a) włączona opcja lokalizacji, ramka otacza poruszający się obiekt, zaznaczony kierunek ruchu i prędkość, b) wyłączona lokalizacja, ramka otacza najaktywniejszy element siatki, c) połączenie sytuacji z punktów a i b, d) sytuacja jak w b) – wewnątrz kratki o największej aktywności znajduje się poruszający się pieszy

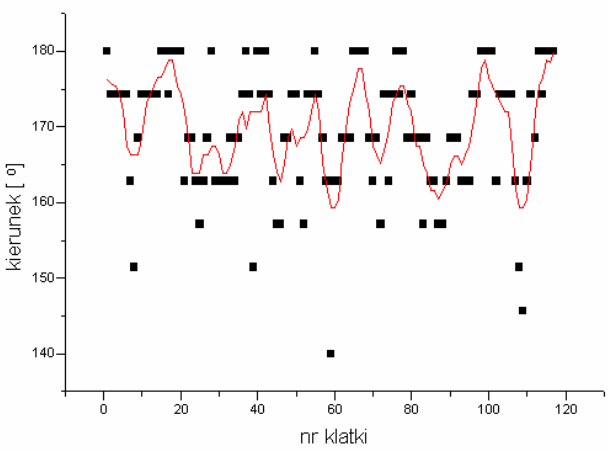
Wyniki ilościowe zostały przedstawione na wykresach (por. rys. 6, 7, 8, 9).

Wykres na rys. 6 pokazuje, jak zmienia się kierunek ruchu samochodu. Jest on przez większą część obserwacji (klatki 1-30) stały i wynosi około 275 stopni. Podczas dojeżdżania do skrzyżowania (klatki 30-40) kąt ten zwiększa się. Wyniki te dokładnie odwzorowują rzeczywisty kierunek ruchu samochodu. Kierunek ruchu pieszego (rys. 7) jest stały – oscyluje wokół wartości 170 stopni. Wynik ten jest zgodny z rzeczywistością obserwowaną na animacji.

Dzięki zarchiwizowaniu kolejnych klatek omawianej sekwencji stało się możliwe uzyskanie danych o rzeczywistych prędkościach – zarówno samochodu, jak i pieszego. Dane te zostały wyznaczone „ręcznie" w wyniku analizy przesunięć obiektów na kolejnych klatkach filmu. W ten sposób stało się możliwe odniesienie prędkości uzyskanych w wyniku symulacji do danych rzeczywistych. Na rysunku 8 zestawiono obie krzywe prędkości chwilowej (obliczoną i rzeczywistą) dla nadjeżdżającego samochodu. Jak widać, oba wykresy prawie się pokrywają.



Rys. 6. Wykres kierunku ruchu samochodu. Zaznaczono kierunki chwilowe oraz krzywą aproksymującą (średnia 5. punktowa)



Rys. 7. Wykres kierunku ruchu pieszego. Zaznaczono kierunki chwilowe oraz krzywą aproksymującą (średnia 5. punktowa)

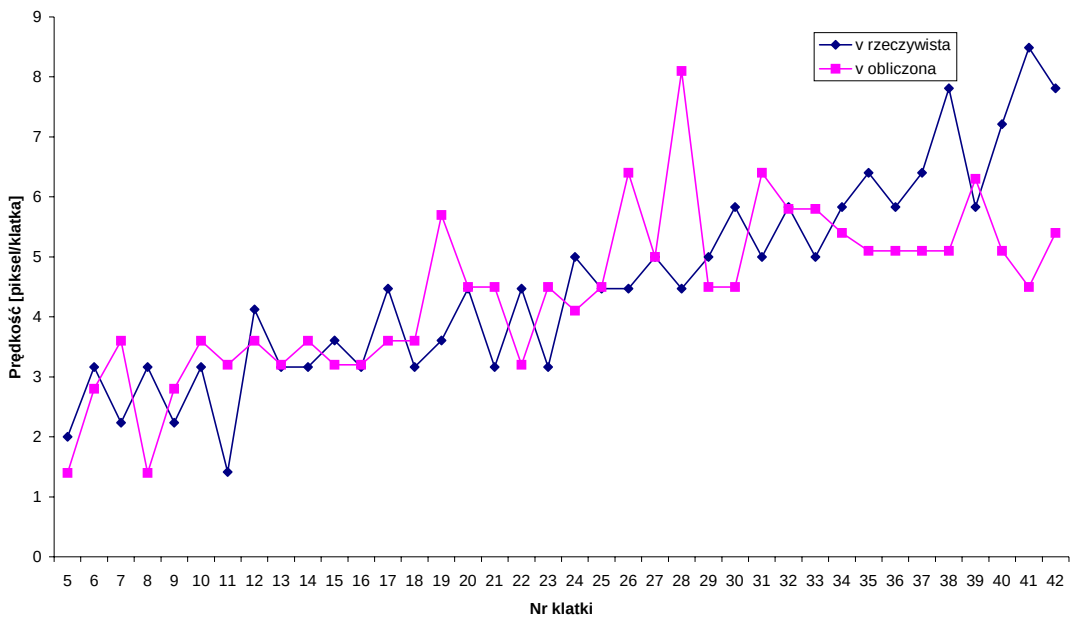
Krzywe na rysunku 9 przedstawiają analogiczne wykresy dla ruchu pieszego. Pewna liczba pomiarów prędkości pieszego odbiega jednak znacznie od wartości rzeczywistych. Przyczyną jest bardziej skomplikowany charakter ruchu pieszego (na każdej klatce pieszy był rejestrowany w trochę innej pozie) oraz zbyt mała liczba *aktywnych* fotoreceptorów. W przypadku, gdy poruszający się obiekt jest tak mały, że podczas jego ruchu uaktywnia się niewielka ilość węzłów v1, wtedy różnice pomiędzy aktywnościami węzłów v1, dostrojonych do różnych prędkości, są bardzo małe. Może to powodować, że podczas wyznaczania przy pomocy reguły konkurencji aktualnej prędkości czasami dochodzą do głosu wartości czysto przypadkowe.

4. Podsumowanie

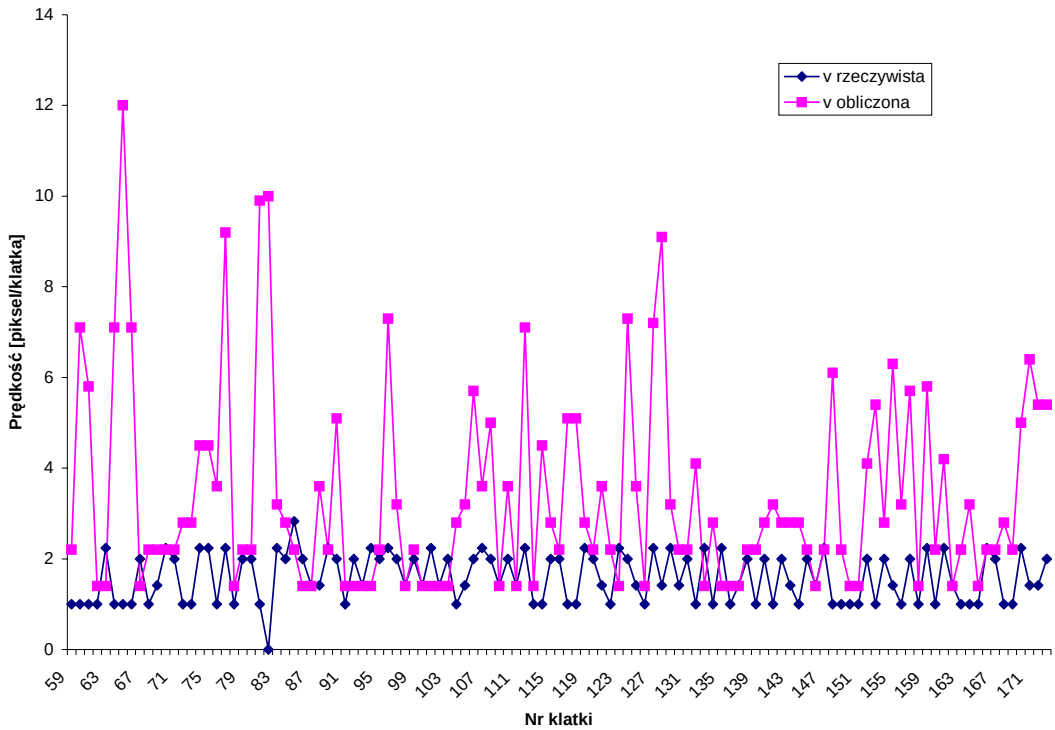
Zaproponowane rozwiązanie opiera się głównie na modelu systemu wzrokowego muchy [4], [3], [2]. Został on rozszerzony o mechanizmy lokalizacji oraz wyznaczania prędkości (por. rys. 1). W tym ostatnim wykorzystano modele pół-detektora Reichardta.

Po implementacji modelu w postaci programu komputerowego, poddano ocenie jego działanie oraz stopień przydatności. Analiza przeprowadzona na kilku przykładach wykazała, że przedstawiony model spełnił założone funkcje. Godnym podkreślenia jest fakt, że detekcja kierunku i prędkości ruchu odbywała się tylko na podstawie analizy *lokalnych* zmian natężenia rejestrowanego światła. Niepotrzebne było stosowane w technice komputerowej analizy obrazów, rozpoznawanie całego obiektu. Zmniejsza to znacznie stopień komplikacji systemu oraz zwiększa jego szybkość działania.

W trakcie testowania dokonano także analizy błędów. Do najbardziej istotnych należy zaliczyć:



Rys. 8. Porównanie chwilowych prędkości samochodu zarejestrowanych przez model z wartościami rzeczywistymi, wyznaczonymi na podstawie ręcznego porównania klatek



Rys. 9. Porównanie chwilowych prędkości pieszego zarejestrowanych przez model z wartościami rzeczywistymi, wyznaczonymi na podstawie ręcznego porównania klatek

- błędy grupowania związane z tym, że obiekt może czasami nie znajdować się w całości w obszarach detekcji (sztywna kratka nałożonej siatki),
- zbyt mała liczba aktywnych węzłów (mały obiekt),
- niedostateczny kontrast obrazu (dokładność wyznaczania prędkości wzrosła po zwiększeniu kontrastu),
- szybkie i duże zmiany oświetlenia powodujące błędne wykrycie ruchu przez lokalne detektory.

Podczas ewentualnych dalszych modyfikacji modelu należałoby w pierwszej kolejności poprawić sposób grupowania węzłów oraz jakość detekcji kierunku i prędkości. W tym celu można spróbować zastąpić równanie (1) określające aktywność węzłów zmiany zależnością opisaną w pracy [4], uwzględniającą krótkotrwałe podtrzymanie aktywności tych sygnałów.

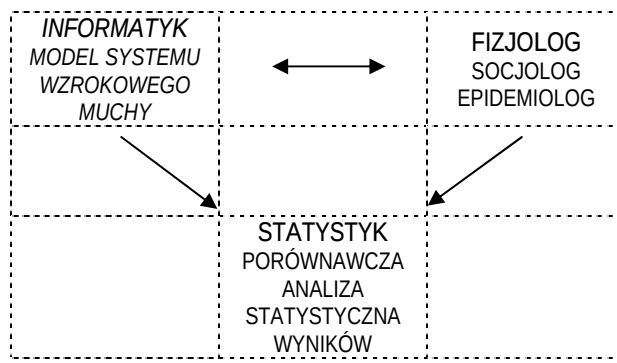
Model został przetestowany dla konfiguracji: procesor - AMD Duron 850 Mhz, pamięć RAM – 128 MB. Prędkość działania aplikacji zależna była od włączonych opcji detekcji oraz wielkości analizowanych obrazów. Dla rozdzielczości 320x240 pikseli i przy włączonych wszystkich opcjach program działał z prędkością około półtorej klatki na sekundę. Po wyłączeniu opcji lokalizacji obiektów, prędkość działania wzrosła do dwóch klatek na sekundę. W implementacji przedstawionego rozwiązania skupiono się głównie na poprawności działania programu. Z tego powodu istnieją jeszcze spore możliwości optymalizacji pod względem szybkości działania oraz ilości niezbędnej pamięci.

Podziękowania

Praca została częściowo wykonana w ramach badań własnych (umowa AGH nr 10.10.120.39)

Literatura cytowana

1. Blanchard M., Rind F. C., Verschure P. M. M. J.: Collision avoidance using a model of the locust LGMD neuron, Robotics and Autonomous Systems, 30: 17-38, 2000.
2. Borst A., Egelhaaf M.: Principles of visual motion detection, Trends NeuroScience, 12: 297-306, 1989.
3. Borst A., Egelhaaf M.: Visuomotor Coordination in Flies, w: Arbib M. A.: The Handbook of Brain Theory and Neural Networks, The MIT Press, Cambridge, Massachusetts 1995.
4. Missler J. M., Kamangar F. A.: A Neural Network for Pursuit Tracking Inspired by the Fly Visual System, Neural Networks 8(3): 463-480, 1994.
5. Zanker J., Srinivasan M., Egelhaaf M.: Speed tuning in elementary motion detectors of correlation type, Biological Cybernetics, 80: 109-116, 1999.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 247-252.

ONTOLOGIA MAMMOGRAFII – STRUKTURA MODELU, DEFINICJE I INSTANCJE POJĘĆ

MAMMOGRAPHIC ONTOLOGY – STRUCTURE OF THE MODEL, DEFINITIONS AND INSTANCES OF CONCEPTS

TERESA PODSIADŁY-MARCZYKOWSKA*, ANNA GUZIK**

*Instytut Biocybernetyki i Inżynierii Biomedycznej, Warszawa, teresa@ibib.waw.pl
**Zakład Radiologii Szpitala Grochowskiego, ul. Grenadierów 51/59, 04-073 Warszawa

Streszczenie. W pracy przedstawiono problemy występujące w dziedzinie mammografii, takie jak: zmienność interpretacji mammogramów, konieczność sfomułowania bardziej efektywnych kryteriów do oceny i opisu mammogramów i niepokojąco wysoki odsetek pominiętych lub błędnie zinterpretowanych przypadków raka piersi. Stwierdzono, że dla pełnego wykorzystania potencjału tkwiącego w badaniu mammograficznym, konieczna jest systematyzacja i objaśnienie struktury wiedzy tego fragmentu medycyny – ontologia mammografii. Podano definicję ontologii, opisano metodologię jej tworzenia i użyty edytor – Protege-2000. Opisano ogólne zasady tworzenia ontologii mammografii, podział modelu na moduły i hierarchię klas w poszczególnych modułach. Przedstawiono definicje podstawowych klas modelu i ich instancje.

Słowa kluczowe: mammografia, ontologia, modelowanie pojęć medycznych

Abstract. The paper presents mammographic ontology model of restricted domain of radiology. Method used to build the model, ontology editor, the structure of the model, class hierarchy, definitions of the main notions in the domain and their instances have been presented.

Key words: mammography, ontology, medical concept modeling

1. Wstęp – potrzeba modelu w dziedzinie mammografii

Nowotwory złośliwe są drugą (po chorobach układu krążenia) przyczyną śmierci kobiet w Polsce. Taka struktura zgonów utrzymuje się od 1963 r. [4]. Wśród chorób nowotworowych rak piersi ma pierwsze miejsce zarówno jeśli chodzi o umieralność, jak i zapadalność [4]. Fakty te wystarczają, aby uznać nowotwory piersi za ważny problem społeczny.

Najczęściej używanym badaniem mającym na celu wykrycie nowotworów piersi we wczesnym stadium jest aktualnie skrining mammograficzny [24]. Każde badanie obrazowe jest ze swojej natury wysoce subiektywne, w mammografii dodatkowo sytuację utrudnia brak tzw. „normy". Nie dziwią więc prace świadczące, że radiolodzy często dość znacznie różnią się w swojej ocenie wyników tego podstawowego w diagnostyce raka piersi badania, i w miarę upływu czasu sytuacja nie ulega poprawie. Dowodzą tego następujące fakty:

1. Rok 1996, stwierdzono dużą zmienność dokładności interpretacji mammogramów w populacji radiologów amerykańskich [1] mimo zadowalającej jakości zdjęć i sprzętu radiologicznego [1] (wg najlepszej wiedzy autorów podobnej oceny nigdy nie przeprowadzono w Polsce),
2. Rok 1998, nadal stwierdza się [17] znaczną zmienność interpretacji mammogramów wśród amerykańskich radiologów. Postulowana jest konieczność sformułowania bardziej efektywnych kryteriów dla oceny i opisu mammogramów,
3. Rok 2002, stwierdzono, że od 10% - 30% przypadków raka widocznych retrospektywnie było pominiętych lub błędnie zinterpretowanych w pierwszej ocenie radiogramów [20].

Należy sobie też uświadomić, że diagnostyka i leczenie chorób piersi wymaga współpracy lekarzy wielu specjalności. Ginekologa lub lekarza pierwszego kontaktu wykonującego badanie kliniczne piersi i zlecającego mammografię, specjalisty radiologa, który ją interpretuje, często innego radiologa wykonującego biopsję cienkoigłową, chirurga wykonującego rozszerzoną biopsję chirurgiczną lub operację usunięcia zmiany, patologa oceniającego materiał biopsyjny, wreszcie radioterapeuty i onkologa planujących radio- i chemioterapię. Dla powodzenia procesu diagnostycznego

i terapeutycznego jasna i jednoznaczna komunikacja pomiędzy wszystkimi specjalistami jest niemal tak ważna, jak ich zawodowa kompetencja. Integralną częścią badania mammograficznego jest raport mammograficzy. Niestety, niska jakość informacji medycznych zawartych w raportach uniemożliwia ich użycie w systemach wspomagania decyzji klinicznych lub jako zbioru indeksów dla obrazowej bazy danych złożonej z mammogramów [22].

Wszystkie przytoczone powyżej fakty wskazują, że dla pełnego wykorzystania potencjału tkwiącego w badaniu mammograficznym konieczna jest systematyzacja i objaśnienie struktury wiedzy tego fragmentu medycyny oraz wprowadzenie mechanizmu umożliwiającego współdzielenie wiedzy zarówno przez ludzi, jak i systemy komputerowe. Wymagania te spełnia pojęciowy model mammografii – jej ontologia.

2. Definicja ontologii

Istnieje wiele definicji artefaktu, jakim jest ontologia. W niniejszej pracy przyjęto najczęściej cytowaną, a zatem subiektywnie najpopularniejszą definicję, podaną w 1993 roku przez Thomasa Grubera [12], wg której ontologia jest formalną specyfikacją wspólnej warstwy pojęciowej.

Przez warstwę pojęciową należy rozumieć abstrakcyjny model zjawisk w ograniczonym wycinku rzeczywistości, otrzymany poprzez identyfikację istotnych pojęć z nim związanych. Oznacza to, że dla pewnej dziedziny zidentyfikowano występujące w niej pojęcia (obiekty, zdarzenia, stany itp.) i relacje pomiędzy nimi. Dla przykładu, wynikiem bardzo pobieżnej ontologicznej analizy medycyny są takie pojęcia, jak: choroba, symptom, terapia i relacje pomiędzy nimi, takie jak: „choroba wywołuje symptomy", „terapia leczy choroby". Przymiotnik formalna oznacza, że model musi być czytelny dla maszyny, co wyklucza użycie języka naturalnego. Specyfikacja oznacza, że definicje pojęć i relacji muszą być jednoznacznie sformułowane, wspólna odnosi się do faktu, że wiedza zawarta w ontologii powinna być akceptowana przez ogół użytkowników. Definicja Gruber'a w swej prostocie stara się ująć jak najwięcej charakterystycznych cech ontologii.

3. Cel ontologii mammografii

Podstawowe scenariusze zastosowań ontologii zostały opisane w [14]. Jednym z nich jest użycie ontologii jako specyfikacji przy projektowaniu i tworzeniu oprogramowania dla danej dziedziny. Korzyści odnoszone przy tym podejściu to: prawidłowa dokumentacja, wiarygodność i niezawodność oraz możliwość ponownego użycia wiedzy.

Zgodnie z powyższym scenariuszem, podstawowym celem ontologii mammografii jest stworzenie specyfikacji dla następujących aplikacji:

1. baza danych przechowująca dane kliniczne pacjentów ze schorzeniami piersi i treść raportów mammograficznych,
2. interfejs do bazy danych wspomagający tworzenie, wyszukiwanie i prezentację raportów mammograficznych. Zastosowanie tego interfejsu ma na celu poprawę jakości informacji medycznych w raportach mammograficznych.

Dla realizacji głównego celu niezbędne jest w pierwszym rzędzie:

1. identyfikacja podstawowych pojęć z dziedziny mammografii i ich hierarchii,
2. stworzenie formalnych definicji pojęć używanych w mammografii,
3. uzgodnienie wspólnego słownika używanego do opisu mammogramów i redagowania raportów mammograficznych.

3. Budowa ontologii

Sytuację w dziedzinie budowy modeli ontologicznych najtrafniej podsumowuje [21] stwierdzając, ze w chwili obecnej konstrukcja ontologii ma więcej wspólnego z opartą na doświadczeniu inżynierią niż z nauką, ale że istniejące metodyki stanowią niepodważalnie duży wkład na drodze w kierunku powstania spójnego podejścia do projektowania ontologii.

Ponieważ nie ma jak dotąd jedynej, zalecanej przez wszystkich metody budowania ontologii [21], przeanalizowano cztery metody budowy modeli ontologicznych: metodę KBSI, TOVE, EMA i metonologię, i na tej podstawie zaproponowano następujące etapy budowy modelu mammogarfii:

1. Cel ontologii.
2. Akwizycja wiedzy.
3. Budowa ontologii.
   a. Identyfikacja klas.
   b. Budowa struktury klas.
   c. Budowa definicji klas.
   d. Budowa instancji.
   e. ** Specyfikacja ograniczeń.
4. **Ocena wyników.
5. **Wdrożenie.
6. **Utrzymanie.

Decyzję, jakie pojęcia należy uznać za istotne w danej dziedzinie, zależą od projektowanych aplikacji i decyzję tę pozostawia się zespołowi modelującemu dziedzinę [2]. Przy tworzeniu modelu używano manualnych metod akwizycji wiedzy, takich jak przegląd literatury [5], [6], [13], [18], analiza raportów mammograficznych i wywiady z radiologiem.

Etapy oznaczone ** nie są omawiane w aktualnym opracowaniu.

4. Narzędzie – edytor Protege-2000

Edytory ontologii to klasa programów umożliwiających zapis ontologii w formie czytelnej dla komputera, jak również podstawowe funkcje takie, jak: jego przeglądanie, modyfikację i wizualizację. Obecnie istnieje kilkadziesiąt takich edytorów. Opis i ocenę najbardziej znanych można znaleźć w [7]. Spośród opisanych tam programów wybrano edytor Protege-2000 w wersji 2.1.2. Protege-2000 został stworzony 15 lat temu w Departamencie Informatyki Medycznej Uniwersytetu Stanford. W swojej pierwszej wersji był to specjalistyczny program wspomagający akwizycję wiedzy dla medycznych systemów ekspertowych (planowanie terapii onkologicznych). Obecnie jest używany do tworzenia modeli z różnych dziedzin wiedzy (od medycyny i genetyki po elektronikę, geografię czy awiację, i ma ponad 8000 zarejestrowanych użytkowników). Może służyć zarówno do tworzenia samych ontologii, jak również na ich podstawie Baz Wiedzy. Protege-2000 jest narzędziem darmowym typu Open-Source, językiem amplementacji jest Java.

4.1. Model wiedzy Protege-2000

Model wiedzy Edytora Protege-2000 [10] jest oparty na ramach [8], [19], [15]. Ramowe modele reprezentacji wiedzy (RMRW) zostały z sukcesem zastosowane w wielu dziedzinach [11], [3], [16]. W RMRW pojęcia reprezentują zbiory obiektów o wspólnych własnościach. Rama reprezentująca pojęcie zawiera sloty opisujące jego atrybuty. Fasety opisują właściwości slotów, takie jak np. zbiór dozwolonych wartości lub kardynalność.

Zbiory klas, slotów i faset i ograniczeń są rozłączne, modelowany obiekt może być albo klasą, albo atrybutem, albo fasetą. Baza wiedzy Protege-2000 zawiera model i instancje klas wraz z właściwymi wartościami dla atrybutów. Reasumując, ontologia w Protege-2000 składa się z następujących obiektów:

- klas odpowiadających pojęciom i ich instancji,
- slotów – opisujących własności klas i instancji,
- faset opisujących własności slotów i nakładających na nie ograniczenia,
- aksjomatów pozwalających nakładać dodatkowe ograniczenia na wartości atrybutów.

5. Ogólne zasady tworzenia modelu i hierarchii klas

W [23] postulowana jest modularna budowa modelu jako podstawowy warunek dla osiągnięcia ponownego użycia wiedzy. Model mammografii składa się z rozdzielnych modułów. Dla każdego modułu:

- zidentyfikowano istotne w nim pojęcia,
- stworzono ich hierarchię,
- dokonano separacji pojęć na elementarne (niepodzielne) i złożone, używając wiedzy z dziedziny medycyny i mammogarfii [5], [6], [13], [18],
- zdefiniowano pojęcia złożone przy użyciu pojęć elementarnych.

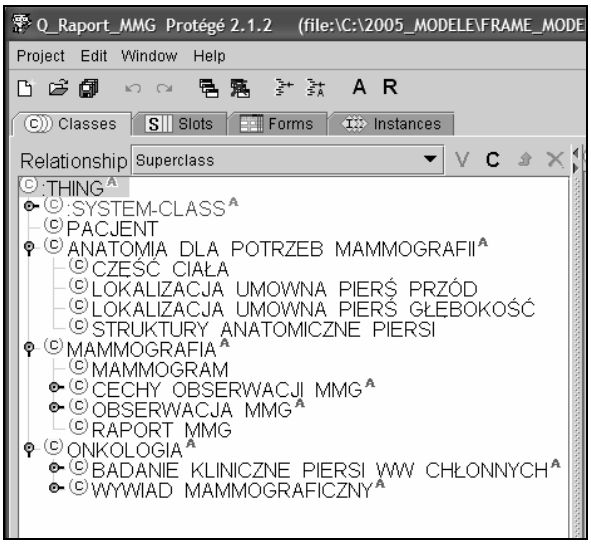
Organizacja pojęć w postaci struktury hierarchicznej jest możliwa dzięki relacji „is-a", interpretowanej jako logiczna subsumcja oraz atrybutowi „rola", określanemu dla każdej klasy.

Relacje w modelu Protege-2000 są to sloty, dla których zbiorem dozwolonych wartości są ramy. Atrybut „rola", określany dla wszystkich klas definiowanych w modelu, może przybierać jedną z wartości: abstrakcyjna i konkretna. Klasy abstrakcyjne nie mają bezpośrednich instancji i służą do tworzenia hierarchii klas modelu, klasy typu konkretnego mają bezpośrednie instancje składające się na bazę wiedzy modelu.

6. Opis modelu mammografii

6.1. Ogólna struktura modelu

Pierwszy poziom modelu określa, jakie obiekty zawiera opisywany wycinek rzeczywistości. W modelu mammografii na pierwszym poziomie znajdują się rozdzielne klasy-pojęcia: PACJENT, ANATOMIA DLA POTRZEB MAMMOGRAFII, MAMMOGRAFIA i ONKOLOGIA połączone klasą THING, będącą korzeniem hierarchii klas w Protege-2000. Klasa PACJENT typu konkretnego ma bezpośrednie instancje, pozostałe klasy typu abstrakcyjnego są klasami nadrzędnymi trzech modułów składających się na model mammografii. Pierwszy poziom modelu i fragment poziomu drugiego przedstawiono na poniższej rycinie.



Rys. 1. Fragment zrzutu ekranowego edytora Protege-2000, przedstawiający w postaci struktury drzewiastej. hierarchię pierwszego poziomu modelu mammografii i fragment drugiego

Moduł ANATOMIA DLA POTRZEB MAMMOGRAFII zawiera klasy, których instancje służą do określenia lokalizacji obserwacji mammograficznych. Moduł ONKOLOGIA zawiera niezbędne do prawidłowej interpretacji mammogramu wyniki klinicznego badania piersi i węzłów chłonnych oraz wywiad pacjenta. Moduły te nie będą dokładnie opisywane w niniejszym opracowaniu.

Podstawowe znaczenie w ontologii mammografii ma moduł MAMMOGRAFIA, zawierający formalne definicje kluczowych klas CECHY OBSERWACJI MMG, OBSERWACJA MMG i RAPORT oraz klasy MAMMOGRAM.

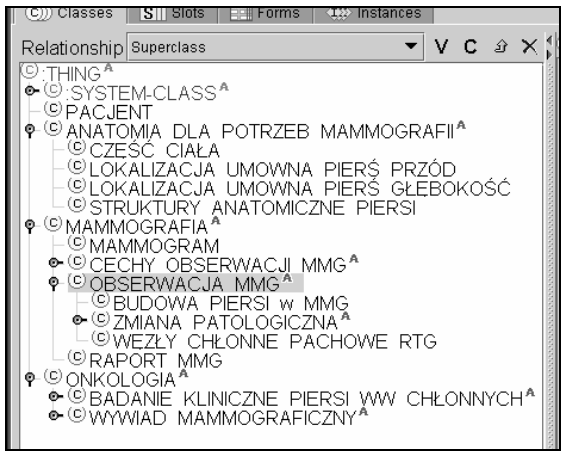
6.2. Struktura modułu MAMMOGRAFIA

Pierwszy poziom modułu MAMMOGRAFIA tworzą klasy: MAMMOGRAM, CECHY OBSERWACJI MMG, OBSERWACJA MMG I RAPORT (rys. 1).

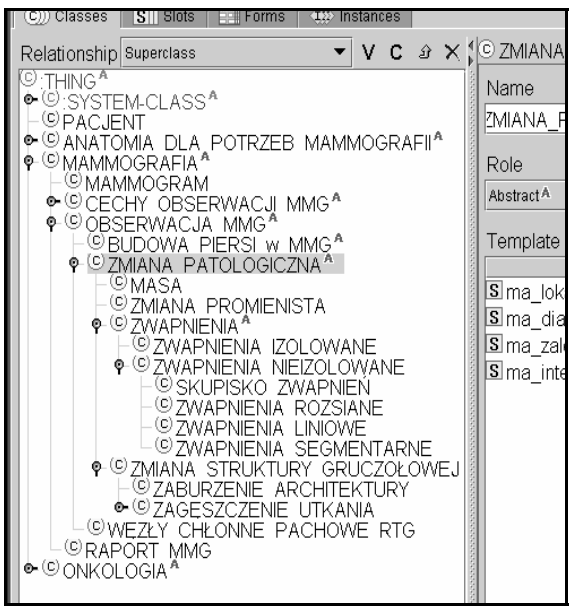
Klasa OBSERWACJA MMG

Klasa OBSERWACJA MMG zawiera jako swoje podklasy typy obserwacji występujące w raportach mammograficznych. Typowy raport rozpoczyna się od opisu budowy piersi w RTG, następnie opisywane są uwidocznione zmiany patologiczne, kończy się oceną stanu węzłów chłonnych. Strukturę klasy OBSERWACJA MMG przedstawia rys. 2.

Klasy BUDOWA PIERSI i WĘZŁY CHŁONNE są klasami konkretnymi. Abstrakcyjna klasa ZMIANA PATOLOGICZNA jest nadrzędną klasą hierarchii wszystkich zmian patologicznych w modelu mammografii. Hierarchię tę, podyktowaną przez cechy morfologiczne zmian w piersi, przedstawia rys. 3.



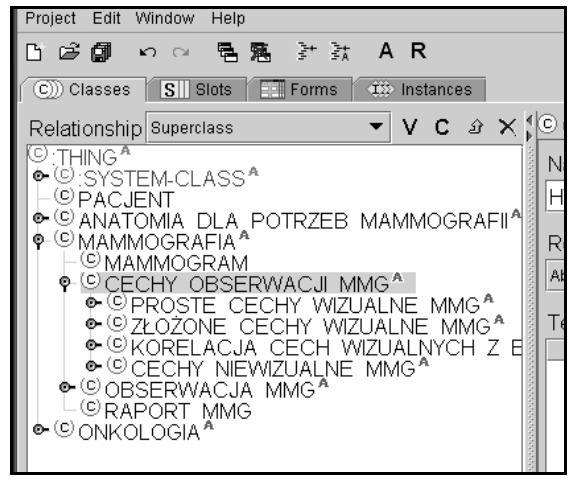
Rys. 2. Struktura klasy OBSERWACJA MMG stanowi odbicie zawartości raportu mammograficznego



Rys. 3. Klasyfikacja zmian patologicznych w piersi

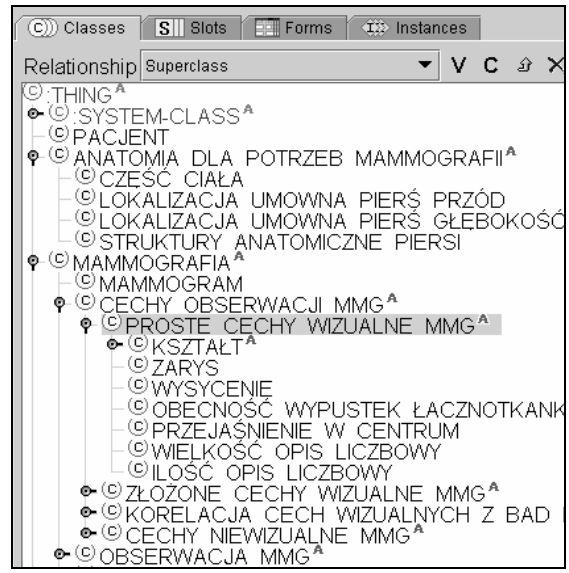
Klasa CECHY OBSERWACJI MMG

Klasa CECHY OBSERWACJI MMG jest klasą typu abstrakcyjnego. Podstawą stworzenia hierarchii jej podklas jest spostrzeżenie, że cechy, jakie opisują zmianę mammograficzną, różnią się przede wszystkim poziomem wiedzy potrzebnej do jej określenia. Ilość wiedzy potrzebna do określenia wartości cechy jest najmniejsza dla instancji klasy PROSTE CECHY WIZUALNE MMG, rośnie kolejno dla klas ZŁOŻONE CECHY WIZUALNE I KORELACJA CECH WIZUALNYCH Z BADANIEM KLINICZNYM I WYWIADEM i jest największa dla klasy CECHY NIEWIZUALNE MMG. Instancje klas opisujących cechy zmian mammograficznych tworzą zbiór atrybutów dla obserwacji mammograficznych, będących instancjami klas z modułu OBSERWACJA MMG. Na poniższej rycinie przedstawiono hierarchię podklas klasy CECHY OBSERWACJI MMG.



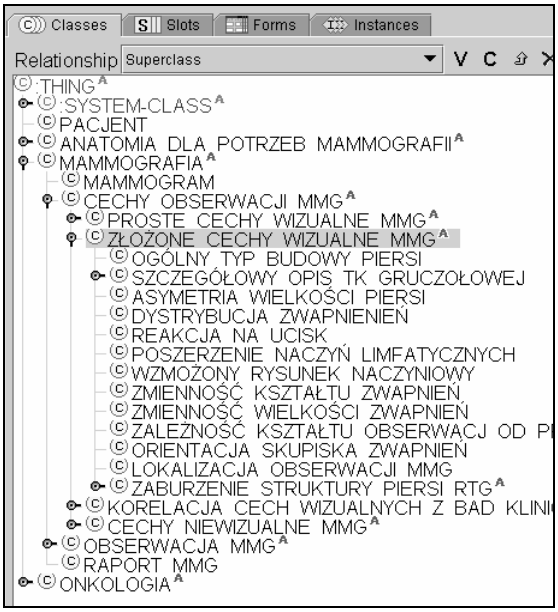
Rys. 4. Klasa CECHY OBSERWACJI MMG wraz z podklasami

Klasa PROSTE CECHY WIZUALNE MMG zawiera jako podklasy zawiera elementarne cechy obserwacji mammograficznych, takie jak: kształt, wysycenie, charakter zarysu czy wielkość zmiany.



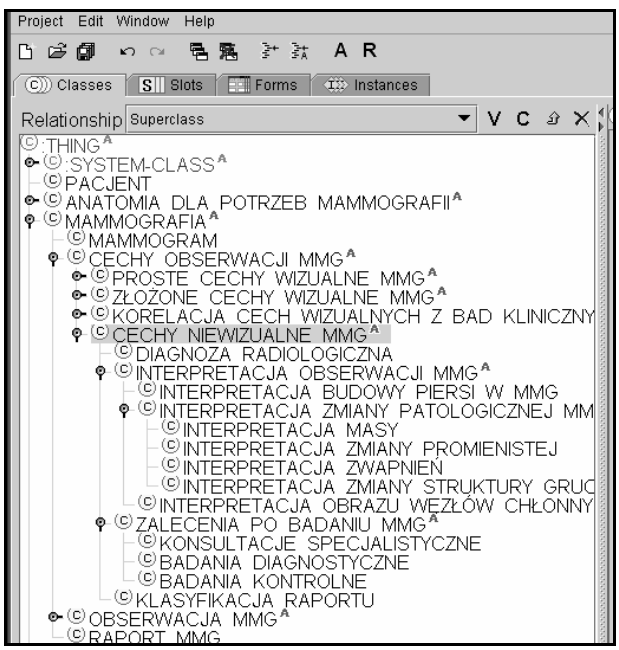
Rys. 5. Proste cechy wizualne obserwacji mammograficznych

Klasy ZŁOŻONE WIZUALNE MMG i KORELACJA CECH WIZUALNYCH Z BADANIEM KLINICZNYM I WYWIADEM grupują cechy, których powstanie wymaga od radiologa większej wiedzy niż dla prostych cech wizualnych i często złożonych operacji wizualnych, np. porównania pewnych obserwacji obu piersi, porównania rodzajów kształtów i wielkości zwapnień czy porównania pewnych obrazów z normą (np. wzmożony rysunek naczyniowy). Podklasy klasy ZŁOŻONE CECHY WIZUALNE MMG przedstawia rysunek 7.



Rys. 7. Złożone cechy wizualne obserwacji mmg

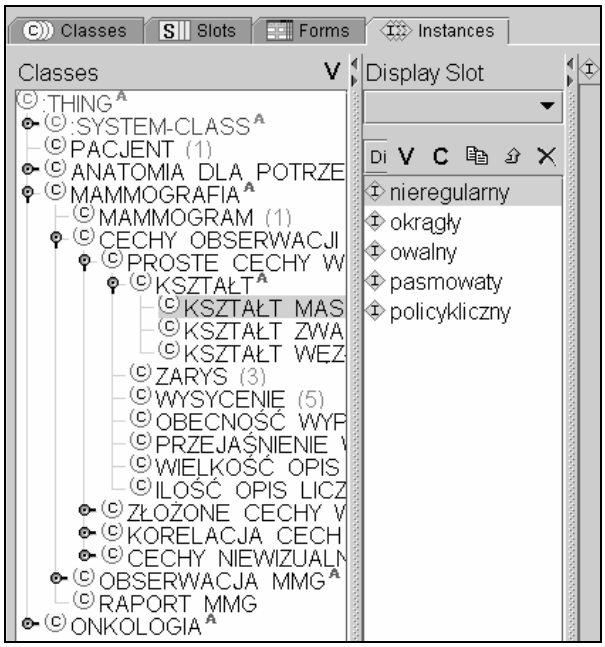
Najważniejszymi dla lekarza zlecającego mammografię cechami obserwacji mammograficznych są ich cechy niewizualne. Do cech tych zaliczono: diagnozę radiologiczną, interpretację poszczególnych typów obserwacji i zalecenia dotyczące dalszej diagnostyki lub kontroli. Powstanie wartości tych cech dla zmiany wymaga od radiologa użycia wiedzy medycznej i doświadczenia. Strukturę klasy CECHY NIEWIZUALNE MMG przedstawiono na rysunku 8.



Rys. 8. Klasy reprezentujące cechy niewizualnych obserwacji mmg

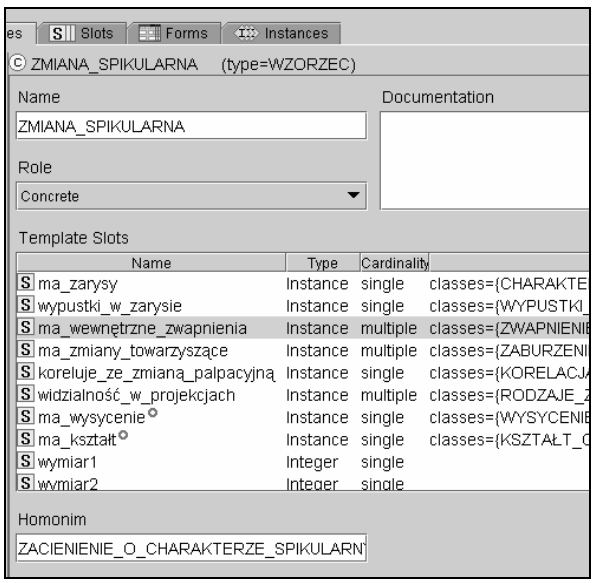
7. Definicje i instancje klas modelu

Atrybuty pojęć reprezentujących zmiany w mammogramie są slotami typu *instance,* tzn. mogą przyjmować jako wartość instancje pewnej klasy zadeklarowanej jako dziedzina slotu. Na przykład instancje klasy KSZTAŁT MASY I ZMIANY SPIKULARNEJ przyjmują wartości: okrągły, owalny nieregularny itp. (patrz poniższy rysunek)



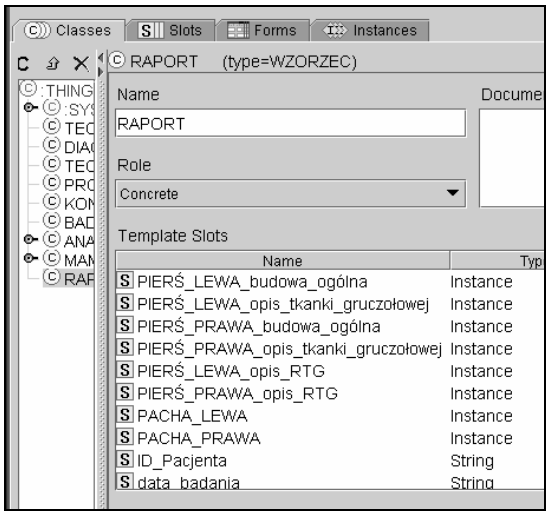
Rys. 9. Prawy panel – zbiór instancji dla klasy KSZTAŁT wyróżnionej w lewym panelu

Definicja klasy ZMIANA SPIKULARNA składa się w większości z atrybutów typu *instance* (patrz rys. 10), wyjątkami są atrybuty opisujące wymiary zmiany. Atrybut ma_kształt jest również atrybutem typu *instance,* którego dziedziną są instancje klasy KSZTAŁT.



Rys. 10. Definicja klasy ZMIANA SPIKULARNA Większość atrybutów zmiany to atrybuty typu *instance*

Definiując klasę RAPORT, posłużono się tą samą zasadą. Oprócz takich atrybutów, jak identyfikator pacjenta, data badania czy nazwisko oceniającego badanie radiologa, większość atrybutów przyjmuje jako wartości instancje klas występujących w module CECHY OBSERWACJI MMG.



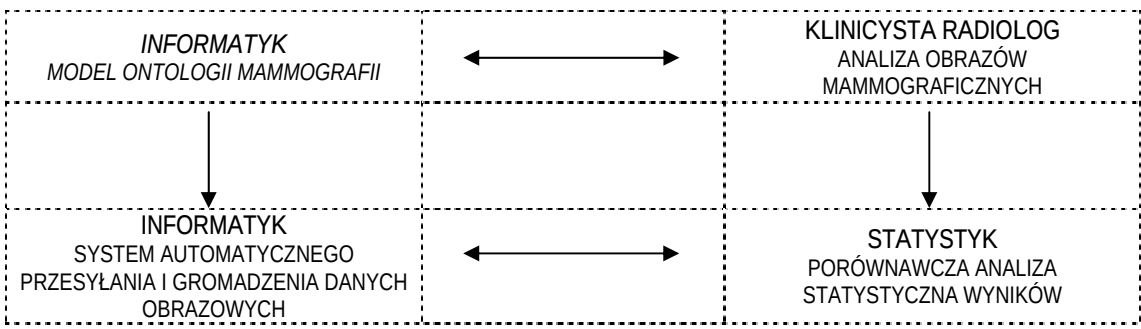
Rys. 10. Definicja klasy RAPORT

8. Podsumowanie

Jak stwierdzono w [2], już sam proces analizy prowadzący do stworzenia modelu ontologicznego systematyzuje i objaśnia strukturę wiedzy w danej dziedzinie. Jednak aby wykorzystać model jako specyfikację musi on zostać zweryfikowany i zaakceptowany przez gremium radiologów

Literatura cytowana

1. Beam C. A., Loyde P. M., Sullivan D. C.: Variability in Screening Mammography by US Radiologists. Findings from a National Sample, Archives of Internal Medicine vol 145 No.2 January 22 1996.
2. Chandrasekaran, B., Josephson, J. R., Benjamins V. R.: What Are Ontologies, and Why We Need Them ?, IEEE Inteligent Systems pp. 20-26, 1999.
3. Ciminio J. J.: From Data to Knowledge through Concept Orinted Terminologies:Experience with the Medical Entities Dictionary, Journal of American Medical Infromation Association, vol 7, pp. 288-297.
4. Didkowska J., Wojciechowska U., Tarkowski W., Zatoński W.: Nowotwory Złośliwe w Polsce w 2003 roku, Centrum Onkologii – Inst. Im. M. Skłodowskiej-Curie Zakład Epidemiologii i Prewencji Nowotworów, Krajowy Rejestr Nowotworów, Warszawa 2003.
5. D'Orsi C. J., Kopans D. B.: Mammogarphic Feature Analysis, Seminars in Roentgenology, 28, pp. 204-229, 1998.
6. Dziukowa J.: Mammografia w Diagnostyce Raka Sutka Copyright BEL CORP Scientific Publ. Co., Warszawa 1998.
7. Duineveld A. J., Stoter R., Weiden M. R., Kanepa B., Benjamins V. R., WonderTools ? A Comparative Study of Ontological Engineering Tools. http://ncs.scienceuva.nl/wondertools/html/paper.htm
8. Fikes R., Kechler T.: The Role of Frame-Based Repreasentations in Reasoning Commun, ACM, vol. 28, pp. 904-920, 1985.
9. Pandha H. S., et. all.: Breast cancers missed in the prevalent screening round, Radiology 2002.
10. Friedman Noy N., Fergerson R. W., Musen. M. A.: The Knowledge Model of Protege-2000: Combining Interoperability and Flexibility, SMI technical report, Stanford University, School of Medicine, 2000.
11. Felbaum C. (Ed.): WordNet an Electronic Lexical Database, MIT Press, Cambridge, MA, 1998.
12. Gruber T. R.: A transactional Approach to Portable Ontology Specification, Knowledge Acquisition 5 (2), pp. 199-221, 1993.
13. Iwaszkiewicz K.: Atlas Mammografii, PZW, Warszawa 1993.
14. Jasper R., Uschold A.: A Framework for Understanding and Classifying Ontology Applications. W Twelfth Workshop on Knowledge Acquisition Modeling and Menagement KAW' 99, 1999.
15. Karp P. D: The Design Space of Frame Knowledge Representation Systems, SRI International Artificial Intelligence Center 1992.
16. Karp P. D., Riley M. et all.: The EcoCyc Database, Nucleic Acid. Res., vol. 30, pp. 56-62, 2002.
17. Kerlikowske K. et all.: Difference in Mammography Reading Journal of the National Cancer Institute, 1998.
18. Kopans D. B., D'Orsi C. J., Adker D. D.: Breast Imaging Reporting and Database System (BI-RADS), American College of Radiology, 1993.
19. Minsky M.: A Framework for representing Knowledge. Readings In Knowledge Representation., Levesque, H. (ed.), Morgan Kaufmann, Los Altos CA, pp. 246-262 , 1987.
20. Pinto H. S., Martins J. P.: Ontologies: How can they be built, Knowledge and Information Systems, (6) pp. 441–464, 2004.
21. Podsiadły-Marczykowska T.: Ocena Jakości Opisowych Danych Medycznych na Przykładzie Raportu Mammograficznego, Acta Bio-Optica Medica et Informatica Medica, vol. 9, pp. 105-110, 2003.
23 Rector A. L.: Modularisation of Domain Ontologies Implemented in Description Logics and Related Formalisms including OWL, in Proceedings K-CAP'03, October 23-25 pp. 121-128, 2003.
24. Tabar L. et all.: Beyond randomized Controlled Trials: organized Mammographic Screening Substantially Reduces Breast Carcinoma Mortality, Cancer 2001.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 253-258.

NEURONOWY MODEL DOŚWIADCZENIA PAWŁOWA. BADANIE ZESPOŁU ODRUCHÓW WARUNKOWYCH

NEURAL MODEL OF THE PAVLOV EXPERIMENT. INVESTIGATION OF COMPLEX CONDITIONED REFLEXES

KRZYSZTOF SOKALSKI

Katedra Matematyki i Informatyki, Akademia Polonijna, Częstochowa,
Instytut Informatyki, Politechnika Częstochowska, sokalski@el.pcz.czest.pl

Streszczenie. Przetestowano neuronowy model doświadczenia Pawłowa. Model opisuje mechanizm odruchu warunkowego i jest podstawą do budowy modeli sterowania odruchami. Przedstawiono uogólnienie modelu na zespół odruchów warunkowych. Zbadano różnice między własnościami pojedynczego odruchu warunkowego a własnościami zespołu odruchów warunkowych.

Słowa kluczowe: odruchy warunkowe, odruchy bezwarunkowe, doświadczenie Pawłowa, sztuczny neuron, reguła Hebba, odruch warunkowy drugiego rzędu, zespół odruchów warunkowych

Abstract. Neural model of the Pavlov experiment has been tested. The model explains the mechanism of conditioned reflex and it is a basis for the construction of models controlling the reflexes. An extension of the model into complex conditioned reflexes has been presented. Differences between properties of the single conditioned reflex and properties of the complex conditioned reflexes have been investigated.

Key words: conditioned and unconditioned reflexes, the Pavlov experiment, artificial neuron, the Hebb rule, conditioned reflex of the second order, complex conditioned reflexes

1. Wstęp

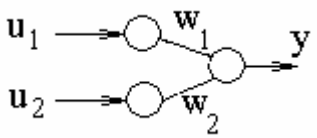
Neuronowy opis doświadczenia Pawłowa, oparty o pojęcia *instar* i *outstar* (wprowadzone przez Grossberga [3]), prowadzi do układu nieliniowych równań różniczkowych zwyczajnych. Rozwiązywanie tych równań jest trudne z racji ich nieliniowości. Z tego też powodu koncepcje te są mało przydatne dla biologów, psychologów, socjologów czy też ekonomistów w wyjaśnianiu zjawisk zachodzących w odpowiadających im dziedzinach nauki i charakteryzujących się pewnego rodzaju histerezą (*histereza typu odruchu warunkowego*). Stąd powrót do prostego modelu rekurencyjnego, wyjaśniającego między innymi mechanizm powstawania odruchu warunkowego. Mimo swojej prostoty model umożliwia projektowanie doświadczeń z zakresu neurobiologii, przewidywanie wyników oraz ich interpretację. Jako przykłady zastosowań przedstawiono klasyczne doświadczenie Pawłowa, powstawanie odruchu warunkowego drugiego rzędu oraz blokowanie odruchu warunkowego drugiego rzędu.

W części badawczej przedstawiono wykorzystanie opisanego modelu do projektowania i badania zespołu odruchów warunkowych.

Wszystkie przedstawione w pracy wyniki mają charakter jakościowy. Z tego też powodu wszędzie tam, gdzie efekt zapominania nie jest istotny w wyjaśnianiu zjawisk, został on pominięty.

2. Minimalny model odruchów warunkowych

Minimalny model mogący wyjaśnić mechanizm odruchu warunkowego to perceptron Rosenblatta [4], [5] (rys. 1) składający się z dwóch neuronów wejściowych i jednego neuronu wyjściowego (przetwarzającego).



Rys. 1. Minimalna sieć neuronowa do modelowania doświadczenia Pawłowa

Neurony wejściowe są tutaj namiastką obwodowego systemu nerwowego, natomiast neuron przetwarzający można uznać za namiastkę centralnego systemu nerwowego.

Sygnały wejściowe u_1 i u_2 odpowiadają odpowiednio bodźcom smakowym i akustycznym, odpowiednio w_1 i w_2 są wagami połączeń między neuronami wejściowymi i neuronem wyjściowym, y oznacza sygnał sterujący gruczołami wydzielającymi ślinę.

Zakłada się, że wartości sygnałów u_1, u_2 i y są binarne, a funkcja aktywacji jest skokiem jednostkowym. Przetwarzanie

sygnałów wejściowych podlega regułom działania warstwowych sieci jednokierunkowych:

$$y = F(u_1 \cdot w_1 + u_2 \cdot w_2) \qquad (1)$$

gdzie *F* jest funkcją aktywacji.

Modyfikacja wag połączeń podlega regule Hebba [6]:

$$w_i(n+1) = w_i(n) + \alpha\, u_i\, y \qquad (2)$$

gdzie: $w_i(n)$ jest *i*-tą wagą w *n*-tym etapie doświadczenia Pawłowa, ***α*** jest parametrem określającym szybkość modyfikacji wag (parametr uczenia).

Powstaje problem związany z ustaleniem wartości początkowych w_1 i w_2. Zakłada się, że „pies" rodzi się z przypadkowymi wartościami tych wag (cechy wrodzone). Ich wartości decydują o zdolnościach adaptacyjnych młodego „organizmu". Do doświadczenia Pawłowa wybiera się osobnika, którego wartości początkowe wag predysponują go do przeżycia oraz stabilnej egzystencji.

3. Wybór osobnika do doświadczeń

Rozważa się cztery kategorie osobników chrakteryzowanych przez wartości początkowe wag I, II, III, IV:

Tab. 1. Wartości początkowe wag połączeń dla czterech kategorii osobników

	I	II	III	IV
w_1	0	1	0	1
w_2	0	0	1	1

Spośród typów wymienionych w tab.1 wybiera się do doświadczeń osobnika najbardziej predysponowanego do przeżycia. W tym celu bada się reakcje osobników na różne kombinacje sygnałów wejściowych A, B, C, D:

Tab. 2. Tabela kombinacji sygnałów wejściowych

	A	B	C	D
u_1	0	1	0	1
u_2	0	0	1	1

Podając na wejście sieci z wagami początkowymi I, II, III i IV kolejno sygnały A, B, C i D, otrzymuje się 16 odpowiedzi *y*:

Tab. 3. Odpowiedzi typów I, II, III i IV na sygnały A, B, C i D

	I	II	III	IV
A	0	0	0	0
B	0	1	0	1
C	0	0	1	1
D	0	1	1	1

Z tab. 3 wynika, że najlepszym osobnikiem z punktu widzenia doświadczenia Pawłowa jest typ II. Uzasadnienie tego wyboru jest następujące: typy I i III nie reagują na bodźce smakowe i tym samym są gorzej przystosowane do życia niż typy II i IV, które na ten bodziec reagują prawidłowo. Spośród dwóch ostatnich typów wybiera się typ II, ponieważ typ IV reaguje na bodziec akustyczny tak samo jak na bodziec smakowy i tym samym nie nadaje się do badań odruchu warunkowego.

4. Doświadczenie Pawłowa [2]

Poszczególne etapy neuronowego modelu doświadczenia Pawłowa przedstawiane są na tle odpowiadających im etapów doświadczenia klasycznego.

Etap 1.
bb: $u_1 = 1, u_2 = 0$ podanie bodźca smakowego *bb=B*,
y=1, $w_1(1) = 2, w_2(1) = 0$ „pies" wydziela ślinę (odruch bezwarunkowy), wzrasta wrażliwość „psa" na bodziec *bb=B*.

Etap 2.
bw: $u_1 = 0, u_2 = 1$ podanie bodźca akustycznego *bw=C*,
y=0, $w_1(2) = 2, w_2(2) = 0$ brak odruchu.

Etap 3.
bb+bw: $u_1 = 1, u_2 = 1$ podanie bodźca smakowego i akustycznego,
y=1, $w_1(3) = 3, w_2(3) = 1$ „pies" wydziela ślinę (odruch warunkowy), wzrasta wrażliwość „psa" na bodziec *B* oraz pojawia się wrażliwość na bodziec C ($w_2(3)>0$).

Etap 4.
bw: $u_1 = 0, u_2 = 1$ podanie bodźca akustycznego *bw=C*,
y=1, $w_1(4) = 3, w_2(4) = 2$, "pies" wydziela ślinę (odruch warunkowy), wzrasta wrażliwość „psa" na bodziec C.

Tab. 4. Tabela zbiorcza przedstawiająca wszystkie etapy doświadczenia Pawłowa

n	in	IN	out y	OUT	$w_1(n)$	$w_2(n)$
1	B	bb	1	ob.	2	0
2	C	bw	0	-	2	0
3	D	bb+bw	1	ow	3	1
4	C	bw	1	ow	3	2

n – numer kolejny etapu, in – sygnały wejściowe, IN – interpretacja wejścia, y – sygnały wyjściowe, OUT – interpretacja sygnałów wyjściowych, $w_{1,2}(n)$ – wartości wag w n-ty etapie, bb – bodziec bezwarunkowy, ob – odruch bezwarunkowy, bw – bodziec warunkowy, ow – odruch warunkowy.

Pawłow szukał odpowiedzi na następujące pytanie: Skoro odruchy warunkowe nie są wrodzone, to w jaki sposób się je nabywa [7]? Neuronowy model jego doświadczenia daje następującą odpowiedź:

Powstanie odruchu warunkowego jest wynikiem wykształcenia się połączenia pobudzającego ($w_2 > 0$) pomiędzy receptorem bodźca neutralnego a efektorem odpowiedzialnym za odruch. W wyniku równoczesnego podawania bodźca bezwarunkowego i bodźca warunkowego D (etap 3) reguła Hebba generuje różną od zera wartość wagi w_2 odpowiedzialnej za to połączenie.

Połączenie to umożliwia w ostatnim etapie doświadczenia wywołanie odruchu za pomocą bodźca neutralnego C.

5. Sterowanie odruchami warunkowymi

Odruchy warunkowe nie zawsze są pożądane [9]. Powstaje, zatem pytanie, czy można nimi sterować? Oczywistym jest, że częste podawanie bodźca bezwarunkowego powoduje wzrost wartości wagi w_1 (tab. 4):

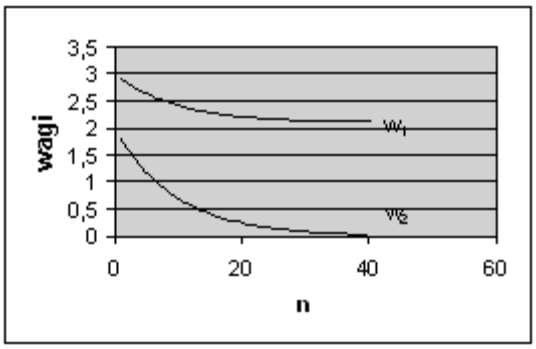
$w_1(0) = 1, w_1(1) = 2, w_1(2) = 2, w_1(3) = 3$

Oznacza to wzrost wrażliwości efektora na ten bodziec. W przypadku stosowania innych modeli (np. Oja [8]) wzrost wag będzie ograniczony i wówczas podawanie bodźca bezwarunkowego będzie podtrzymywać wysoką wrażliwość efektora na ten bodziec. Jeżeli w tym czasie nie będzie podawany bodziec warunkowy, to wrażliwość efektora na ten bodziec będzie słabła z powodu efektu zapominania [3]. Efekt zapominania uwzględnia się w obliczeniach neuronowych poprzez modyfikację wzoru (2) [3]:

$$w_i(n+1) = (1-\beta)\, w_i(n) + \alpha\, y\, u_i\,, \qquad (3)$$

gdzie β jest parametrem określającym prędkość zapominania.

Rys. 2 przedstawia proces pozbywania się odruchu warunkowego przez cykliczne podawanie tylko bodźca bezwarunkowego B. Liczba zastosowanych bodźców bezwarunkowych zależy od założonej maksymalnej wrażliwości na bodziec warunkowy.

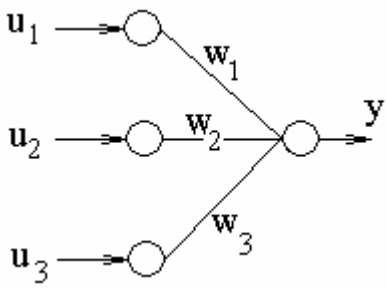


Rys. 2. Zapominanie odruchu warunkowego. Wartości wag w_1, w_2 w zależności od liczby bodźców bezwarunkowych

Wykres przedstawiony na rys. 2 odpowiada wartościom początkowym wag $w_1 = 3, w_2 = 2$ oraz następującym wartościom parametrów $\alpha = 0.22, \beta = 0.1$.

6. Odruch warunkowy drugiego rzędu

Przedstawiony tutaj uproszczony neuronowy model doświadczenia Pawłowa wyjaśnia wszystkie bardziej skomplikowane doświadczenia związane z odruchami warunkowymi. Przykładem takiego doświadczenia jest badanie odruchu warunkowego drugiego rzędu [10]. Minimalną sieć do badania tego odruchu przedstawia rys. 3.



Rys. 3. Minimalny układ do badania odruchów warunkowych drugiego rzędu

Do doświadczenia kwalifikuje się osobnika posiadającego następujące wartości wag początkowych:

$$w_1 = 1, w_2 = 0, w_3 = 0$$

Doświadczenie przebiega w następujących etapach:

Etap 1.
Podanie bodźca bezwarunkowego *bb*:

$$u_1 = 1, u_2 = 0, u_3 = 0$$

y=1,. $w_1(1) = 2, w_2(1) = 0, w_3(1) = 0$

Etap 2.
Podanie pierwszego bodźca warunkowego razem z bodźcem bezwarunkowym *bb*+*bw₁*:

$$u_1 = 1, u_2 = 1, u_3 = 0$$

y=1, $w_1(2) = 3, w_2(2) = 1, w_3(2) = 0$

Etap 3.
Podanie pierwszego bodźca warunkowego razem z drugim bodźcem warunkowym $bw_1 + bw_2$:

$$u_1 = 0, u_2 = 1, u_3 = 1\,,$$

y=1, $w_1(3) = 3, w_2(3) = 2, w_3(3) = 1$

Etap 4.
Podanie drugiego bodźca warunkowego bw_2:

$u_1 = 0, u_2 = 0, u_3 = 1\,,$

y=1, $w_1(3) = 3, w_2(3) = 2, w_3(3) = 2\,.$

Wniosek: bodziec warunkowy bw_2 wywołuje odruch warunkowy (*y*=1) mimo iż nigdy nie był podawany razem z bodźcem bezwarunkowym *bb*. Powstanie odruchu warunkowego drugiego rzędu jest wynikiem wyidukowania różnej od zera wartości wagi w_3.

7. Blokowanie odruchu warunkowego drugiego rzędu

Etap 1.
Niech „osobnik" przedstawiony na rys. 3 posiada nabyty odruch warunkowy na bodziec bw_1:

$$w_1 = 1, w_2 = 1, w_3 = 0$$

Etap 2.
Wielokrotne podanie bodźców *bb*+bw_2:

$$u_1 = 1, u_2 = 1, u_3 = 0$$

powoduje znikanie wartości wagi w_2 do zera, a tym samym zanik wrażliwości osobnika na bodziec bw_2:

$$u_1 = 0, u_2 = 0, u_3 = 1\,.$$

Etap 3.
Wielokrotne podawanie tylko bodźca *bb*:

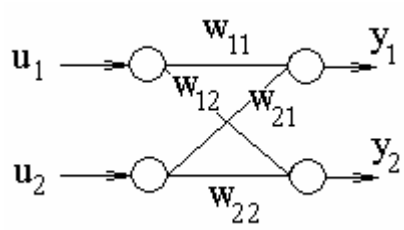
$$u_1 = 1, u_2 = 0, u_3 = 0$$

powoduje zanik wrażliwości osobnika na obydwa bodźce warunkowe bw_1 i bw_2, co jest spowodowane znikaniem obydwu wag w_1 i w_2.

8. Zespół odruchów warunkowych

Wprowadzony tutaj model odruchu warunkowego daje się uogólnić na zespół odruchów warunkowych. Najprostszy

przypadek takiego zespołu może powstać w następującym układzie:



Rys. 4. Minimalny układ do badania zespołu odruchów warunkowych

gdzie u_1 jest bodźcem bezwarunkowym (bb_1) wywołującym odruch bezwarunkowy (ob_1) oznaczonym na rys. 4 przez y_1. Analogicznie u_2 jest bodźcem bezwarunkowym (bb_2) wywołującym odruch bezwarunkowy (ob_2) oznaczonym na rys. 4 przez y_2. Przyjęte oznaczenia wynikają z założenia, że do doświadczenia zakwalifikowano osobnika, który w chwili „narodzin" miał następujące wartości wag:

$$w_{11} = 1, w_{12} = 0, w_{21} = 0, w_{22} = 1$$

Etap 1.
Doświadczenie rozpoczyna się od podania bodźca bb_1 (tab. 2, kombinacja B). Badany osobnik odpowiada odruchem ob_1 $(y_1 = 1, y_2 = 0)$ i modyfikacją wag:
$w_{11}(1) = 2, w_{12}(1) = 0, w_{21}(1) = 0, w_{22}(1) = 1.$

Etap 2.
Podaje się bodziec bb_2 (tab. 2, kombinacja C). Badany osobnik odpowiada odruchem ob_2 $(y_1 = 0, y_2 = 1)$ i modyfikacją wag:
$w_{11}(2) = 2, w_{12}(2) = 0, w_{21}(2) = 0, w_{22}(2) = 2.$

Etap 3.
Podaje się bodziec bezwarunkowy bb_1 + bb_2 (tab. 2, kombinacja D). Osobnik odpowiada odruchem y_1=1 i odruchem y_2 =1, oraz modyfikacjami wag:
$w_{11}(3) = 3, w_{12}(3) = 1, w_{21}(3) = 1, w_{22}(3) = 3.$

Etap 4.
Podaje się bodziec warunkowy bw_1 (tab. 2, kombinacja B). Osobnik odpowiada odruchem $ow_2 + ob_1$: y_1=1, y_2 =1, oraz modyfikacjami wag:
$w_{11}(4) = 4, w_{12}(4) = 2, w_{21}(4) = 1, w_{22}(4) = 3.$

Etap5.
Podaje się bodziec warunkowy bw_2 (tab. 2, kombinacja C). Osobnik odpowiada odruchem $ow_1 + ob_2$: y_1=1, y_2 =1, oraz modyfikacjami wag:
$w_{11}(5) = 4, w_{12}(5) = 2, w_{21}(5) = 2, w_{22}(5) = 4.$

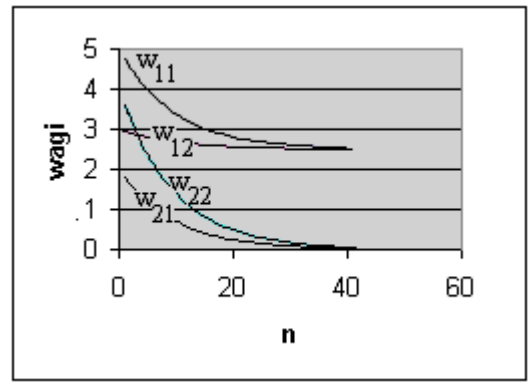
Wniosek z etapów 1-3 jest następujący:
w etapie 3 powstają zawsze równocześnie dwa odruchy warunkowe

$$(w_{12}(5) = 2, w_{21}(5) = 2)$$

Wniosek z etapów 4-5:
jeżeli w etapie 3 powstają równocześnie dwa odruchy warunkowe to podanie bodźca warunkowego powoduje równocześnie odruch bezwarunkowy i odruch warunkowy (*Etap 5*).

Ciekawe wyniki otrzymamy, jeżeli do sieci neuronowej przedstawionej na rys. 4 zastosujemy terapię mającą na celu zlikwidowanie odruchów warunkowych przez wielokrotne podawanie bodźców bezwarunkowych, analogicznie do doświadczenia opisanego w rozdziale 5. W celu likwidacji odruchu warunkowego ow_2 należy cyklicznie podawać bodziec B (tab. 2). Umownie tę procedurę nazwano dalej terapią TB. Analogicznie, w celu likwidacji odruchu warunkowego ow_1 należy cyklicznie podawać bodziec C (tab. 2). Umownie tę procedurę nazwano dalej terapią TC.

Terapia TB



Rys. 4. Zapominanie odruchu warunkowego ow_2 przez cykliczne podawanie bodźca bb_1. Wartości wag $w_{11}, w_{12}, w_{21}, w_{22}$ w zależności od liczby bodźców bezwarunkowych

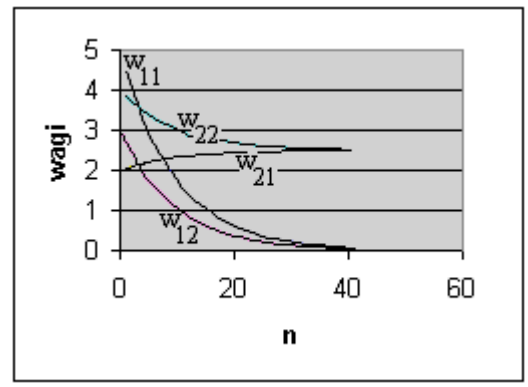
Sieć poddaną terapii TB testuje się przez podanie bodźców B, C i D.

Tab. 5. Tabela zbiorcza przedstawiająca reakcje sieci po poddaniu jej terapii TB na bodźce B, C, D

	B	C	D
y_1	1	0	1
y_2	1	0	1

Osobnik po terapii TB nie wykazuje odruchu warunkowego na bodziec C (y_1=0). Jako skutek uboczny obserwuje się brak odruchu bezwarunkowego na bodziec C (y_2=0).

Terapia TC



Rys. 5. Zapominanie odruchu warunkowego ow_1 przez cykliczne podawanie bodźca bb_2. Wartości wag $w_{11}, w_{12}, w_{21}, w_{22}$ w zależności od liczby bodźców bezwarunkowych

Sieć poddaną terapii TC testuje się przez podanie bodźców B, C i D.

Tab. 6. Tabela zbiorcza przedstawiająca reakcje sieci po poddaniu jej terapii TC na bodźce B,C,D

	B	C	D
y_1	0	1	1
y_2	0	1	1

Osobnik po terapii TC nie wykazuje odruchu warunkowego na bodziec B (y_2=0). Jako skutek uboczny obserwuje się brak odruchu bezwarunkowego na bodziec B (y_1=0).

9. Podsumowanie

Perceptronowy model doświadczenia Pawłowa pozwala na odtworzenie wszystkich jego etapów oraz interpretację wyników. Zauważmy tutaj, że powstawanie odruchu warunkowego związane jest z generowaniem różnej od zera wartości wagi odpowiadającej bodźcowi warunkowemu. Uwzględniając w regule Hebba efekt zapominania, model poprawnie odtwarza proces zapominania odruchu warunkowego przez wielokrotne podanie bodźca bezwarunkowego.

Drugi test polegający na odtworzeniu odruchu warunkowego drugiego rzędu wypada również pozytywnie. Zaproponowany model wyjaśnia mechanizm powstawania tego odruchu oraz jego zanikanie (blokowanie odruchu warunkowego drugiego rzędu).

Problem zespołu odruchów warunkowych przedstawiono jako przykład projektowania nowego eksperymentu, przewidywań i interpretacji jego wyników.

Wyniki przedstawionych tutaj doświadczeń można podsumować następująco. Eksperymentator mający do dyspozycji jedynie bodźce i nie mogący bezpośrednio oddziaływać na wagi połączeń (np. „farmakologicznie") może programować sieć neuronową na dwa sposoby:

1) Sposób polegający na generowaniu niezerowych wag przez podawanie dwóch lub więcej bodźców równocześnie. Ważne jest to, że jeden bodziec podawany jest przez połączenie aktywne z wagą różną od zera. Natomiast pozostałe bodźce podawane są przez programowane, i w danej chwili nieaktywne, połączenia z wagami o wartościach zerowych. Zgodnie z regułą Hebla, połączenia dotąd nieaktywne stają się aktywne przez podwyższenie wartości odpowiadających im wag. Zmiany wartości wag objawiają się powstaniem odpowiadających im odruchów warunkowych.

2) Sposób polegający na częstym podawaniu jednego i tego samego bodźca bezwarunkowego. Na skutek efektu zapominania obniżają się wartości wag połączeń odpowiadających bodźcom warunkowym, co skutkuje zanikaniem odruchu warunkowego.

W przypadku zespołu odruchów obserwuje się skutek uboczny polegający na zanikaniu lub osłabianiu odruchów bezwarunkowych związanych z innymi bodźcami bezwarunkowymi.

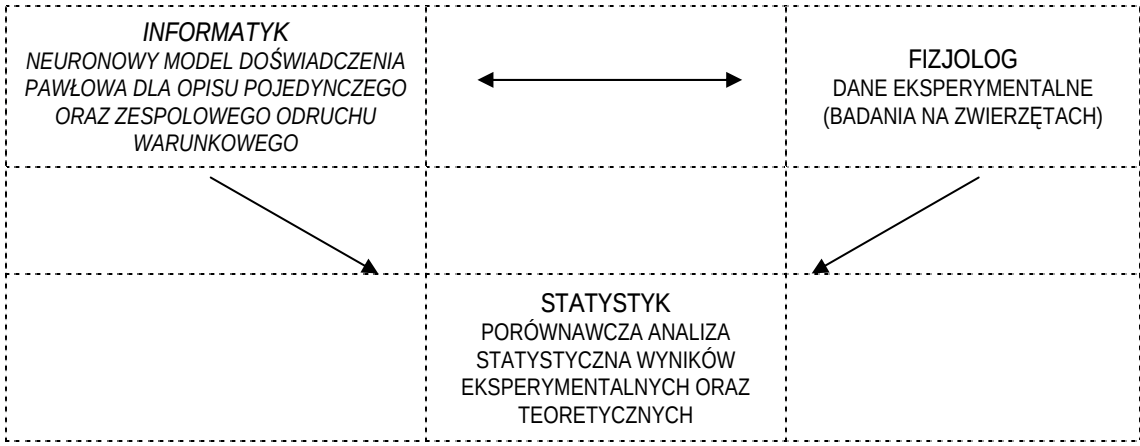
Literatura cytowana*

1. Linsker R.: Self-organization in a perceptual network, Computer, 21: 105-117, 1988.
2. Pawłow I. P.: Conditioned reflexes, Oxford University Press, London 1927.
3. Grossberg S.: Nonlinear Difference-Differential Equations in Prediction and Learning Theory, Proceedings of the National Academy of Sciences, 58: 1329-1334, USA 1967.
4. Rosenblatt F.: Principles of Neuro-dynamics, Spartan, New York 1962.
5. Rosenblatt F.: The perceptron. A probabilistic model for information storage and organization in the brain, Psych. Rev., 65: 386-408, 1958.
6. Hebb D. O.: The Organization of Behaviour, Wiley, New York 1949**.
7. Hock R. R.: Forty studies that changed psychology: Explorations into the history of psychological research, Personal Education, In: Upper Saddle River, New Jersey 2002.
8. Oja E.: A simplified neuron model as a principal component analyzer, J. Math. Biology, 15: 267-273, 1982.
9. Zimbardo P. G., Ruch F. L.: Psychology and Live, Scott, Foresman and Company, Glenview 1977.
10. Mobus G. E.: Toward a Theory of Learning and Representing Causal Inferences in Neural Networks, In: Levine D. S., Aparicio M. (Eds.): Neural Networks for Knowledge Representation and Inference, Lawrence Erlabaum Associates, New Jersey 1994***.
11. Mobus G. E.: Adapting Robot Behavior to a Nonstationary Environment: A Deeper Biologically Inspired Model of Neural Processing, Sensor Fusion and Department Control in Robotic Systems III, McKee G.T., Schenker P. S., Proc. SPIE Vol. 4196, 2000.
12. Duch W., Korbicz J., Rutkowski L., Tadeusiewicz R.: Sieci neuronowe, Biocybernetyka i inżynieria biomedyczna 2000, tom 6 (Nałęcz M., red.), Polska Akademia Nauk, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2000.

* Pozycja literaturowa [12] jest monografią polskich specjalistów z zakresu sztucznych sieci neuronowych, która zawiera podstawową wiedzę o sztucznych sieciach neuronowych zawartych w pozycjach [1], [3]-[6], [8]. Do tej pozycji nie ma odwołań w tekście pracy. Została ona zamieszczona dla wygody czytelnika. Pozycje [7] i [9] zostały wydane również w języku polskim.

** Reguła Hebba podana została rok wcześniej:
Konorski J.: Conditioned Reflexes and Neuron Organisation, CUP, Cambridge 1948.

*** http://faculty.washington.edu/gmobus/Adaptrode/causal_representation.html


INFORMATYK
NEURONOWY MODEL DOŚWIADCZENIA PAWŁOWA DLA OPISU POJEDYNCZEGO ORAZ ZESPOLOWEGO ODRUCHU WARUNKOWEGO
FIZJOLOG
DANE EKSPERYMENTALNE (BADANIA NA ZWIERZĘTACH)
STATYSTYK
PORÓWNAWCZA ANALIZA STATYSTYCZNA WYNIKÓW EKSPERYMENTALNYCH ORAZ TEORETYCZNYCH

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 259-266.

BISTABILITY IN EPILEPTIC PHENOMENA

BISTABILNOŚĆ W UKŁADACH EPILEPTYCZNYCH

PIOTR SUFFCZYŃSKI[*,**], STILIYAN KALITZIN[**], FERNANDO LOPES DA SILVA[**,***]

[*]*Zakład Fizyki Biomedycznej, Instytut Fizyki Doświadczalnej, Uniwersytet Warszawski, Warszawa, suffa@fuw.edu.pl*
[**]*Stichting Epilepsje Instellingen Nederland, Heemstede, Holandia, skalitzin@sein.nl*
[***]*Section Neurobiology, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, silva@science.uva.nl*

Abstract. It is currently believed that the mechanisms underlying spindle oscillations are related to those that generate spike and wave (SW) discharges. The mechanisms of transition between these two types of activity, however, are not well understood. In order to provide more insight into the dynamics of the neuronal networks leading to seizure generation in a rat experimental model of absence epilepsy we developed a computational model of thalamo-cortical circuits based on relevant (patho)physiological data. The model is constructed at the macroscopic level since this approach allows to investigate dynamical properties of the system and the role played by different mechanisms in the process of seizure generation, both at short and long time scales. The main results are the following: (i) SW discharges represent dynamical bifurcations that occur in a bistable neuronal network. (ii) The durations of paroxysmal and normal epochs have exponential distributions, indicating that transitions between these two stable states occur randomly over time with constant probabilities. (iii) The probabilistic nature of the onset of paroxysmal activity implies that it is not possible to predict its occurrence. (iv) The bistable nature of the dynamical system allows an ictal state to be aborted by a single counter-stimulus.

Key words: absencje seizures, epilepsy, bistability

Streszczenie. Mechanizm spontanicznego powstawania oraz wygaszania napadów epileptycznych nie jest dobrze poznany. W celu lepszego zrozumienia tych mechanizmów opracowany został model symulacyjny układu wzgórzowo-korowego, odpowiedzialnego za powstawanie napadów nieświadomości u ludzi oraz zwierząt. Model wykorzystuje podejście populacyjne, co pozwoliło na zbadanie właściwości układu w skali krótko- i długoczasowej. Główne wyniki pracy to: (i) napady nieświadomości powstają w bistabilnej sieci neuronalnej, (ii) długości napadów oraz odcinków pomiędzy napadami mają rozkład eksponencjalny, co sugeruje, że przejścia pomiędzy stanem normalnym a napadem odbywają się losowo w czasie ze stałym prawdopodobieństwem (iii) probabilistyczna natura powstawania napadów sugeruje, że nie jest możliwe ich przewidywanie (iv) napady powstające w układzie bistabilnym mogą być zatrzymane poprzez stymulację pojedynczym impulsem elektrycznym.

Słowa kluczowe: napady nieświadomości, epilepsja, bistabilność

1. Introduction

Absence seizures are paroxysmal losses of consciousness that start and end abruptly and are accompanied by bilaterally synchronous spike and wave (SW) discharges that can be recorded on the electroencephalogram (EEG). The cause of this abnormal behavior is presumably genetic although genetic abnormalities responsible for absence phenotype in humans are not yet clear [4]. The development of animal models of absence seizures and, more recently, computational models of circuits involved in SW generation, advanced our understanding of the basic neuronal mechanisms of this type of epilepsy. It is currently believed that the mechanisms underlying SW discharge may be related to the thalamo-cortical mechanisms of sleep spindle generation ([16]). However the mechanisms that are responsible for the *spontaneous transition* between the normal spindle oscillations and paroxysmal SW activity are not well understood. Similarly, the questions concerning the factors that determine seizure duration and/or inter-seizure interval have not often been addressed either. Experimental studies in animal models have yielded new findings (reviewed in [6]) but these are still fragmentary so that a comprehensive interpretation of the role played by different mechanisms in the process of seizure generation, both at short and long time scales, is still limited. We assumed that a computational model that takes into account the most relevant (patho) physiological experimental findings would be an appropriate tool that could contribute to reach such a comprehensive view. The essence of epilepsy is that a patient displays (long) periods of normal EEG activity (i.e. non-epileptiform) intermingled with epileptiform paroxysmal activity only occasionally. Thus, the main aim of the present study is to find out the mechanisms responsible for

spontaneous transitions from normal activity to paroxysmal SW discharges. This aspect of the dynamical process responsible for epilepsy has, to the best of our knowledge, not been addressed in a computational model of absence epilepsy and therefore the present model provides a novel contribution.

In this study we approach the given problem at the intermediate level between the distributed neuronal network and lumped circuit levels. That is, we do not simulate the explicit behavior of individual neurons but rather model the populations of interacting neurons integrating neuronal and network properties.

In the present study we focus on the SW activity that is commonly recorded from a genetic animal model of absence epilepsy, namely the Wistar albino Glaxo from Rijswijk (WAG/Rij) rat [21], that has similar characteristics to the generalized absence epilepsy rat from Strasbourg (GAERS) model [15]. The main reason to choose these animal data for constructing the computational model is that rat experimental data at the cellular and network levels are available, in contrast to the human case.

2. Materials and Methods

Model's structure

The present model is an extended version of the lumped model initially proposed by Lopes da Silva et al. [13]. The latter model consisted of a single module, i.e., a cluster of neurons containing two interacting populations. Excitatory cells of main population projected to the interneurons while the latter fed back on the main cells, inhibiting them with fast $GABA_A$ receptor mediated inhibitory postsynaptic potentials (IPSPs). Each population was described by the time courses of postsynaptic potentials and sigmoid transfer function, which was used to simulate the conversion between the mean membrane potential of a neuronal population and the firing rate (number of pulses per second (pps)) of the population. We made a number of extensions and added new features to the above described model. The present model is advanced with respect to its precursor in three main points. (i) It consists of two modules, cortical and thalamic ones, that are mutually interconnected. (ii) The transformation between mean membrane potential and firing density in the thalamic populations takes into account the low-threshold I_T calcium current that underlies burst firing in the thalamic cells. (iii) In addition both fast $GABA_A$ and slow $GABA_B$ receptor mediated inhibitions are included and $GABA_B$ receptors are assumed to have nonlinear activation properties. The schematic diagram of the model is shown in Fig. 1.

The thalamic loop consists of a population of thalamocortical (TC) cells that projects to a population of reticular thalamic (RE) cells. The latter inhibits TC population by way of $GABA_A$ and $GABA_B$ types of inhibition. The TC cells receive external excitatory input that represents sensory inputs from the ascending afferents while the RE population receives external inhibitory input. The latter represents the input from the neighboring RE cells, since these are interconnected by mutual inhibitory synapses. The cortical module consists of a negative feedback loop formed by interacting populations of pyramidal cells (PY) and inhibitory interneurons (IN). Pyramidal cells, in addition to projecting to interneurons, send also excitatory connections to the thalamus both to the TC and RE populations. In turn, the TC cells excite both the pyramidal cells and interneurons. The PY population receives also external cortical input that stands for the glutamatergic input from other pyramidal cells, not included in the lump. All excitatory synaptic interactions are mediated by glutamate α-amino-3-hydroxy-5-methyl-4-isoxa-zolepropionic acid (AMPA) receptors. The output of the model, is the mean membrane potential of the pyramidal cell population. We assume that the dynamics of this activity (e.g., dominant frequency) is reflected on the local field potential recordings and EEG signals that have the same dynamical properties. However, in the present model we did not aim to simulate realistic field potential waveforms. A detailed model description is given in [19].

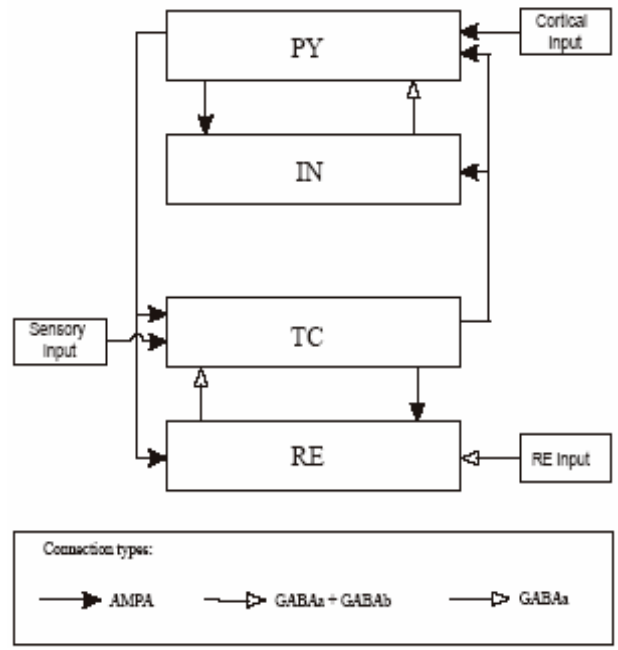


Fig. 1. Schematic structure of connections in the thalamo-cortical network model consisting of cortical and thalamic modules. The cortical module consists of two interconnected populations of PY and IN neurons. The thalamic module consists of two interconnected populations of the TC and RE neurons. TC cells project to both the PY and IN cells, while PY cells project to both the TC and RE cells. The PY population receives external cortical excitatory input, the TC population receives external sensory input and the RE population receives external inhibitory input. Figure's legend shows synaptic connection types.

3. Results

Different types of model behavior

The model may exhibit two qualitatively different types of behavior, such as seen in the experimental animals (WAG/Rij and GAERS). The output signal may display a waxing and waning "spindle-like" oscillation having a spectrum with a peak at approximately 11 Hz or a high amplitude "seizure-like" oscillation at a frequency around 9 Hz. We refer to the former behavior as "normal on-going" activity while to the latter as paroxysmal activity.

For the reference set of parameters, the model is in a bistable regime where it may generate both normal and paroxysmal oscillations and spontaneous transitions between

these two types of behavior. The model output in this regime is shown in Fig. 2, upper panel. The lower panel depicts the corresponding power spectra.

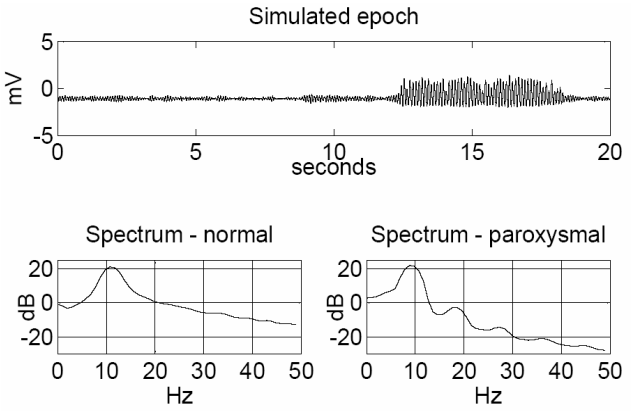


Fig. 2. Example of model output. Upper panel: 20 seconds of a simulation with the occurrence of a spontaneous paroxysmal episode. Lower panel: power spectra of simulated normal and paroxysmal activity. Dominant frequency of normal activity is around 11 Hz while that of paroxysmal activity is around 9 Hz.

Spindle activity in the model is generated by the cyclical interaction between the TC and RE populations. Inhibitory postsynaptic potentials of RE origin facilitate the generation of rebound low threshold spikes in TC population, which in turn activate GABAergic RE neurons. The dominant frequency (~ 11 Hz) of this rhythmic activity is largely determined by the time courses of both AMPA and $GABA_A$ postsynaptic currents and membrane time constants and to a lesser degree by other factors like the characteristics of low threshold spikes (LTS), synaptic coupling constants and sensory input level (see Appendix). Thalamic generated spindle activity induces activity at the same frequency in the cortical populations.

During normal activity, the thresholds for the activation of $GABA_B$ receptors in the TC and PY cells are not reached. When one of these thresholds is surpassed, $GABA_B$ receptor mediated IPSPs may be triggered, which results in the sudden transition of the model's behavior from "normal on-going" activity to "seizure-like" 9 Hz large amplitude oscillations. The frequency and amplitude of paroxysmal oscillations depend on the relative contribution of $GABA_A$ and $GABA_B$ components. A reduction of the contribution of $GABA_A$ and an increase of that of $GABA_B$ slows down the paroxysmal oscillations and increases their amplitude. A complete block of $GABA_A$ receptors in the TC and PY cells (while keeping the $GABA_B$ conductance at the reference level) slows the oscillation down to 4 Hz.

The dynamical mechanisms of spontaneous transitions between normal and paroxysmal behavior are described below.

Bifurcation analysis

First, the model was run with constant input signals, i.e., with the variance of noise of cortical and sensory inputs (see Appendix) equal to zero. Under such conditions the model exhibited either a constant output, or a periodic 9 Hz oscillation, depending on the initial conditions, but it did not exhibit transitions from one state to another. It indicates that transitions in the model behavior, such as shown in Fig. 2, cannot take place when input signals are constant. It suggests that random fluctuations that are introduced into the system by noise in cortical and sensory inputs are necessary for the transitions to occur. This became clearer while analysing bifurcation diagrams.

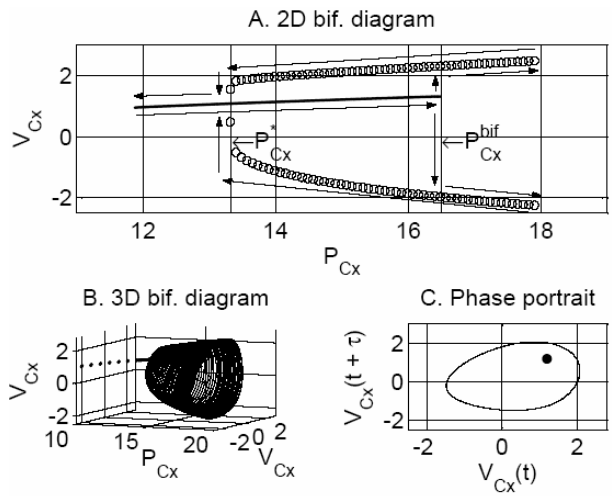


Fig. 3. Bifurcation diagram.

A. Two dimensional bifurcation diagram showing amplitude of the mean membrane potential of the pyramidal cells population (V_{Cx} in mV) along the y-axis as a function of step-wise increasing and decreasing constant (noise free) cortical input (P_{Cx} in pulses per second (pps)) along the x-axis. In the simulation performed to draw the bifurcation diagram the noise component was removed from the inputs. Arrows along the bifurcation diagram depict changes of V_{Cx} as P_{Cx} increases and decreases. To read the bifurcation diagram, fix the P_{Cx} at a particular value and mentally draw a vertical line at that value. Each crossing of this line with a curve in the diagram corresponds to a point attractor (thick line) or a limit cycle (open circles denote maxima and minima of an oscillation). Point attractor corresponds to normal activity while limit cycle corresponds to paroxysmal activity. For increasing DC values of P_{Cx} the transition from point attractor to limit cycle occurs at point P_{Cx}^{bif}. For $P_{Cx} > P_{Cx}^{bif}$ only paroxysmal behavior is possible. For decreasing values of P_{Cx} the transition from limit cycle to point attractor occurs at point P_{Cx}^*. For $P_{Cx} < P_{Cx}^*$ only normal behavior is possible. For $P_{Cx}^* < P_{Cx} < P_{Cx}^{bif}$ the system possesses bistable dynamics: the normal state coexists with the paroxysmal oscillatory state.

B. Three dimensional bifurcation diagram of the system constructed using delay embedding with cortical input P_{Cx} along the x-axis, model's output V_{Cx} along the y-axis and delayed (τ = 40 ms) version of model's output V_{Cx} along the z-axis.

C. Phase portrait of the system, obtained by taking a cross section of figure in B for P_{Cx} = 15 pps, showing coexistence of a point attractor (black dot) with a limit cycle (thin line).

In Fig. 3A we show the two-dimensional bifurcation diagram for the reference set of parameters, as a function of constant (noise free) cortical input P_{Cx}. Its three-dimensional version is presented in Fig. 3B. In the bifurcation analysis the noise component was removed from the inputs to keep the illustration simple. The bifurcation diagram was constructed by first increasing and subsequently decreasing the values of P_{Cx}, the arrows along the diagram follow changes of system's output as the cortical input increases and decreases. The changes in P_{Cx} were made step-wise and allowed the system to reach its stationary behavior. Therefore the diagram depicts

dynamics of the system corresponding to different (fixed) values of the control parameter. Fig. 3A reveals that there are two types of activity possible: an equilibrium point corresponding to "normal on-going" activity (solid line) and a limit cycle corresponding to paroxysmal oscillations (open circles mark its minimal and maximal amplitude). Both equilibrium point and a limit cycle are emergent property of our model for the reference set of parameters that were chosen according to available physiological data. Fig. 3A. reveals further that the transition between the "normal on-going" and the paroxysmal activity modes occurs when the cortical input P_{Cx} increases beyond the bifurcation point at P_{Cx}^{bif}. This is a typical bifurcation of the subcritical Hopf kind (Glass and Mackey, 1988). As the cortical input is subsequently decreased there is a jump from large amplitude limit cycle oscillations back to equilibrium state but this takes place at a value of P_{Cx}^{*} which is less than P_{Cx}^{bif}. This is manifestation of hysteresis. For $P_{Cx}^{*} < P_{Cx} < P_{Cx}^{bif}$, there are two coexisting states: normal and paroxysmal. These two dynamical states are locally stable in the sense that under small perturbations of finite-time duration the system recovers to its original state. Therefore we can call such system a bistable system. The coexistence of two attractors can be seen also in the phase space plot in Fig. 3C, which is a cross-section of the bifurcation diagram shown in Fig. 3B.

In the absence of noise the normal and paroxysmal attractors correspond to equilibrium point and limit cycle, respectively. When noise is introduced to the system (both cortical and sensory inputs are superpositions of two components: a constant that constitutes the DC offset and a Gaussian noise components representing random input fluctuations), the perturbations of the steady state give rise to small amplitude rhythmic (spindle-like) activity, while perturbations of the limit cycle result in amplitude modulation of the SW oscillations (both rhythms can be seen in Fig. 2). Spontaneous (i.e., noise induced) transitions between normal and paroxysmal state occur as follows. Let us assume that the system, initially in the normal stable state, is in the bistable domain, and that all parameters, including the DC levels of the inputs, are constant. If the (randomly fluctuating) value of P_{Cx} becomes greater than P_{Cx}^{bif}, the system undergoes a transition to the paroxysmal state and remains there (because this is a stable state too) until the value of P_{Cx} becomes smaller than P_{Cx}^{*}, bringing the system back to its initial normal state.

The dynamical scenario of transitions between normal and paroxysmal states was described above using the bifurcation diagram in Fig. 3, which was made for only one control parameter (the value of the constant component of the cortical input), and for only one system's variable (model output). A more complete bifurcation analysis that would take into account other parameters and system's variables and simultaneous variation of two or more control parameters lies beyond the scope of this paper. However, the following general conclusion can be formulated. The occurrence of paroxysmal activity in the model is governed by a random process, i.e., both the onset and cessation of paroxysms occur randomly over time with certain probabilities. Accordingly, the distribution of the duration of paroxysmal and normal epochs can be predicted to be exponential [7]. Therefore the construction of histograms of the durations of paroxysmal and normal epochs from simulated and experimental time series offers a way to verify these theoretical predictions regarding the dynamics of the process leading to spontaneous paroxysmal discharges. An exponential distribution of lengths of paroxysmal and normal epochs would favor the scenario presented above; different kinds of distributions may indicate that other mechanisms have to be taken into consideration.

Distributions of paroxysmal and normal epochs durations

First, we constructed histograms of paroxysmal and normal epochs durations from the simulated signals. The detection of paroxysmal epochs was based on two criteria: the level of activation of $GABA_B$ receptor in the TC cells and the standard deviation of the output signal. The choice of the former criterion was based on the well documented fact that activation of $GABA_B$ inhibition in thalamic relay nuclei is essential for paroxysmal discharges in animal models of absence epilepsy [10], [12]. In our model, the beginning of a paroxysmal epoch was detected when $GABA_B$ conductance in the TC cells was greater than 3% of the maximal conductance value and the end of a paroxysmal epoch was detected when $GABA_B$ conductance in the TC cells fell below 0.5% of the maximal conductance value. This choice of the detection thresholds gave very good results in comparison to visual inspection of the output signals. To improve detection performance further, we introduced a second criterion that took into account the standard deviation of the model's output amplitude. The threshold for detection of paroxysmal activity was set at three times the standard deviation of the normal, background activity. The latter was calculated on the basis of a long (100 seconds), paroxysmal free, signal. The epoch was considered a paroxysm when both criteria were fulfilled. The histograms of durations of paroxysmal and normal epochs detected during 24 hours of simulated time, for a reference set of parameters, are shown in Fig. 4, upper panel.

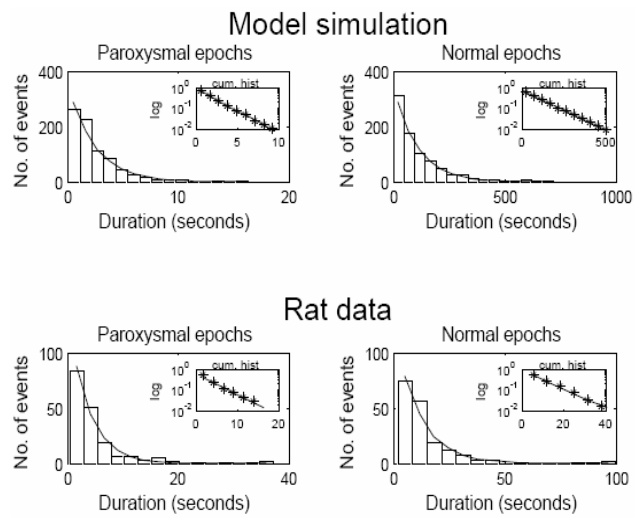


Fig. 4. Distributions of lengths of normal and paroxysmal epochs of the simulated signal (upper panel) and of real EEG signal recorded from the rat (lower panel). On each plot an exponential function fitted to the histogram is shown. Model histograms were obtained by simulating 24 hours of activity using the reference parameters set. Histograms of rat data were obtained from 30 min. recording of EEGs of WAG/Rij rat after administration of high dose of vigabatrin. In all four distributions shown, the divergence between the observed

histogram and fitted function is small and we were not able to reject the null hypothesis that the observed histogram scores came from the fitted exponential distribution even when tested at high significance level (α = 0.2). Insets in each window depict cumulative histograms shown on logarithmic scale and straight lines fitted to histogram points. In all four plots the lines fit well the histograms' points further confirming exponential distributions.

The histograms were fitted with exponential functions. Second, we computed equivalent distributions of experimental data. These distributions were derived form data of WAG/Rij rat six hours after administration of vigabatrin (500 mg/kg), which significantly enhances (~ five times) the incidence of paroxysms in these rats and therefore enables to obtain more data for analysis in a relatively short time. Data of durations of normal and paroxysmal epochs in WAG/Rij rats were obtained by an independent experimenter and details of data acquisition are given in [3]. The experimental histograms and fitted exponential functions are shown in Fig. 4, lower panel.

The fitted exponential functions appear to describe well both the simulated and experimental distributions. In all four examples shown, the Kolmogorov – Smirnov goodness of fit test failed to reject the null hypothesis (H_0) that the observed histogram scores came from the fitted exponential distribution even when setting a high probability, α = 0.2, of falsely rejecting H_0 (critical values for significance levels greater than 0.2 were not tabulated). Additional confirmation of exponential law comes from the cumulative histograms shown in inset of each panel. In all four examples straight lines seem to fit well cumulative histograms on the logarithmic scale.

Counter-stimulation

There is a general interest to develop methods able to control the occurrence or the evolution of seizures. Thus we investigated, using the model, how a paroxysmal oscillation could be aborted. Since the model's dynamics displays bistability where a limit cycle co-exists with a steady state (Fig. 3C), it should be possible to terminate a paroxysmal oscillation by an appropriate stimulus [9]. We verified this theoretical prediction as follows. A paroxysmal oscillation was triggered by a pulse of 40 pps (pulses per second) amplitude, 10 ms duration delivered to the cortical input around second 1.5. If no stimulation was applied afterwards, the paroxysmal discharge continued as time progressed (Fig. 5A). If a counter-stimulation pulse with the same parameters as the trigger but shifted forward in time by 62 ms was delivered to the cortical input, the paroxysmal oscillation may be aborted (Fig. 5B).

The paroxysmal oscillation, however, was annihilated only if the stimulus was delivered at a specific phase of the oscillation. Close inspection of the output signal showed that a depolarizing pulse was effective if it was delivered near the time at which the output signal reached maximal negative polarity (Fig. 5B, inset). Counter-stimulation performed equally well on spontaneous paroxysms as on those triggered by an external stimulus. Phase – amplitude analysis of the effectiveness of the counter-stimulus is shown in Fig. 5C. It shows that the abnormal rhythm can be stopped either by a positive stimulus applied on the descending slope of the PY population amplitude or by a negative stimulus applied on the rising slope of the PY population signal.

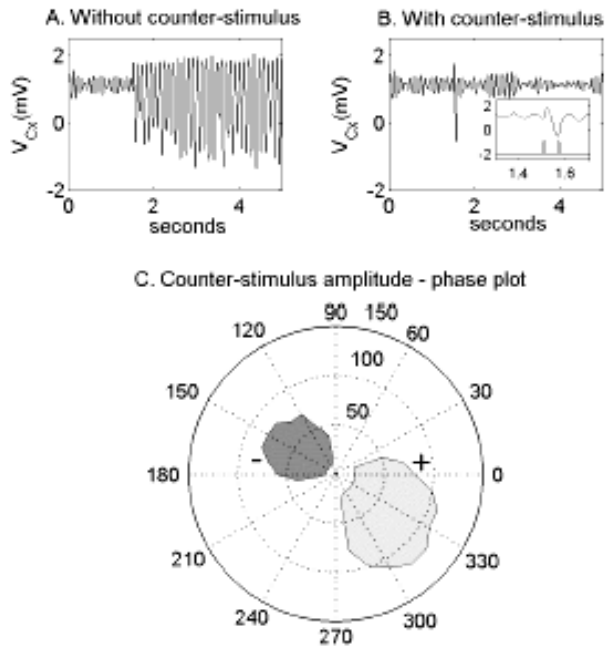


Fig. 5. Counter-stimulation phenomenon

A. Output epoch showing paroxysmal oscillations initiated by the triggering stimulus (40 pps (pulses per second) amplitude, 10 ms duration) delivered to the cortical input around second 1.5. Paroxysmal activity is prolonged when there is no counter-stimulation pulse applied afterwards.

B. Output epoch showing paroxysmal oscillations initiated by the triggering stimulus as in A. The triggering stimulus is followed by a counter-stimulation pulse (40 pps amplitude, 10 ms duration) delivered to the cortical input 62 ms after the trigger. In this case only one period of simulated paroxysmal discharge is present and the paroxysm is aborted. The inset displays the output signal and both triggering and counter-stimulation pulses in an extended time scale. It shows that the stimulus effective in annihilating the paroxysmal rhythm was delivered when the PY population was maximally inhibited.

C. Amplitude – phase analysis of the effective counter-stimulus. The plot in polar co-ordinates shows two patches, one (light gray, marked with a '+' sign) corresponding to positive pulses, the other (dark gray, marked with a '-' sign) corresponding to negative pulses. Each patch shows the counter-stimulation parameters (amplitude and phase) that were effective in aborting paroxysmal oscillations. Zero phase corresponds to maximal negative polarity of the output signal. Patch plots were obtained by model simulations with noise components removed and paroxysmal oscillations triggered by an external stimulus as in A, B. The counter-stimulus had 10 ms duration, its phase was sampled form 0 to 360 deg in steps of 10 deg (assuming cycle frequency 9.28 Hz) while its amplitude was sampled from 0 to 150 pps in steps of 10 pps. The counter-stimulation was delivered after 9 cycles of the paroxysm (~ 1 second) to avoid transient effects influencing the analysis.

4. Discussion

Basic properties of the model

We constructed a model of interconnected thalamic and cortical populations with parameters directly related to synaptic and cellular properties in order to investigate the mechanisms of generation of synchronized oscillations in a thalamo-cortical neuronal network. We used animal data obtained in rat genetic models of absence epilepsy (WAG/Rij and GAERS) as the basic experimental data to be modeled.

The simulation studies of the dynamical behavior of the model and of its sensitivity to a number of basic parameters led to the following main results:

(i) Paroxysmal discharges characterised by 9 Hz large amplitude oscillations represent bifurcations that occur in a neuronal network with bistability properties. This means that two stable states co-exist, one corresponding to the "normal on-going" EEG activity and the other to the paroxysmal oscillations, and that the system may undergo transitions from one state to another. These transitions can emerge spontaneously and are not induced necessarily by any parameter changes.

(ii) The distributions of lengths of paroxysmal and "normal on-going" epochs are exponential, indicating that transitions between these two stable states occur randomly over time and that the probabilities for the transition between both states can be defined.

(iii) Probabilities of transitions between "normal on-going" neuronal activity and paroxysmal oscillations depend on a number of model parameters; therefore these probabilities can be controlled in a not unique way.

(iv) Paroxysmal oscillations can be annihilated by a well-timed pulse.

(v) Since random fluctuations in control parameters and/or dynamic variables can lead to the sudden onset of large amplitude paroxysmal activity, this implies that the occurrence of this type of phenomena is unpredictable per definition.

Time distributions of paroxysmal events

Here we discuss how our model results compare with experimental findings on rat models of absence epilepsy. A first observation is that paroxysmal activity in rats arises abruptly with a change of signal's amplitude to large values [21], [18]. This sudden appearance of paroxysmal activity is consistent with the hypothesis that underlying neuronal populations behave as bistable systems. In monostable systems oscillations usually build up since gradual parameter changes are required to reach the paroxysmal oscillatory state. The evidence for bistability in both GAERS and WAG/Rij rats exists, since SW discharges are immediately interrupted by strong sensory (e.g., acoustic) stimulation [8]. The role of random fluctuations in triggering the transitions between normal and paroxysmal state in WAG/Rij rats is demonstrated here by the histograms of durations of both ictal and inter-ictal periods of activity (Fig. 4, lower panel). The exponential shape of both distributions indicates that in these animals seizure initiation and termination are random processes with constant probabilities over time. However, we should note that the experimental rat data analysed here were obtained under vigabatrin, a condition where there is an increase of the occurrence of SW discharges, an increase of their mean duration and a decrease of their peak-frequency as compared to saline conditions [3]. These experimental distributions provide 'proof of existence' of the bistability phenomenon predicted in our computer model studies. Another confirmation of our predictions comes from analysis of recordings of GAERS in drug-free conditions [20]. In some rats the distributions of ictal and seizure-free epochs are exponential while in other they can be described by gamma distributions (generalization of exponential distributions) which are also typical for random transitions in bistable systems in more general conditions. Therefore, we do not exclude that other physiological mechanisms such as progressive and persistent activation of the hyperpolarization-activated depolarizing (I_h) current in TC cells [2], or a change in the ionic composition of the extra-cellular space [17], [1], may play also a role in the termination or initiation of a paroxysmal burst.

In addition to the fact that the present model explains the dynamics of spontaneous appearance of SW seizures, the model study allows also the identification of a neuronal mechanism underlying these dynamical properties. Random fluctuations may arise due to a number of reasons including quantum noise in opening and closing of ionic membrane channels, axons membrane voltage fluctuations and the probabilistic nature of synaptic transmission ([14]). Bistability in our thalamo-cortical network model results from the specific properties of $GABA_B$ receptors and I_T current in TC neurons. $GABA_B$ receptors, having nonlinear activation properties, require a strong stimulus intensity to evoke a response. Accordingly, in the normal state $GABA_B$ receptor mediated inhibition is not active. Activation of $GABA_B$ receptors in TC cells results in prolonged hyperpolarization which is effective in removing the inactivation of the I_T current in these cells leading to increased burst firing of TC population. Large TC bursts strongly activate RE cells – directly and indirectly (through PY cells) which results in subsequent activation of $GABA_B$ receptors and this cycle tends to repeat itself. Failure to reinitiate the cycle results in seizure cessation.

Can the model account for human absences?

The model was developed to account for absence rat animal data but it is interesting to ask the question whether the model is likely to be relevant to human absence seizures. Human absences and rat absence-like non-convulsive seizures share many common features like presumed genetic origin, behavioral and EEG patterns and pharmacological profile, but there are also some differences. In typical human absence epilepsy seizures start in childhood and usually disappear in adulthood. In epileptic rats seizures appear in adult animals and continue to occur through the entire life. Another difference is that the SW frequency among humans is between 2.5 – 4 Hz while among absence epilepsy rats the frequency varies form 7 to 11 Hz. This difference in frequency can be easily accounted for by minor parameter changes in the model: e.g., a reduction of $GABA_A$ conductance and an increase of $GABA_B$ conductance in PY and TC cells and a small increase of membrane time constant in all types of cells can transform the model's output into paroxysmal episodes of frequency 4 Hz arising also from small amplitude, waxing and waning background of dominant frequency about 9 – 10 Hz. These results are similar to those of Destexhe [5] showing that thalamo-cortical circuit can generate both fast (5 – 10 Hz) and slow (~ 3 Hz) SW oscillations depending on balance between GABAergic conductances in thalamic relay cells. Since the long term properties of our model apply also for the parameter setting with the characteristics of human SW frequencies, we may predict that the distributions of absence seizure durations and inter-ictal intervals in humans should be exponential as they are in rats. Interestingly, exponential shape of histograms of durations of ictal events has been

found in human recordings from absence patients [20] and in other types of human epilepsy [22].

Can absence seizures be predicted?

Following the hypothesis that absence-like seizures in rats and human absences correspond to the same basic mechanisms, an interesting question arises whether the model can help to understand whether the occurrence of absence seizures may be predicted some time in advance or not. The fact that we found that random fluctuations can be responsible for the sudden onset of the paroxysmal oscillations implies that the occurrence of these seizures cannot be predicted since fluctuations are unpredictable by definition. This conclusion is consistent with classic clinical observations in humans as expressed by Lennox and Lennox [11]: 'If warning occurs, the diagnosis of petit mal may be questioned'. While the possibility of prediction of the exact timing of absence seizures is rejected by our model, quantifying the probability for an instantaneous seizure transition, e.g., by indirectly measuring some control parameters, might still be a realistic option.

Can paroxysmal oscillations be aborted?

In a bistable systems where a stable steady state coexists with a stable limit cycle the transitions between the two can be triggered by a single pulse. In addition, the oscillation may be terminated if a brief stimulus of sufficient magnitude is delivered at a critical phase of the oscillation [9]. The results of the present model confirm this theoretical prediction as shown in Fig. 6A, B. A detailed analysis of the effective pulse parameters is shown in Fig. 6C. It shows that paroxysms can be aborted either by an excitatory pulse delivered before the time at which the main pyramidal cells population reaches maximal negative polarity or by an inhibitory pulse delivered before the time at which the main population is maximally depolarized. Experimental application of these results may be feasible, although in biological networks spatially distributed degrees of freedom must be taken additionally into account.

In conclusion, we can argue that our model, inspired by neurophysiological observations while constructed with low computational complexity, has allowed us to test a number of hypotheses and the formulation of new ones concerning mechanisms responsible for the spontaneous occurrence of paroxysmal episodes in rat models of absence seizures and their main dynamical properties.

Acknowledgements

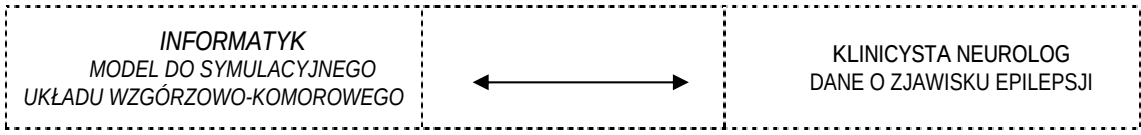
Piotr Suffczyński was partly sponsored by grant 8 T11E 003 17 from Polish Committee for Scientific Research and by Lopes da Silva Fellowship from the Christelijke Vereniging voor de Verpleging van Lijders aan Epilepsie (Christian Society for the Management of Patients with Epilepsy), Heemstede.

References

1. Amzica F., Massimini M., Manfridi A.: Spatial buffering during slow and paroxysmal sleep oscillations in cortical networks of glial cells in vivo. J. Neurosci. 22(3): 1042-1052, 2002.
2. Bal T. and McCormick D. A.: What stops synchronized thalamocortical oscillations? Neuron 17: 297-308, 1996.
3. Bouwman B. M., van den Broek, P. L. C., van Luijtelaar E. L. J. M., van Rijn C. M.: The effects of vigabatrin on type II spike wave discharges in rats. Neurosci. Lett. 338: 177-180, 2003.
4. Crunelli V., Leresche N.: Childhood absence epilepsy: genes, channels, neurons and networks. Nature Rev. Neurosci. 3: 371-382, 2002.
5. Destexhe A.: Can $GABA_A$ conductances explain the fast oscillation frequency of absence seizures in rodents? Eur. J. Neurosci. 11: 2175-2181, 1999.
6. Destexhe A., Sejnowski T. J.: Interactions Between Membrane Conductances Underlying Thalamocortical Slow-Wave Oscillations. Physiol. Rev. 83: 1401-1453, 2003.
7. Doob J. L.: Stochastic Processes. John Wiley & Sons, New York 1953.
8. Drinkenburg W. H. I. M., Coenen A. M. L., Vossen J. M. H., van Luijtelaar E. L. J. M.: Spike-wave discharges and sleep-wake states in rats with absence epilepsy. Epilepsy Res. 9: 218-224, 1991.
9. Glass L., Mackey M. C.: From Clocks to Chaos: The Rhythms of Life. Princeton University Press, Princeton 1988.
10. Hosford D. A., Clark S., Cao Z., Wilson Jr. W. A., Lin F. H., Morriset R. A., Huin A.: The role of $GABA_B$ receptor activation in absence seizures of lethargic (lh\lh) mice. Science 257: 398-401, 1992.
11. Lennox W.G., Lennox M. A.: Epilepsy and related disorders. Vol.1. Little, Brown and Co., Boston 1960.
12. Liu Z., Vergnes M., Depaulis A., Marescaux C.: Involvement of intrathalamic $GABA_B$ neurotransmission in the control of absence seizures in the rat. Neuroscience 48: 87-93, 1992.
13. Lopes da Silva F. H., Hoeks A., Smits H., Zetterberg L. H.: Model of Brain Rhythmic Activity. The Alpha-Rhythm of the Thalamus. Kybernetic 15: 27-37, 1974.
14. Lopes da Silva F. H., Pijn J. P., Velis D., Nijssen P. C. G.: Alpha rhythms: noise, dynamics and models. Int. J. Psychophysiol. 26: 237-249, 1997.
15. Marescaux C., Vergnes M., Depaulis A.: Genetic absence epilepsy rat from Strasbourg – A review. J. Neural. Transm. (Suppl.) 35: 37-69, 1992.
16. McCormick D. A., Contreras D.: On the cellular and network bases of epileptic seizures. Annu. Rev. Physiol. 63: 815-846, 2001.
17. Pumain R., Heinemann U.: Stimulus- and amino acid-induced calcium and potassium changes in rat neocortex. J. Neurophysiol. 53(1): 1-16, 1985.
18. Snead O. C., Depaulis A., Vergnes M., Marescaux Ch.: Jasper 's Basic Mechanisms of the Epilepsies. In: Delgado-Escueta, A.V., Wilson, W.A., Olsen, R.W. and Porter, R.J. (Eds.) Advances in Neurology, Vol. 79: 253-279, Lippincott Williams &Wilkins, Philadelphia, 1999.
19. Suffczyński P., Kalitzin S., Lopes da Silva F. H.: Dynamics of non – convulsive epileptic phenomena modeled by a bistable neuronal network. Neuroscience, 126(2): 467-484, 2004.
20. Suffczyński P., Lopes da Silva F., Parra J., Velis D., Bouwman B., van Hese P., Carlen P., Kalitzin S.: Dynamics of epileptic phenomena determined from statistics of ictal transitions. Submitted to IEEE TBME, 2005.
21. Van Luijtelaar E. L. J. M., Coenen A. M. L.: Two types of electrocortical paroxysms in an inbred strain of rats. Neurosci. Lett. 70: 393-397, 1986.

22. Velazquez J. L. P., Khosravani H., Lozano A., Bardakijan B. L., Carlen P. L., Wennberg R.: Type III intermittency in human partial epilepsy. Eur. J. Neurosci. 11(7): 2571-2576, 1999.

.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 267-272.

ADD, HKD I ADHD W ŚWIETLE HIPOTEZY DIPOLOWEJ

ADD, HKD AND ADHD IN DIPOLE HYPOTHESIS

PIOTR WALECKI, JAN TRĄBKA

Zakład Bioinformatyki i Telemedycyny, Collegium Medicum Uniwersytetu Jagiellońskiego, 31-501 Kraków, ul. Kopernika 17

Streszczenie. W pracy zostało zaprezentowane oryginalne podejście do diagnozy oraz różnicowania zaburzeń aktywności i uwagi (ADD, HKD i ADHD) w kontekście analizy minimalnych uszkodzeń mózgu (MBD) i klinicznej teorii dipolowej.
Zaburzona elektrogeneza mózgowa, która przejawia się w zapisie hrEEG, tłumaczona w świetle hipotezy dipolowej pozwala na szczegółową analizę zmian w elektrycznej czynności mózgu pod wpływem farmakoterapii i terapii neurofeedback (NF). Dipol przedstawia patologiczną organizację elektrogenezy mózgu, ujmowanego jako całość. Gdy elektrogenezę rozpatrujemy w kategoriach sieci neuronalnej, czyli w płaszczyźnie dwuwymiarowej, to „skok napięciowy" w przestrzeni trójwymiarowej przekształca się w makroskopową organizację dipola. Uszkodzenie anatomiczne OUN o małych rozmiarach (MBD), które nie manifestuje się pod postacią objawów ubytkowych, staje się miejscem powstania dipola elektromagnetycznego. Od umiejscowienia oraz orientacji geometryczno-przestrzennej dipola zależy rozkład pola elektromagnetycznego, który warunkuje sektory lub kwadranty mózgu zakłócone, a nawet wyłączone z normalnego planu czynnościowego. Rozszerzające się w głębi mózgu pole elektromagnetyczne staje się czynnikiem hipersynchronizującym, zakłócającym autochtoniczną elektrogenezę korową, z kolei to zakłócenie produkcji fal mózgowych, uwidaczniając się w obrazie QEEG oraz MEG, staje się wyrazem czynnościowych zaburzeń.
Słowa kluczowe: ADD, HKD, ADHD, hipoteza dipolowa, EEG

Abstract. This paper presents innovative approach to the diagnosis and differentiation of activity and attention disorders (ADD, HKD and ADHD). Research was undertaken in the perspective of analysis of minimal brain dysfunctions (MBD) and clinical dipole hypothesis.
Disorder of brain electrogenesis, which manifests itself in high resolution EEG, explained in dipole hypothesis makes possible detailed analysis of changes in electrical brain activity under the influence of pharmacotherapy or neurofeedback therapy (NF). The dipole shows pathological electrogenesis of the brain, considered as a whole. As the electrogenesis was considered in the neural nets category, that is a two dimensional plane, then "voltage jump" in three dimensional space transforms into the macroscopic organization of the dipole. Minimal brain damage that shows no loss function symptoms became the electromagnetic dipole. The dipole localization and three-dimensional orientation determine distribution of electromagnetic field, which specifies brain sectors or quadrants exclusion with normal neuronal activity. Electromagnetic field spreading in the depth of brain becomes a hipersynchronizating feature that disrupts native cortex electrogenesis. Disruption of brain waves (demonstrated in QEEG and MEG) causes activity disorders.
Key words: ADD, HKD, ADHD, dipole hypothesis, EEG

1. Zarys problematyki

Zespół zaburzeń uwagi (Attention Deficit Disorder – ADD) oraz zespół nadpobudliwości (Hyperkinetic Disorder – HKD), po raz pierwszy zostały opisane przez Dr. Heinricha Hoffmana w „The Story of Fidgety Philips" w 1845 roku (*Fidgety Phil, won't sit still, he wriggles and jiggles and swings bank and forth*). Obecnie zespół nadpobudliwości psychoruchowej z zaburzeniami uwagi (*Attention Deficit Hyperactivity Disorder* – ADHD) jest jednym z najczęściej diagnozowanych zaburzeń dzieci w wieku szkolnym. ADHD w Polsce występuje szacunkowo od 3 do 12% populacji i przejawia się różnym stopniem nasilenia objawów oraz charakteryzuje się zindywidualizowanym obrazem syptomatologicznym. Badania populacyjne, które zostały przeprowadzone w USA stwierdzają, że na ADHD cierpi ok. 15 milionów Amerykanów. Zaburzanie to występuje zarówno u dzieci, jak i u dorosłych, dotyka wszystkie grupy etniczne i społeczne. Z badań statystycznych wynika, że częstość występowania zaburzenia większa jest w populacji mężczyzn, niż kobiet, u których znacznie częściej przybiera postać ADD, czyli zaburzeń uwagi i koncentracji bez komponenty nadpobudliwości. Wcześniej przyjmowano, że jest to głównie zaburzenie wieku dziecięcego, jednak dokładniejsze badania dowodzą, iż u wielu osób cierpiących na ADHD w dzieciństwie, zaburzenie przejawia się także w życiu dorosłym, choć przy nieco zmienionych objawach [1]. Zmiana symptomatologii dla ludzi dorosłych związana jest z nakładającymi się objawami zaburzeń wtórnych, będącymi pośrednim wynikiem sytuacji powodowanych przez ADHD oraz konsekwencją negatywnych oddziaływań społecznych (np. napiętnowanie, odrzucenie) [2].

Pierwsze badania naukowe nad zaburzeniami zachowania wyróżniającymi się nadpobudliwością i chwiejnością uwagi przeprowadził George F. Still publikując je w renomowanym

czasopiśmie naukowym Lancet (4, 19, 1902). Impulsywność i złe zachowanie postawiono wtedy w nowej perspektywie doszukując się jego przyczyn w zaburzeniach organicznych i genetycznych, zamiast odwoływać się tradycyjnych wskazań na lenistwo, nieprawidłowości w wychowaniu lub słabą wolę.

Badania nad zaburzeniami uwagi występującymi wraz ze wzmożoną motoryką rozwijały się w kierunku odnalezienia ich neurobiologicznego podłoża. Dlatego przez wiele lat określane były nazwami odwołującymi się do tak zakreślonego spektrum badawczego tj. zespół minimalnego uszkodzenia mózgu, zespół minimalnej dysfunkcji mózgu (*minimal brain damage, minimal brain dysfunction* – MBD) lub wczesnodziecięcy zespół psychoorganiczny [3], [4], [5]. Funkcjonowały również nazwy odwołujące się do objawów tj. syndrom hiperkinetyczny (*hyperkinetic syndrome*) czy zespół zaburzeń hiperkinetyczno-odruchowych. Obraz objawów był jednak bardzo zróżnicowany, wyróżniono około siedemdziesiąt różnych charakterystyk opisujących dany zespół zaburzeń. Dopiero w 1980 roku powstała ujednolicona kategoria diagnostyczna w trzecim wydaniu klasyfikacji chorób i zaburzeń psychicznych Amerykańskiego TowarzystwaPsychiatrycznego – *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) [6].

W 1994 roku wydano DSM-IV, gdzie ADHD (kod 314.01), opisano za pomocą bardziej ustandaryzowanych kategorii diagnostycznych. Opis nozologiczny zespołu nadpobudliwości ruchowej z deficytem uwagi charakteryzuje ADHD za pomocą triady objawów osiowych:

1) zaburzenia koncentracji uwagi (*inattention*),
2) impulsywność (*impulsivity*),
3) nadruchliwość czy nadmiar energii (*hyperactivity*).

W ramach tych trzech grup wyróżniono osiemnaście objawów występujących najczęściej. Wystąpienie minimum 12 objawów utrzymujących się w dłuższym okresie czasu może dawać podstawę diagnostyczną [7], [8], [9].

Zaburzenia uwagi:

1. opuszczanie szczegółów,
2. trudności u podtrzymaniu uwagi,
3. trudności w słuchaniu ze zrozumieniem,
4. nie kończenie rozpoczętych czynności,
5. trudności organizacyjne,
6. unikanie dłuższego wysiłku,
7. gubienie przedmiotów,
8. łatwe rozpraszanie się pod wpływem bodźców z zewnątrz,
9. słaba pamięć.

Impulsywność:

10. odpowiadanie na pytania zanim się je usłyszy do końca,
11. trudności z przestrzeganiem instrukcji, zasad bądź ról,
12. przerywanie innym, wtrącanie się w rozmowę lub przeszkadzanie.

Nadruchliwość:

13. nadmierna gadatliwość,
14. niemożność usiedzenia na krześle, częste nerwowe poruszanie rękoma lub stopami,
15. trudności z pozostawaniu na swoim miejscu, kiedy się tego wymaga,
16. niecierpliwe chodzenie tam i z powrotem,
17. hałaśliwe zachowanie,
18. nadmiar motoryki.

W dziesiątym wydaniu międzynarodowej klasyfikacji chorób i zaburzeń psychicznych ICD-10, propagowanym przez Światową Organizację Zdrowia (WHO), ADHD jest określane jako zespół nadpobudliwości ruchowej (*Hyperkinetic Disorder* – HKD, kod zaburzenia F90) [10]. Obecnie klasyfikacje DSM i ICD przyjmują tą sama triadę objawów osiowych oraz zespół osiemnastu wyżej wyszczególnionych objawów jako podstawowe kategorie diagnostyczne [11]. Różnice pozostają w nomenklaturze oraz wytycznych dotyczących diagnozowania chorób współwystępujących (w ICD diagnozowane są osobno, a w DSM mogą być łączone z ADHD jako jego podtypy) [12], [13].

Wobec różnorodności skojarzonych z tym zaburzeniem objawów towarzyszących właściwa diagnostyka wydaje się podstawowym problemem medycznym. Występowanie licznych podtypów ADHD utrudnia i tak skomplikowane rozpoznanie, które jest warunkiem poprawnego procesu terapeutycznego. U prawie 70% ludzi cierpiących w dzieciństwie z powodu ADHD, objawy utrzymują się nadal w życiu dorosłym [14], [15]. Najczęściej jednak łatwe do wychwycenia symptomy nadpobudliwości bywają stłumione, a właściwe objawy osiowe są maskowane przez zaburzenia współtowarzyszące, które początkowo bardziej rzucają się w oczy, np. depresja, alkoholizm, uprawianie hazardu, ryzykowny lub nieuporządkowany styl życia [16], [17], [18], [19].

Podstawowe znaczenie ma zatem diagnostyka różnicowa pozwalająca wykluczyć zbliżone objawowo zachowania, zespoły zaburzeń lub jednostki chorobowe, tj.: zaburzenia lękowe, depresję, chorobę afektywną dwubiegunową, zaburzenia zachowania u dzieci, zaburzenia kontroli impulsów psychicznych (np. kleptomania, piromania), stany chronicznego zmęczenia, płodowy zespół alkoholowy (FAS), nadczynność tarczycy, zatrucie ołowiem, organiczne zaburzenia uczenia się (np. dysleksja, dysgrafia, dyskalkulia), lekomania (np. uzależnienie od barbituranów), uzależnienie od substancji pobudzających (np. kawy, kofeiny) lub psychoaktywnych (np. kokainy, alkoholu), zaburzenia obsesyjno-kompulsyjne (nerwica natręctw), zaburzenia osobowości (np. borderline, narcyzm), zaburzenia opozycyjno-buntownicze, padaczka (napad absencyjny), stany dysocjacji lub nadmierny stres, brak snu i przemęczenie [20], [21], [22].

2. Metody diagnostyczne

Diagnoza ADHD odbywa się przede wszystkim na podstawie analizy historii zaburzeń, nieprawidłowego zachowania w szkole i problemów w rodzinie. Trudno przeprowadzić wyraźną linię oddzielającą normalne zachowanie od zaburzonego, rozpoznanie przede wszystkim opiera się na porównaniu dziecka z jego grupą rówieśniczą (lub analogicznie dorosłego z jego grupą odniesienia). Jeżeli dziecko wyraźnie odstaje od reszty poprzez swoje zachowanie lub wyniki w nauce oraz wykluczone zostały inne przyczyny chorobowe lub społeczne, to wtedy można postawić podejrzenie o ADHD. Oprócz wywiadu przedmiotowego i podmiotowego oraz środowiskowego ważnych wskazówek diagnostycznych dostarczają testy psychologiczne i neurobiologiczne.

Pewne wyniki u dzieci testu inteligencji WISC (*Wechsler Intelligence Scale for Children*) mogą sugerować wystąpienie ADHD, są to: słabe rezultaty rozwiązywania zadań z arytmetyki i kombinatoryki, niskie wyniki z testów dotyczących funkcjonowania pamięci, a także rozbieżność pomiędzy wynikiem z części werbalnej i niewerbalnej. Najbardziej istotne informacje diagnostyczne pochodzą z wyspecjalizowanych testów do badania procesów uwagi w zakresie zaburzeń jej ciągłości

oraz selektywności i odporności na zakłócenia. Najważniejsze tego typu testy dostępne w wersji komputerowej to tzw. testy ciągłego wykonywania (*Continuum Performance Test* – CPT), np. „Test zmienności uwagi" (*Test of Variables of Attention* – TOVA) [23] oraz testy oceniające sprawność pamięci operacyjnej i funkcji wykonawczych, np. „Test sortowania kart Wisconsin" (*Wisconsin Card Sorting Test* – WCST) czy „Test Stroopa" (*Color-Word Interference Test*) [24]. Osobną grupę stanowią skale kliniczne (*rating scales*) oraz wywiady standaryzowane (*standardized interviews*) dostarczające bardzo dane na temat historii, symptomatologii oraz nasilenia zaburzeń. Skale kliniczne są dostępne i odpowiednio zindywidualizowane zarówno dla przypadków ADHD u dzieci, jak i u dorosłych. Do najbardziej rozpowszechnionych należą: *Child Attention Problem Rating Scale*, *ADHD Rating Scale* (ADHDRS), *Conners Adult ADHD Rating Scale* (CAARS) [25], *Adult ADHD Self-Report Scale* (ASRS) [26] oraz *Attention Deficit Disorders Evaluation Scale* (ADDES) [27]. Cenne dane mogą pochodzić także ze skal klinicznych służących do oceny ogólnego poziomu funkcjonowania życiowego pacjenta (*Global Assessment of Functioning* – GAF) oraz skal dotyczących nasilenia i przebiegu obserwowanych u pacjenta zaburzeń (*Comprehensive Assessment of Symptoms and History Scale* – CASH) [28].

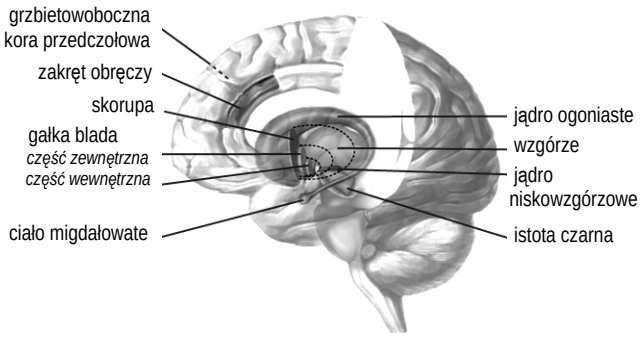
W niektórych przypadkach wskazane może być wykonanie testów do oceny funkcji werbalnych. Problemy z rozwojem umiejętności językowych występujące w ADHD mogą przejawiać się na wiele sposobów i dotykać różnych mechanizmów procesów poznawczych. Podstawowe dwie grupy odnoszą się do: 1) problemów z przyswajaniem słów lub zdań oraz umiejętności formułowania wypowiedzi, 2) problemy z ekspresją werbalną, które zaburzają zdolność pisania lub mówienia oraz tworzenia konceptualizacji (wyobrażeń w mózgu).

Ze względu na możliwą etiopatogenezę organiczną (uszkodzenia w OUN np. guzy, krwiaki itp.) obserwowanych zaburzeń konieczne może być przeprowadzenie odpowiednich badań neurologicznych i testów neuropsychologicznych (*The Halstead-Reitan battery of psychological organic tests*, *Cambridge Neuropsychological test Automated Battery* – CANTAB).

Właściwa diagnostyka jest ściśle powiązana z procesem terapeutycznym [29], i to nie tylko w momencie rozpoznania lecz także przy monitorowaniu przebiegu leczenia i ustalaniu jego końcowego efektu [30], [31], [32]. Wszystkie powyżej zaprezentowane metody testowe są dobrym predykatorem niekorzystnych zmian i zaburzeń zachowania oraz czynności poznawczych [33]. Jednakże bywają sytuacje, iż metody te zawodzą lub niemożliwe staje się korzystanie z nich zgodnie z ich wymaganiami. Częstym przypadkiem jest, że dzieci z ADHD potrafią się bardzo dobrze koncentrować na wykonywanym zadaniu podczas badania w lekarskim gabinecie, ponieważ w takiej sytuacji pojawia się element nowości wzbudzający ciekawość. Maskowanie objawów ADHD w sytuacjach ciekawych poznawczo, a czasem nawet stresujących i lękowych, które niejako wymuszają stan skupienia, bywa przyczyną mylnej diagnozy lekarskiej [34]. Również wysoki iloraz inteligencji może opóźnić rozpoznanie, ponieważ dzięki zdolnościom których jest on wyznacznikiem łatwiej przychodzi kompensowanie ubytków sfery poznawczej czy wykonaniowej [35], [36]. Dlatego ciągle trwają poszukiwania metody diagnostycznej mniej wrażliwej na zakłócające warunki środowiska oraz niezależnej od zdolności, wykształcenia i specyficznych cech osobowości lub kontekstu społeczno-kulturowego. Metody takie muszą bazować na psychofizjologicznych i neuroanatomicznych przesłankach, i jako takie mogą oferować bardziej obiektywną i stabilną diagnozę [37]. Jednymi z metod budzących największe nadzieje jest wykorzystanie urządzeń służących do zapisu aktywności bioelektrycznej lub magnetycznej mózgu, które bazują na wiedzy o fizjologicznych i anatomoklinicznych przyczynach ADHD.

3. Neurofizjologia ADHD/HKD

Diagnostyka ADHD/HKD tworzona jest na bazie symptomatologii zaburzeń, gromadzone są jednak dane pochodzące z nauk neurologicznych na temat biologicznych przyczyn tych zaburzeń. Wcześniejsze studia zajmujące się neuropsychologią, neuroanatomią, neurochemią i biologią molekularną ADHD/HKD opisują nieprawidłowości w określonych obszarach neuroanatomicznych i neurochemicznych szlakach [38], [39]. Badania z wykorzystaniem techniki rezonansu magnetycznego (MRI) pokazują pewne rejony płatów czołowych oraz jader podstawy (jadro ogoniaste i gałka blada), które są o ok. 10 % mniejsze w grupie dzieci z ADHD niż w grupie kontrolnej [40], [41], [42], [43]. Obecnie większą uwagę poświęca się metodom funkcjonalnym jak fMRI, SPECT, PET, EEG, EPR czy MEG, które ukazują dynamicznie zmieniający się stan czynnościowy wyróżnionych obszarów anatomicznych podczas wykonywania testów [44].



Rys .1. Neuroanatomiczne regiony biorące udział w ADHD/HKD

Przez wiele lat jedną z podstawowych neurofizjologicznych hipotez nadpobudliwości była hipoteza katecholaminowa [45], która w zmodyfikowanej wersji mówi, że wynikiem zaburzeń zachowania i koncentracji są: zmniejszona aktywność noradrenargiczna i zaburzone dopaminergiczne presynaptyczne funkcjonowanie układu [46]. Jednakże badania biochemiczne i testy kliniczne metabolitów neuroprzekaźników obecnych w moczu prowadzone przez dwadzieścia lat nie udokumentowały postulowanej roli katecholamin w powstaniu ADHD. Defekt w pojedynczym systemie neuroprzekaźnikowym nie może być wyłącznym regulatorem zaburzeń hiperkinetycznych, ponieważ w neuronach może dochodzić do konwersji dopaminy (DA) w noradrenalinę, a poza tym wiele leków działających na katecholaminy działa również na serotoninę (5-HT) [47], [48]. Ponadto możliwy jest również wpływ na konkretny układ innych neuroprzekaźników jak kwas γ-aminomasłowy – GABA. Oparta na tej hipotezie farmakoterapia polega na stymulacji dopaminy i norepinefryny, które zwiększają aktywność słabiej działających obszarów przedczołowych kory mózgowej wzmacniając procesy uwagi i hamowania korowego sieci ruchowych w obszarach koro-

wych i podkorowych [49], [50]. Transmisja katecholaminergiczna związana jest z pęczkiem środkowym przodomózgowia (MFB), stanowi podstawową część układu nagrody. Strukturami związanymi z układem nagrody są: kora przedczołowa, jądro wtrącone, część grzbietowa pnia mózgu oraz brzuszna część nakrywki śródmózgowia (*ventral tegmental area* – VTA), od której bierze początek dopaminergiczny szlak mezolimbiczny do kory czołowej, ciała migdałowatego i hipokampa. Amfetamina aktywuje układ nagrody poprzez wyrzut DA powodując odczuwanie przyjemności, a podobny do niej strukturalnie lek – metylfenidat, stosowany u chorych z ADHD, przyczynia się do podniesienia koncentracji i obniżenia nadpobudliwości [51].

Potrzeba ciągłego wzmocnienia jest jednym z kluczowych elementów ADHD, z tego powodu wydaje się, iż funkcjonowanie ośrodków nerwowych odpowiedzialnych za procesy nagrody może mieć podstawową rolę w powstaniu i przebiegu tych zaburzeń. Patomechanizm ADHD oparty o anatomokliniczne struktury układu nagrody regulowanego przekaźnictwem katecholaminergicznym [52] jest szczególnie użyteczny w farmakoterapii i terapii behawioralnej tych zaburzeń.

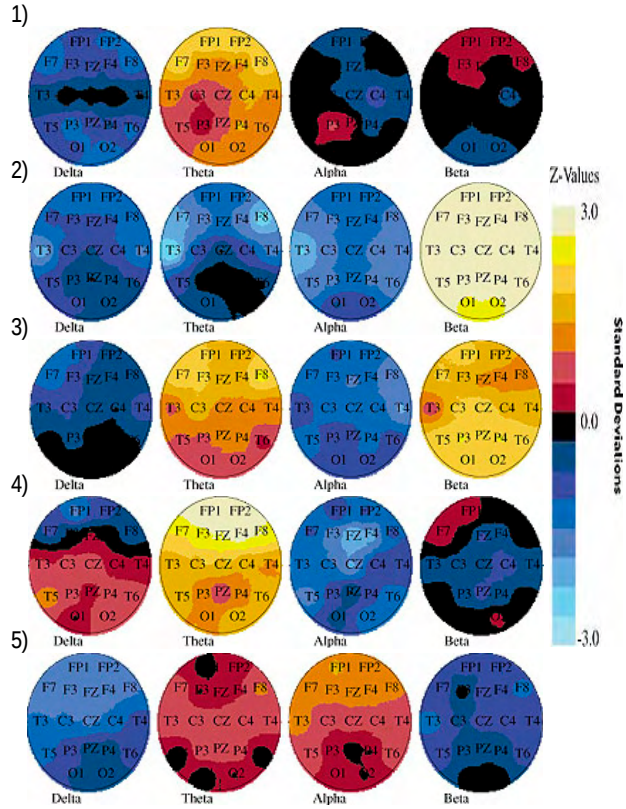
Również studia z genetyki molekularnej pokazują, że diagnoza ADHD może być związana z polimorfizmem w niektórych genach dopaminowych, które kodują m.in.: receptor dopaminy D4 (DRD4), receptor dopaminy D5 (DRD5), transporter dopaminy, beta hydroksylazę dopaminy (DBH) oraz białko 25kD (SNAP-25) [53], [54], [55].

4. Hipoteza dipolowa

Hipoteza dipolowa jest szeroko używanym modelem neuronalnego źródła (*source model*) zaburzeń zachowania dla aplikacji diagnostycznych wykorzystujących EEG i MEG. Różne rozwinięcia i zastosowania tej hipotezy przedstawiane są niemal na każdym międzynarodowym kongresie dotyczącym badań nad topografią mózgu [56], [57].

Lokalizacja aktywności bioelektrycznej mózgu bazująca na zapisie EEG przyczyniła się do powstania kilkunastu zawansowanych metod obrazowania w naukach neurologicznych (*neuroscience*), z których wyróżnić należy: analizę grafoelemetów klasycznych zapisów hrEEG za pomocą metod nieliniowych, analizę ilościową i analizę spektralną: QEEG i mapowanie EEG dajace topograficzny rozkład wybranego parametru (np. amplitudy, częstotliwości). BEAM i Brainmapping są sposobami przedstawienia w formie mapy składowych elektrycznej czynności mózgu, przekształconych różnymi metodami analiz matematycznych np. FFT (szybka transformata Fouriera). Mogą być one stosowane w połączeniu z analizą automatyczną CEEG, której zadaniem jest wyszukiwanie synchronizacji i desynchronizacji fal wolnych oraz zespołów iglic lub fal ostrych z falą wolną w wyróżnionych ośrodkach neuroanatomicznych. Za ich pomocą możliwa staje się obiektywna diagnostyka zaburzeń czynności poznawczych opierająca się na neurobiologicznej aktywności mózgu.

W przypadku ADHD/HKD opisano charakterystyczny współczynnik theta/beta (*theta/beta ratio* – TBR) sygnalizujący udział fal wolnych w elektrogenezie mózgowej. Współczynnik ten dla osoby dorosłej wynosi 1-2, a dla dzieci – najczęściej 2-3, natomiast osoby posiadające kłopoty z koncentracją uwagi mają TBR powyżej 3 [58]. Na rys. 2 pokazano różnice w aktywności bioelektrycznej mózgu między osobami z ADHD i z grupy kontrolnej [59], [60].

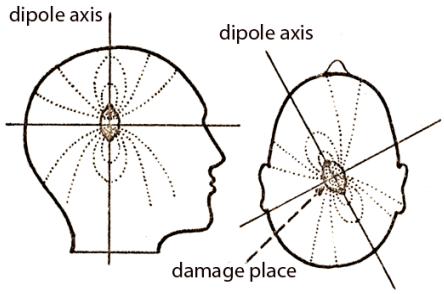


Rys. 2. Obrazy wykonane za pomocą techniki QEEG pokazują odchylenia od normalnego poziomu (kolor czarny) fal delta, theta, alpha i beta w przypadku pięciu podtypów ADHD: 1) Wysoki poziom theta, 2) Wysoki poziom beta właściwy dla ADHD powiązanego z alkoholizmem, zaburzeniami lękowymi, bezsennością, 3) Wysoki poziom theta właściwy dla ADHD powiązanego z alkoholizmem i zaburzeniami lękowymi, 4) Wysoki poziom theta właściwy dla ADHD powiązanego z zaburzeniami uczenia się, 5) Wysoki poziom theta właściwy dla ADHD powiązanego z depresją [59], [60]

Istotne informacje dostarczają badania nad reakcjami OUN na bodźce odbierane przez narządy zmysłów, które uwidaczniają się w postaci serii fal nazywanych potencjałami wywołanymi (*evoked potentials* – EP). Szczególne znaczenie mają składowe endogenne (*event releted potentials* – EPR), które pochodzą z pobudzenia ośrodków korowych i podkorowych gdzie przetwarzane są dane dotyczące oceny bodźca i wykonywanego zadania [61]. Za pomocą elektrod odbierane są sygnały pokazujące średnią aktywność grupy blisko położonych neuronów, które reprezentują synchroniczną zmianę potencjału polowego powstałego w wyniku prezentacji bodźca zmysłowego [62]. Metody te są szczególnie obiecujące ponieważ umożliwiają eksperymentalne określenie dynamiki aktywności sieci neuronalnej w miejscu rejestracji, w zmieniających się sytuacjach behawioralnych [63]. W zależności od rodzaju odbieranych bodźców można wyróżnić potencjały wywołane czuciowe (*somato-sensory evoked potentials* – SEPs) [64], słuchowe (*auditory evoked potentials* – AEPs) [65] i wzrokowe (*visually evoked potentials* – VEPs), natomiast ze względu na czas trwania: potencjały o krótkiej latencji (*short-latency*), długiej latencji (*long-latency*) i trwające w równowadze (*steady-state*) [66].

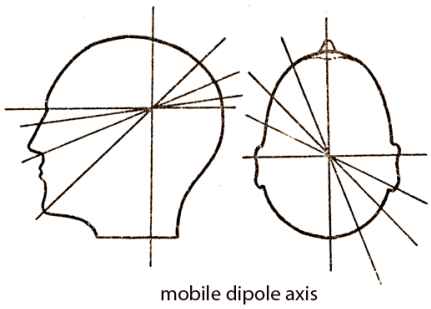
Dipol składa się z dwóch pól elektrycznych o znaku przeciwnym, oddzielonych od siebie linia zerową, która przebiega prostopadle do osi dipola, czyli linii łączącej obydwa bieguny dipola. Miarą dipola jest jego moment elektryczny *pe*=*q**1. Elektryczny moment dipolowy jest równy iloczynowi odległości

miedzy różnoimiennymi biegunami dipola i wartości ładunku dodatniego [67], [68]. Linie izoelektryczne pola są nierównomiernie rozłożone w przestrzeni, największe ich zagęszczenie, obrazujące najwyższy gradient napięciowy pola elektrycznego, znajduje się w sąsiedztwie linii zerowej. Od umiejscowienia oraz orientacji geometryczno-przestrzennej dipola zależy rozkład pola elektrycznego (patrz rys. 3), który warunkuje sektory lub kwadranty mózgu zakłócone, a nawet wyłączone z normalnego planu czynnościowego [69].



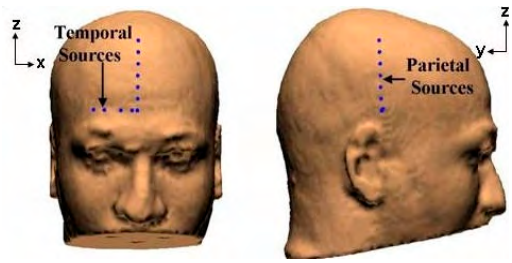
Rys. 3. Rozkład linii pola elektrycznego w dipolu [69]

W ADHD zakładamy istnienie głębokiej lokalizacji ogniska dipola elektrycznego w obszarze środkowym (*basal ganglia*, *hippocampus*). Rozszerzające się w głębi mózgu elektryczne pole staje się czynnikiem hipersynchronizującym, zakłócającym autochtoniczną elektrogenezę korową, z kolei to zakłócenie produkcji fal mózgowych, uwidaczniając się w obrazie EEG, staje się wyrazem czynnościowych zaburzeń.



Rys. 4. Ruchome osie dipola [69]

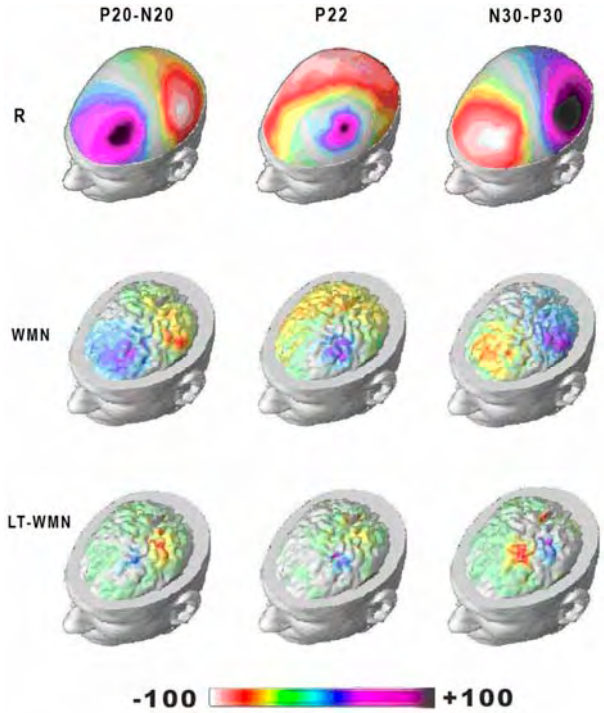
Dipol elektryczny mózgu jest układem dynamicznym, w którym może następować zmiana orientacji przestrzennej linii izoelektrycznych pola (rys. 4).



Rys. 5. Lokalizacja dipoli w płacie ciemieniowym i skroniowym umieszczona na realistycznym modelu głowy. Na rysunku pokazano tylko 5 z 23 skroniowych i 7 z 23 ciemieniowych lokalizacji dipoli. Każda lokalizacja może być wykorzystana do symulacji potencjałów dipoli wektorów stycznych (*tangential*) i promieniowych (*radial*) [71]

Dipol przedstawia patologiczną organizację elektrogenezy mózgu, ujmowanego jako całość. Gdy elektrogenezę rozpatrujemy w kategoriach sieci neuronalnej, czyli w płaszczyźnie dwuwymiarowej, makroskopowej to „skok napięciowy" w przestrzeni trójwymiarowej przekształca się w makroskopową organizację dipola [69]. Uszkodzenie anatomiczne OUN o małych rozmiarach [70], które nie manifestuje się pod postacią objawów ubytkowych, staje się miejscem powstania dipola elektrycznego.

Również zaburzenia u osób z ADHD w obszarze grzbietowobocznej korze przedczołowej (DLPFC), której funkcje związane są ze skupieniem i podtrzymywaniem uwagi [55], [72], mogą być rejestrowane jako sumaryczna aktywność postsynaptyczna w komórkach piramidowych ułożonych równolegle do siebie z dendrytami szczytowymi skierowanymi prostopadle do powierzchni kory, które tworzą podczas pobudzenia wyraźnie zorientowane dipole elektryczne [62].



Rys. 6. Rozkład dipoli elektrycznych trzech grup załamków (P20-N20, P22, N30-P30) czuciowych potencjałów wywołanych (SEPs) nałożonych na model głowy uzyskany metodą MRI. Analiza wykonana została przy pomocy trzech metod matematycznych: *raw potentials* (R), *weighted minimum norm* (WMN), *Laplacian-transformed weighted minimum norm* (LT-WMN) [64]

Aktywność bioelektryczna mózgu w przypadkach ADHD wykazuje podwyższony współczynnik TBR (rys. 2), dlatego szczególnie ważne dla diagnostyki wydają się być wysokoamplitudowe (do 500 mV) fale θ, które są przejawem synchronizacji hipokampa podczas reakcji orientacyjnej, co wykazuje na ich związek z mechanizmami wzbudzenia i uwagi [55], [58], [60], [73].

Podsumowując można stwierdzić, iż zarówno w ośrodkach korowych (DLPFC), jak i podkorowych (*basal ganglia*) odpowiedzialnych za powstanie ADHD może występować patologiczna aktywność bioelektryczna manifestująca się charakterystyczną organizacją dipolarną. Umiejętność jej detekcji i rozpoznania stanowi jedno z głównych założeń rozwojowych dla diagnostyki ADHD.

Literatura cytowana

1. Wender P.H.: Attention-deficit hyperactivity disorder in adults. New York: Oxford University Press. 1995.
2. Weiss M., Hechtman L.T., Weiss G.: ADHD in adulthood: A guide to current theory, diagnosis, and treatment. Baltimore: Johns Hopkins Univ Press. 1999.
3. Clement S.D.: Minimal Brain Dysfunction in Children. NINDS Monograph No. 3, U.S. Department of Health, Education, and Welfare 1966.
4. Paine R.: Syndrome of minimal cerebral damage, Ped. Clin. Of N Am, 15, 1968.

5. Small L.: The Minimal Brain Dysfunctions. Diagnosis and Treatment, Collier Macmillan Publishers, London 1982.
6. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (3th ed.). Washington, DC: 1980.
7. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (4th ed.). Washington, DC: 1994.
8. Faraone S.V., Biederman J., Friedman D.: Validity of DSM-IV subtypes of attention-deficit/hyperactivity disorder: A family study perspective. Journal of the American Academy of Child and Adolescent Psychiatry, 39: 300-307, 2000.
9. Weiss G., Hechtman L.T.: Hyperactive children grown up: ADHD in children, adolescence, and adults (2nd ed.). New York: Guilford Press. 1993.
10. The ICD-10 Classification of Mental and Behavioral Disorders: Clinical Descriptions and Diagnostic Guidelines, tenth revision. Geneva: WHO, 1992.
11. Murphy K.R.: Psychological counseling of adults with ADHD. In R.A. Barkley (ed.), Attention deficit hyperactivity disorder: A handbook for diagnosis and treatment (pp.582-591). New York: Guilford Press. 1998.
12. Murphy K., Barkley R.: Attention deficit hyperactivity disorder adults: Comobidities and adaptive impairments. Comprehensive Psychiatry, 37(6), 393-401, 1996.
13. Millstein R.B., Wilens T.E., Biederman J., Spencer T.J.: Presenting ADHD symptoms and subtypes in clinically referred adults with ADHD. Journal of Attention Disorders, 2 (3), 159-166, 1997.
14. Lahey B.B., Applegate B., McBurnett K., Biederman J., Greenhill L., Hynd G.W., Barkley R.A., Newcorn J., Jensen P., Richters J., Garfinkel B., Kerdyk L., Frick P.J., Ollendick T., Perez D., Hart E. L., Waldman I., Schaffer D.: DSM-IV field trials for attention deficit hyperactivity disorder in children and adolescents. American Journal of Psychiatry, 151 (11), 1673-1685. 1994.
15. Heiligenstein E., Conyers L.M., Berns A.R., Smith M.A.: Preliminary normative data on DSM-IV attention deficit hyperactivity disorder in college students. Journal of American College Health, 46: 185-188, 1998.
16. Ratey J.J., Greenberg M.S., Bemporad J.R., Lindem K.J.: Unrecognized attention-deficit hyperactivity disorder in adults presenting for outpatient psychotherapy. Journal of Child and Adolescent Psychopharmacology, 2 (4): 267-275, 1992.
17. Mannuzza S., Klein R.G., Bessler A., Malloy P., LaPadula M.: Adult outcome of hyperactive boys: Educational achievement, occupational rank, and psychiatric status. Archives of General Psychiatry, 50: 565-576, 1993.
18. Mannuzza S., Klein R.G., Bessler A., Malloy P., LaPadula M.: Adult psychiatric status of hyperactive boys grown up. Am J of Psychiatry, 155: 493-498, 1998.
19. Silver L.: Attention deficit/hyperactivity in adult lives. Child & Adolescent Psychiatric Clinics of North America, 9: 511-523, 2000.
20. Barkley R.A., ed. Attention Deficit Hyperactivity Disorder: A Handbook for Diagnosis and Treatment. New York, NY: Guilford Press; 2001: 596–606.
21. Goldman L.S., Genel M., Bezman R.J., et al. Diagnosis and treatment of attention-deficit/hyperactivity disorder in children and adolescents: Council on Scientific Affairs, American Medical Association. Journal of the American Medical Association. 279 (14): 1100–1107, 1998.
22. Wolf L.E., Wasserstein J.: Adult ADHD: concluding thoughts. Annals of the New York Academy of Sciences, 931: 396-408, 2001.
23. Greenberg L. M., T.O.V.A. continuous performance test manual. Los Alamitos, CA: Universal Attention Disorders 1996.
24. Pennington B. F., Ozonoff S., Executive functions and developmental psychopathology. J Child Psychol Psychiatry, 1996, 37:51-87.
25. Conners C.K., Ernhardt D., Sparrow E.: Conners' adult ADHD rating scales. Toronto: Multi-Health Systems 1999.
26. Conners C.K., Erhardt D., Epstein J.N., Parker J.D.A., Sitarenios G., Sparrow E.: Self-ratings of ADHD symptoms in adults I: Factor structure and normative data. Journal of Attention Disorders, 3 (3): 141-151. 1999.
27. McCarney, S. B., Attention Deficit Disorders Evaluation Scale. Columbia, MO: Hawthorne Press. 1995.
28. Andreasen N. C., Flaum M., Arndt S., The comprehensive assessment of symptoms and history (CASH). An instrument for assessing diagnosis and psychopathology. Arch. Gen. Psychiatry, 49: 615-623. 1992.
29. American Academy of Child and Adolescent Psychiatry. Practiceparameters for the assessment and treatment of children, adolescents, and adults with attention-deficit/hyperactivity disorder. Journal of the American Academy of Child and Adolescent Psychiatry, 36 (suppl), 85S-121S, 1997.
30. Bussing R., Zima B.T., Belin T.R.: Variations in ADHD treatment among special education students. Journal of the American Academy of Child and Adolescent Psychiatry. 37 (9): 968–976, 1998.
31. Clinical practice guideline: treatment of the school-aged child with attention-deficit/hyperactivity disorder. Pediatrics, 108: 1033-44, 2001.
32. Greenfield B., Hechtman L., Bouffard R., Minde K., Looper K., Ochs E., Eakin L.: Correspondance between self and significant other reports of childhood and current adult ADHD symptoms. Poster presented at the International Society for Research in Child and Adolescent Psychopathology, Vancouver 2001.
33. Multimodal Treatment Study of Children With ADHD (MTA) Cooperative Group. A 14-month randomized clinical trial of treatment strategies for attentiondeficit/hyperactivity disorder, Arch. Gen. Psychiatry, 56, 1073-86, 1999.
34. Zarin D.A., Suarez A.P., Pincus H.A., et al. Clinical and treatment characteristics of children with attention-deficit/hyperactivity disorder in psychiatric practice. J Am Acad of Child and Adolescent Psychiatry. 37 (12): 1262–1270, 1998.
35. Nadeau K.: Psychotherapy for women with ADHD. In K. Nadeau & P. Quinn (Eds.), Understanding Women with ADHD. Silver Spring, Advantage Books 2002.
36. Bemporad J., Zambenedetti M.: Psychotherapy of adults with attention-deficit disorder. Journal of Psychotherapy Practice and Research, 5 (3), 228-237, 1996.
37. Wasserstein J., Wolf L.E., LeFever F. (Eds.), Attention deficit disorder: Brain mechanisms and life outcomes. The New York Academy of Sciences, 2001.
38. Berger A., Posner M.I.: Pathologies of brain attentional networks. Neurosci Biobehav Rev, 24(1): 3-5. 2000.
39. Wasserstein J., Wolf L.E., LeFever F. (Eds.), Attention deficit disorder: Brain mechanisms and life outcomes. The New York Academy of Sciences, 2001.
40. Aylward E. H., Reiss A. L..: Reader M.J., et al.: Basal ganglia volumes in children with attention-deficit hyperactivity disorder, J Child Neurol, 11(2): 112-115, 1996.
41. Filipek P.A., Semrud-Clikeman M., Steingard R.J., et al.: Volumetric MRI analysis comparing subjects having attention-deficit hyperactivity disorder with normal controls, Neurology, 48(3): 589-601, 1997.
42. Liu X., Matochik J. A., Cadet J.-L., London E. D.: Smaller Volume of Prefrontal Lobe in Polysubstance Abusers: A Magnetic Resonance Imaging Study, Neuropsychopharmacology, 18: 243–252, 1998.
43. Castellanos F.X., Lee P.P., Sharp W., et al.: Developmental trajectories of brain volume abnormalities in children and adolescents with attention-deficit/hyperactivity disorder. JAMA 288(14): 1740-1748, 2002.
44. Casey B.J., Trainor R.J., Orendi J.L., Schubert A.B., Nystrom L.E., Giedd J.N., et al.: A developmental functional MRI study of prefrontal activation during performance of a go-no-go task. J Cogn Neurosci, 9: 835–847, 1997.
45. Schildkraut J.J.: The catecholamine hypothesis of affective disorders: A review of the supporting evidence. Am J Psychiatry, 122: 509–522, 1965.
46. Kornetsky C.: Psychoactive drugs in the immature organism, Psychopharmacologia, Vol. 17: 105-136, 1970.
47. Pliszka S.R., McCracken J.T., Maas J.W.: Catecholamines in attention-deficit hyperactivity disorder: current perspectives, J Am Acad Child Adolesc, Psychiatry. 35, 264-72, 1996.
48. Arnstern A. F.: ADHD, Part 2: Norepinephrine has a critical modulatory influence on prefrontal cortical function, J Am Acad Child Adolesc Psychiatry, 39, 2000.
49. Arnsten A.F., Steere J.C., Hunt R.D.: The contribution of alpha2-noradrenergic mechanisms to prefrontal cortical cognitive function. Arch General Psychiatr, 53: 448-455, 1996/
50. Roth R., Elsworth J.: Biochemical pharmacology of midbrain dopamine neurons. In Psychopharmacology: The Fourth Generation of Progress. Edited by Bloom F., Kupfer D., New York: Raven Press; 1995: 227-243.
51. Castellanos F.X.: Toward a pathophysiology of attention deficit/hyperactivity disorder, Clin Pediatr, 36: 381-393, 1997.
52. Wender P.H., Reimherr F.W., Wood D., Ward M.: A controlled study of methylphenidate in the treatment of attention deficit disorder, residual type, in adults. American Journal of Psychiatry, 142 (5): 547-552, 1985.
53. Civelli O.: Molecular biology of the dopamine receptor subtypes. In Psychopharmacology: The Fourth Generation of Progress. Edited by Bloom F., Kupfer D., New York: Raven Press; 1995: 155-161.
54. Meador-Woodruff J.: Differential expression of autoreceptors in the ascending dopamine systems of the human brain. Neurobiology, 91: 8297-8301, 1994.
55. Mrzljak L., Bergson C., Pappy M., Levenson R., Goldman-Rakic P.: Localization of dopamine D4 receptors in GABAergic neurons of the primate brain. Nature, 381: 245-248, 1996.
56. Frontiers in Human Brain Topography - 15th World Congress of the International Society for Brain Electromagnetic Topography (ISBET 2004), Japan 11-14 April 2004, Eds. M. Nakagawa, et al. (Otsubo H., et al.: Dipole source localization of epileptic discharges in EEG and MEG; Horia J., Heb B.: Artefact elimination in spatiotemporal cortical dipole layer imaging with parametric projection filter).
57. Brain Topography Today - III Pan-Pacific Congress on Brain Topography (BTOPPS), Tokyo Bay, Japan, 1-4 April 1997 (Watanabe S., et al.: Dipole source modelling of pain-related brain responses: a magneto-encephalographic study; Sakurai A., et al. A study on dipole of wave V in auditory brainstem response).
58. Demos J.: Getting Started With Neurofeedback, New York, 112-119, 2005.
59. http://www.add-clinic.com/neuro/set1.html
60. Chabot, R.J., Merkin, H., Wood, L.M., Davenport, T.L., Serfontein, G.: Sensitivity and Specificity of QEEG in children with attention deficit or specific developmental learning disorders, Clinical Electroencephalography, Vol. 27, No. 1: 26-34, 1996.
61. Szelenberger W.: Potencjały wywołane, Wyd. Elmiko, 2001.
62. Wypych M., Wróbel A.: Identyfikacja stanu funkcjonalnego mózgu przy pomocy nowych metod analizy potencjałów wywołanych, Konferencja „Nowe metody w neurobiologii", 15 grudnia 2004: 49−55.
63. Wróbel A., Kublik E., Musiał P.: Gating of sensory activity within barrel cortex of the awake rat, Exp. Brain Res., 123: 117–123, 1998.
64. Babiloni F., Babiloni C., Locche L., Cincotti F., Rossini P. M., Carducci F.: High Resolution EEG: Source Estimates of Laplacian-Transformed Somatosensory-Evoked Potentials Using a Realistic MRI-Constructed Subject's Head Model, Med Biol Eng & Comput, 38: 512-519, 2000.
65. Trąbka W., Stanuch H., Trąbka J.: Automatic analysis of the long latency auditory evoked potentials using harmonical analysis. XII International Congress of Electroencephal. nad Clinical Neuroph., Rio de Janeiro, Brazil 1990.
66. Pipingas A.: Steady-State Visually Evoked Potential Correlates of Object Recognition Memory, Brain Sciences Institute, Swinbourne 2003.
67. De Munck J.C., et al.: Mathematical Dipoles are Adequate to Describe Realistic Generators of Human Brain Activity. IEEE Trans Biomed Eng, 35: 960-6. 1988.
68. Schimpf P.H., et al.: Dipole Models for the EEG and MEG. IEEE Trans Biomed Eng, 49(5):409-18. 2002.
69. Trabka J., et al.: The Otiatric Syndrome of Minimal Cerebral Damage (OSMCD). Part I. Folia Medica Cracoviensis, 16, 45., 1974.
70. Paine R.: Syndrome of minimal cerebral damage, Ped. Clin. Of North America, 15, 1968.
71. Whittingstall K., Stroink G., Gates L., Connolly J.F., Finley A.: Effects of dipole position, orientation and noise on the accuracy of EEG source localization, BioMedical Engineering OnLine 2003, 2: 14. http://www.biomedical-engineering-online.com/content/2/1/14
72. Goldman-Rakic P.S.: Topography of cognition: parallel distributed networks in primate association cortex. Annu Rev Neurosci, 11: 137-156. 1988.
73. Vieth J., et al.: Slow wave and beta wave activity associated with white matter structural brain lesions, localized by the dipole density plot. In: Baumgartner C., et al., Biomagnetism: fundamental research and clinical applications. Amsterdam: Elsevier/IOS-Press; 1995, pp. 50-54.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 273-276.

MODELOWANIE TRANSPORTU OBRAZU MODYFIKOWANYMI WŁÓKNAMI POLIMEROWYMI

MODELLING OF IMAGE TRANSPORTATION USING MODIFIED POLYMER FIBRES

MARIUSZ WÓJCIK

Katedra Biomateriałów Akademii Górniczo-Hutniczej, Kraków, wojmar@uci.agh.edu.pl

Streszczenie. W artykule przedstawiono próbę aplikacji teorii informacji dla procesu jej przetwarzania i przekazu na drodze oko-mózg i zaproponowano pewne rozwiązanie równań Maxwella w postaci równań hiperbolicznych jako podejście do rozpoznania procesu przekazu obrazu we włóknie nerwowym. Otrzymane rozwiązanie pokazuje, w jaki sposób żywy organizm może przesyłać obraz od receptora (oko) poprzez włókno nerwowe do mózgu. Odpowiednio zmodyfikowane włókno polimerowe po zadaniu na wejściu sterowalnym sygnałem elektrycznym powinno dać odpowiedź w postaci prądu elektrycznego o zmienionej częstotliwości i pola magnetycznego odpowiadającego natężeniu prądu. Ten model daje szansę na lepsze poznanie i zrozumienie mechanizmu przesyłania informacji jako historii zewnętrznego świata w procesie komunikowania się organizmów żywych.

Słowa kluczowe: przekaz obrazu, włókna nerwowe, włókna polimerowe, modelowanie matematyczne, równania Maxwella

Abstract. A certain solution of Maxwell's equations using hyperbolic formula was proposed as an approach to the recognition of the process of image transfer from a receptor (eye) trough neurone fibres into the brain; a trial application of an information theory for the image transfer is presented in the paper. Solution of the equation explains the way in which human body transfers an image from the eye into the brain. Suitably modified polymer fibres after inputting controlled electrical signal can give an answer as an electrical current with changed frequency and a magnetic field corresponding with an intensity of the current. This model gives an opportunity of better recognition and understanding of the mechanism of image transfer as an external history at the communication process between living organisms.

Key words: image transportation, neurone fibres, polymer fibres, mathematical modelling, Maxwell's equations

Wstęp

Szwajcarski uczony Dr Alexis Carrel w swojej znakomitej książce „Człowiek istota nieznana", za którą otrzymał nagrodę Nobla w roku 1932 [1], jeszcze w latach dwudziestych XX wieku sugerował, że bodźce zmysłowe pobudzając odpowiednie komórki w układach zmysłowych kory mózgowej, tworzą w nim pewną historię świata zewnętrznego i wydaje się, że mózg jako całość nie zwraca uwagi na nośnik informacji, jakim jest fala świetlna lub akustyczna, ale analizuje jego zawartość. Analizując zawartość przesyłanej za pomocą tych zmysłów informacji tworzącej właśnie historię świata zewnętrznego, człowiek realizuje sens swojego istnienia poprzez proces myślenia.

John Dewey, twórca interpretacji funkcjonalnej psychologii myślenia i psychodydaktyki, napisał, *że „myślenie jest to «umysłowa rewolucja» przebiegająca w sferze symbolicznych prób i błędów pomiędzy koncepcjami rywalizującymi o «życie» i jest traktowane jako proces modyfikowania niepewności hipotez pod wpływem uzyskiwanych informacji*" [2]. Ta definicja Dewey'a sugeruje istnienie tzw. przestrzeni rozwiązania, w której przebiega nieustanny proces wymiany informacji pomiędzy organizmami żywymi, a to na pewno jest siłą napędową istnienia gatunku ludzkiego.

Współcześnie Newsweek [3] w popularnonaukowym artykule „Zmysłowe bramy umysłu" napisał, że „*na razie nikt nie potrafi powiedzieć, dlaczego sygnały elektryczne, docierające do kory mózgowej, odbierane są przez mózg jako obrazy, skoro taki sam rodzaj sygnałów w innych częściach mózgu rozpoznawany jest jako dotyk lub dźwięk*".

Naukowy tygodnik New Scientist, informuje, że dr Paul Bach-y-Rita, neurolog z uniwersyteckiej kliniki w Wiscontin, wpadł na pomysł zastępowania jednego utraconego zmysłu innym i w przypadku jednego swoich pacjentów zamiast wzroku wykorzystał dotyk. Zamianę taką umożliwiło skonstruowanie nowego urządzenia nazwanego BrainPort – brama mózgu.

Badania naukowe prowadzone w ciągu ostatnich lat pokazują więc, że istnieją fascynujące powiązania pomiędzy różnymi zmysłami naszego organizmu, a skoro wzrok, słuch

i dotyk komunikują się ze sobą, możliwe staje się wykorzystanie zamiast jednej drogi do mózgu innej. Nasz mózg, znając źródło impulsów wysyłanych przez receptory oczu, uszu czy skóry, potrafi je bezbłędnie zinterpretować, co wskazuje na to, że nasze zmysły porozumiewają się z mózgiem w podobny sposób. Zarówno obrazy zewnętrznego świata, jak i wrażenia dotykowe czy dźwięki wpadające do uszu, zamieniane są w impulsy nerwowe i przewodzone za pomocą włókien nerwowych zbudowanych z neuronów do różnych miejsc kory mózgowej i tam za pomocą innego układu komórek nerwowych, nazwijmy ich komórkami pamięci, rozpoznawane jako informacje wzrokowe, dotykowe lub słuchowe.

Spróbujmy skupić się tylko na zmyśle widzenia. W odbieraniu fal elektromagnetycznych (świetlnych) i ich przekształcaniu na wrażenia zmysłowe bierze udział narząd odbiorczy, czyli oko oraz drogi i ośrodki łączące oko z polem wzrokowym kory mózgowej. Samo oko spełnia złożoną funkcję, zawiera układ optyczny, właściwe receptory, czyli pręciki i czopki oraz drogi dla impulsów nerwowych w samej siatkówce. W literaturze medycznej można znaleźć wiele szczegółowych opisów budowy i czynności oka [4]. Co się jednak dzieje dalej z tą falą na drodze oko-mózg i w jaki sposób następuje przekaz niesionej przez tę falę informacji do mózgu. Czy można w jakiś sposób próbować przeanalizować proces przekazu obrazu do mózgu. Jakiego instrumentu użyć do takiej analizy?

Autor spróbuje odpowiedzieć na te pytania, stosując pewne podejście matematyczne do równań Maxwela, które opisują teorię rozchodzenia się fal elektromagnetycznych. Prezentowane w niniejszym artykule podejście jest tylko niewielkim fragmentem szerszych rozważań nad przesyłem informacji w postaci impulsów fal elektromagnetycznych za pomocą odpowiednio modyfikowanych polimerowych włókien, stanowiących pewien nowy model służący dla odwzorowania funkcji włókien nerwowych.

Na zakończenie wstępu autor chce wyjaśnić, że zaproponowany tytuł artykułu wychodzi nieco poza ramy prezentowanego materiału, ale jest to świadoma próba zasygnalizowania omawianego zagadnienia dla wywołania dyskusji wśród specjalistów zajmujących się modelowaniem cybernetycznym systemów biologicznych.

1. Funkcjonowanie organizmu a teoria informacji

Rozważmy najpierw ogólne zagadnienia dotyczące funkcjonowania organizmu człowieka i dopasujmy doń pewne znane pojęcia z teorii informacji. Poznawalne i doświadczalnie wykrywane prawa fizjologiczne, jakim podlega żywy organizm oraz poszczególne jego układy, narządy, tkanki i komórki, określają zarówno warunki, jak i mechanizmy czynnościowe organizmów zapewniające ich prawidłowe funkcjonowanie w środowisku, w którym żyją. Wspólną właściwością wszystkich istot żywych jest przemiana materii, czyli metabolizm, który jako podstawowy warunek życia biologicznego żywego organizmu zachodzi w jego środowisku wewnętrznym i na który oddziałuje dodatkowo środowisko zewnętrzne. Zatem wewnętrzne środowisko organizmów żywych jest ściśle uzależnione od zewnętrznego. Powstałe i rozwinięte w procesie ewolucji wyspecjalizowane narządy i układy organizmów żywych pozwoliły na coraz lepsze przystosowanie ich do zmieniającego się środowiska zewnętrznego. Najlepszym przykładem jest tu organizm człowieka i dzięki rozwojowi ośrodkowego układu nerwowego jest on najlepiej przystosowany do zmian zachodzących w środowisku zewnętrznym. Te doskonale wyspecjalizowane układy przejęły na siebie funkcje pośredniczenia pomiędzy tymi środowiskami, aby zapewnić istnienie organizmu człowieka, a więc po adaptacji do warunków otoczenia zapewnić utrzymanie jego życia osobniczego oraz utrzymanie gatunku.

Konieczność zapewnienia stałego metabolizmu i utrzymania stałego środowiska wewnętrznego organizmu wymaga stałej precyzyjnej kontroli wszystkich czynności życiowych, takich jak odżywianie, oddychanie, wydalanie i krążenie. Taka kontrola środowiska wewnętrznego przypada autonomicznemu układowi nerwowemu, który stanowi końcowy odcinek drogi wzbudzania impulsowego nerwów, biegnącej od receptorów znajdujących się wewnątrz organizmu. Układ nerwowy przekazuje dalej wzbudzenie impulsowe bezpośrednio do tkanek i narządów wewnętrznych. Mając na uwadze fakt zachodzących ciągłych zmian zarówno w środowisku wewnętrznym, jak i zewnętrznym, łatwo jest zauważyć, że nieznaczna nawet zmiana w środowisku jest źródłem bodźców odbieranych przez receptory, a każdy bodziec niesie przecież w sobie informację. Ze środowiska wewnętrznego mogą to być na przykład informacje o prężności parcjalnej tlenu we krwi lub stężenia pH, natomiast z zewnętrznego fale świetlne czy akustyczne, a więc na pewno obraz otaczającego świata. Zatem poprzez receptory zachodzi proces przetwarzania informacji, ale i jej przenoszenia. Proces ten może odbywać się w sposób ciągły, czyli analogowy lub impulsowy, czyli cyfrowy. Przykładem analogowego może być proporcjonalne do wzrostu koncentracji jakiegoś hormonu w organizmie przenoszenie określonej w tym samym okresie czasu informacji w postaci wydzielania do krwi jakiegoś związku chemicznego. Przenoszenie cyfrowe zachodzi natomiast we włóknach nerwowych. Receptory spełniają więc funkcję przetworników zamieniających informację ze środowiska zewnętrznego lub wewnętrznego, natomiast włókna nerwowe stanowią kanały informacyjne. Informacja w tym przypadku jest zakodowana na przykład w częstotliwości, z jaką impulsy przewodzone są przez włókno nerwowe. Informacja przetwarzana jest w organizmie nie tylko przez receptory, ale również przez komórki nerwowe. Odebraną informację cyfrową komórki nerwowe zamieniają na informację analogową w postaci zmian potencjału elektrycznego błony komórkowej albo pobudzającego, albo hamującego.

Adoptując do fizjologii człowieka dla analizowania mechanizmów odpowiadających za procesy życiowe podstawowe pojęcia z teorii informacji takie, jak układy, systemy, wejścia, wyjścia, można za ich pomocą i łącząc je w postaci schematów blokowych, opisywać przetwarzanie i przenoszenie informacji w organizmie człowieka. W organizmie czynność jednych układów stale jest sterowana przez inne układy. Informacja przepływa od wyjścia układu sterującego do wejścia układu sterowalnego, wymuszając na jego wyjściu określony proces. Większość procesów fizjologicznych w organizmie podlega regulacji i kontroli dzięki sprzężeniu zwrotnemu, zapewniając stabilizację funkcji fizjologicznych. Powyższe przesłanki stanowić więc będą podstawę do analizy równań Maxwella dla podjęcia próby interpretacji zawartych w nich informacji, które mogą być zaadaptowane do opisu formalnego przesyłu obrazu.

2. Modelowanie transportu obrazu

Układ podstawowych równań elektromagnetyzmu obejmujących: prawo Gaussa dla elektryczności i magnetyzmu, prawo Ampere'a dotyczące efektu magnetycznego zmieniającego się pola elektrycznego lub prądu oraz prawo indukcji Faradaya dotyczące efektu elektrycznego zmieniającego się pola magnetycznego nosi nazwę równań Maxwella [5]. Wiążą one ze sobą doświadczenia o bardzo szerokim zakresie i pozwalają one przewidywać nowe wyniki doświadczeń. Zakres zastosowań równań Maxwella jest niezwykle duży, obejmuje podstawowe zasady działania wszystkich urządzeń elektromagnetycznych, takich jak silniki, cyklotrony, telewizja i radar.

Pola elektryczne i magnetyczne są ze sobą sprzężone i, występując razem, charakteryzują się ciągłą zmiennością lokalnych wartości w przestrzeni.

Zagadnienie obrazu przedstawmy również w kategoriach rachunkowych. W tym celu rozpatrzymy równania Maxwella w postaci:

$$\frac{\partial H}{\partial t} + rotE = 0, \quad \frac{\partial E}{\partial t} - rotH = j \tag{1}$$

gdzie wektory E, H są natężeniami pól elektrycznego i magnetycznego, j jest gęstością prądu.

Po przejściu do równań hiperbolicznych odpowiadających układowi równań (1), dostaniemy (2):

$$\frac{\partial^2 H}{\partial t^2} - \Delta H = rotj, \quad \frac{\partial^2 E}{\partial t^2} - \Delta H = -\nabla divE + \frac{\partial j}{\partial t}$$

Ponieważ fala elektromagnetyczna jest drganiem poprzecznym, zatem:

$$E_{x_3} = 0, \quad H_{x_3} = 0, \quad x_3$$

przyjmujemy za kierunek rozchodzenia się fali.

Rozważmy falę monochromatyczną. Załóżmy, że własności ośrodka grawitacyjnego nie zależą od kierunku, wobec tego możemy poszukiwać rozwiązań układu równań (2) w postaci funkcji:

$$E_{x_1} = f(x_1, x_2) + \sin(x_3 - t + \varphi_1) \tag{3a}$$

$$E_{x_2} = f(x_1, x_2) + \sin(x_3 - t + \varphi_2) \tag{3b}$$

$$H_{x_1} = g(x_1, x_2) + \sin(x_3 - t + \phi_1) \tag{3c}$$

$$H_{x_2} = g(x_1, x_2) + \sin(x_3 - t + \phi_2) \tag{3d}$$

Przedyskutujmy zapisane pola.

Funkcje f i g są identyczne dla dwu współrzędnych przestrzennych z racji założonej anizotropowości ośrodka grawitacyjnego, amplitudy drgań zależą od skali zjawiska, dlatego uznaliśmy je za jednostkowe. Podobnie zagadnienie propagacji zależy od skali problemu, wobec czego przyjęliśmy jednostkową prędkość impulsu. Funkcje sinus są typowymi rozwiązaniami problemów równań Maxwella, dlatego użyliśmy ich właśnie do opisu zjawiska.

Dodamy, że układ sześciu równań (2) posiada dziewięć niewiadomych, dlatego w ostatecznym wyniku niektóre z nich będą zależeć od innych w granicach pozytywnej odpowiedzi na postawione pytanie.

Otrzymujemy w ten sposób układ równań różniczkowych w postaci:

$$\frac{\partial^2 E_i}{\partial t^2} - \Delta E_i = -\frac{\partial}{\partial x_i}\left(\sum_{k=1}^{2} \frac{\partial E_k}{\partial x_k}\right) + \frac{\partial j_i}{\partial t}$$

$$\frac{\partial^2 H_i}{\partial t^2} - \Delta H_i = -rot_i j \text{, i = 1,2,3} \tag{4}$$

Po podstawieniu kształtu funkcji (3) do równań (4) dostaniemy kolejny układ równań różniczkowych:

$$-\Delta f = -\frac{\partial}{\partial x_1}\left(\frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2}\right) + \frac{\partial j_1}{\partial t} \tag{5a}$$

$$-\Delta f = -\frac{\partial}{\partial x_2}\left(\frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2}\right) + \frac{\partial j_2}{\partial t} \tag{5b}$$

$$\frac{\partial j_2}{\partial x_1} = \frac{\partial j_1}{\partial x_2} \tag{5c}$$

$$\frac{\partial j_3}{\partial t} = 0 \tag{5d}$$

$$\Delta g = \frac{\partial j_3}{\partial x_2} - \frac{\partial j_2}{\partial x_3} \tag{5e}$$

$$\Delta g = \frac{\partial j_3}{\partial x_1} - \frac{\partial j_1}{\partial x_3} \tag{5f}$$

Przeliczanie ostatnich równań jest w tym momencie zbyteczne, dlatego, że znaleźliśmy ideę przenoszenia obrazu przez promieniowanie elektromagnetyczne. Zauważyliśmy, że z dziewięciu równań ubyły dwie niewiadome: natężenia pól elektrycznego i magnetycznego w kierunku rozchodzenia się impulsu (są zerami z racji przyjętego kształtu promieniowania poprzecznego).

Natomiast istotny jest fakt prowadzenia przez promieniowanie prądu w kierunku osi rozchodzenia się promieniowania j_3.

Następnie musimy zauważyć (z równań (5e) oraz (5f)), że wspomniany wyżej prąd jest również funkcją dwu zmiennych niezależnych (x_1, x_2).

Dalej zauważymy (z równań (5a),(5b),(5c)), że prądy j_1, j_2 nie są zerami, ale współrzędnymi również zmiennych (x_1, x_2).

A wobec tego:

Po pierwsze wektor gęstości prądu *j* niesie w sobie obraz nadajnika (w płaszczyźnie prostopadłej do linii, rozchodzenia się impulsu).

Po drugie prąd o zmiennych na płaszczyźnie kierunkach w zależności od gradientu modułów jego składowych jest w stanie z dużą rozdzielczością indukować potencjał na odbiorniku. Obraz tej mapy potencjałów może zostać skierowany już w sposób inny do ośrodka deszyfrującego.

Widzimy więc, że transport obrazu odbywa się w analizowanym przypadku inaczej niż dla nadajników radiowych i telewizyjnych, czyli bez zakłóceń na fali nośnej.

3. Model kompozytowego włókna nerwowego

Inżynieria materiałowa jest w tej chwili jedyną dziedziną rozwijającą intensywnie obszar pojęć fizykalnych znajdujących zastosowanie w budowie koncepcji naruszających pewien porządek semantycznego i aksjomatycznego opisu materii [6]. Za pomocą technik inżynierii materiałowej można bowiem wytworzyć takie materiały, które samoistnie nie występują w przyrodzie. Dla naszych celów podjęto próbę opracowania modelu sztucznego włókna nerwowego w formie kompozytu opartego na tzw. zmodyfikowanym prekursorze polimerowym [7]. W niniejszym artykule ze względu na opracowania patentowe nie zostaną podane informacje o rodzaju zastosowanych materiałów dla wytworzenia opisywanego włókna. Przy wyborze materiałów służących do jego wytworzenia postawiono jednak podstawowy i najważniejszy warunek, że to sztuczne włókno – na podobieństwo włókna nerwowego – można będzie stymulować zarówno prądami elektrycznymi, jak i polem magnetycznym, a więc musi ono przewodzić impulsy elektryczne i posiadać mierzalną podatność magnetyczną. Takie warunki produkcyjne pozwolą na wytworzenie rzeczywistego materiału kompozytowego, który będzie poddawany testom zmiennych pól elektrycznych i magnetycznych zgodnie z wynikami analiz teoretycznych i płynącymi z nich wnioskami praktycznymi. Naturalne włókno nerwowe jest bowiem przykładem wspaniałego żywego kompozytu, jaki zdolna była wytworzyć natura, służącego do przesyłania informacji zarówno w środowisku wewnętrznym organizmu, jak i komunikowania się ze światem zewnętrznym dla osiągnięcia podstawowych czynności fizjologicznych organizmów żywych, o czym wspomniano we wstępnie.

W Katedrze Biomateriałów WIMiC AGH zapoczątkowano badania nad wytworzeniem modyfikowanych nanocząstkami włókien polimerowych i otrzymano grant badawczy KBN.

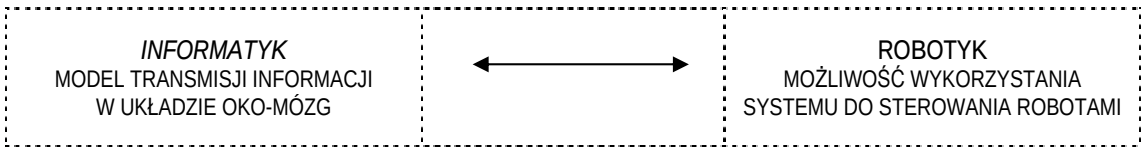
4. Podsumowanie

Analiza formalna wykazała, że obraz w postaci zapisu zmian elektromagnetycznego pola pojawia się w płaszczyźnie prostopadłej do kierunku rozchodzenia się drgań i może być w pewnej formie transportowany wzdłuż przewodzącego włókna. Okazuje się, że ten mechanizm przesyłania obrazu jest najprawdopodobniej taki sam lub bardzo podobny do tego, jaki odbywa się w neuronie. Dzięki mechanizmowi tzw. pompy sodowo-potasowej (metabolicznej), polegającemu na przepompowaniu przez kanały sodowe i potasowe tych jonów w kierunku do środka włókna i na zewnątrz, a więc w płaszczyznach prostopadłych do osi włókna, realizowany jest proces przesyłania sygnału nerwowego. Sygnał nerwowy (impuls) ma postać obszaru o odwróconym znaku ładunku, poruszającego się wzdłuż włókna. Tuż przed sygnałem otwierają się bowiem kanały sodowe, wpuszczając do wnętrza jony sodowe, a tuż za sygnałem kanały potasowe wypuszczają na zewnątrz jony potasowe. Po przejściu sygnału, pompy metaboliczne przywracają stan pierwotny.

Program badań kompozytowych włókien polimerowych, nawiązując do mechanizmu przesyłania sygnału włóknami nerwowymi, zakłada między innymi pomiary rozkładu wybranych wielkości parametrów elektromagnetycznych pól w przestrzeni matematycznego rozwiązania uformowanych włókien polimerowych.

Literatura cytowana

1. Carrel A.: Człowiek istota nieznana, Wydawnictwo Biblioteka Wiedzy, Trzaska, Evert i Michalski S.A. Warszawa 1949.
2. Dewey J.: Jak myślimy, PWM, Warszawa 1988.
3. Kastory B.: Zmysłowe bramy umysłu, Newsweek, 9: 46, 2005.
4. Traczyk W. Z.: Fizjologia człowieka w zarysie, Państwowy Zakład Wydawnictw Lekarskich, Warszawa 1982.
5. Halliday D., Resnik R.: Fizyka dla studentów nauk przyrodniczych i technicznych, PWN, Warszawa 1974.
6. Grabski M. A, Kozubowski J. A.: Inżynieria materiałowa, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2003.
7. Śledziona J.: Podstawy technologii kompozytów, Wydawnictwo Politechniki Śląskiej, Gliwice 1998.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 277-280.

OPTYMALIZACJA KODU MONTE CARLO DO SYMULACJI CZASU PRZELOTU FOTONÓW W STRUKTURACH TKANKOWYCH

OPTIMIZATION OF MONTE CARLO CODE FOR SIMULATIONS OF PHOTON FLIGHT TIMES

NORBERT S. ŻOŁEK, ADAM LIEBERT, ROMAN MANIEWSKI

Instytut Biocybernetyki i Inżynierii Biomedycznej PAN, Warszawa, norbert@ibib.waw.pl

Streszczenie. Symulacje Monte Carlo są często wykorzystywane w modelowaniu transportu światła w strukturach optycznie mętnych. Pozwalają uzyskać dokładny opis rozchodzenia się światła, ale wymagają dużej mocy obliczeniowej. Najbardziej czasochłonnymi operacjami podczas symulacji są: obliczanie wartości funkcji logarytmicznej i trygonometrycznej oraz generowanie liczb pseudolosowych. W tym artykule została przedstawiona weryfikacja i porównanie metody Monte Carlo wykorzystującej aproksymację funkcji logarytmicznej i trygonometrycznej z metodą klasyczną (z dokładnym obliczaniem logarytmu i funkcji cosinus) i z rozwiązaniami równania dyfuzji dla struktur jednorodnych.

Słowa kluczowe: metoda Monte Carlo, transport światła, struktury warstwowe

Abstract. Monte Carlo simulations are frequently used in modeling of light transport in turbid media. They allow precise description of light transport, but require long computation time. Logarithmic and trigonometric functions evaluation, and generation of pseudorandom numbers are main – time consuming procedures in Monte Carlo calculations of light transport. In this paper, parallel Monte Carlo method with approximated logarithmic and trigonometric functions is analyzed, verified, and compared to the original one (with exact computation of logarithm and cosine functions) and with solution of diffusion equation for homogeneous media.

Key words: Monte Carlo method, light transport, layered structures

1. Wstęp

Nowoczesne metody badań właściwości optycznych tkanek wielowarstwowych (np. mózg i tkanki okołomózgowe: skóra, kość czaszki, płyn mózgowo-rdzeniowy, mózg) oparte na czasie przelotu fotonów wymagają informacji o drogach optycznych w poszczególnych warstwach badanej struktury [1]. Metoda Monte Carlo [2], [3], [4], [5], [6] jest często wykorzystywana do teoretycznego opisu tych wielkości, gdyż oferuje uniwersalne podejście do prognozowania transportu fotonów w ośrodkach optycznie mętnych. Wykorzystuje makroskopowe parametry charakteryzujące badany ośrodek i może być użyta dla złożonych geometrii i bez ograniczeń na właściwości optyczne ośrodka, w przeciwieństwie np. do dyfuzyjnego równania transportu radiacyjnego [7], [8]. Niestety, do uzyskania statystycznie wiarygodnych wyników należy śledzić trajektorie dużej liczby fotonów. W związku z tym metoda wymaga dużych mocy obliczeniowych, a stąd konieczność optymalizacji kodu do symulacji.

2. Symulacje Monte Carlo

Modelowanie transportu światła przez tkankę metodą Monte Carlo opiera się na symulowaniu ruchu pakietów fotonów emitowanych do wnętrza tkanki. Pakiet fotonów przemieszcza się wzdłuż prostoliniowych odcinków losowej długości aż do momentu zaabsorbowania wewnątrz struktury lub ostatecznego opuszczenia badanego ośrodka.

Pakiety fotonów są emitowane z punktu emisji w kierunku prostopadłym do powierzchni badanego ośrodka. Długość prostoliniowego odcinka trajektorii jest określana przez zmienną losową i właściwości optyczne ośrodka według zależności:

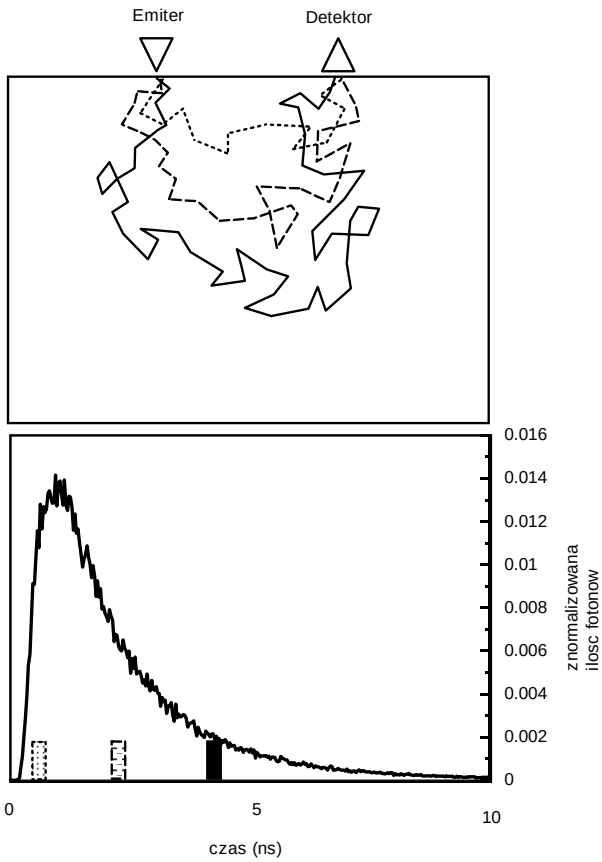
$$s = \frac{-\ln(\xi)}{\mu_a + \mu_s^{'}}, \qquad (1)$$

Gdzie: x jest zmienną losową o rozkładzie jednostajnym w przedziale $(0;1)$, μ_a – współczynnik absorpcji, $\mu_s^{'}$ – zredukowany współczynnik rozpraszania. Rozpatrywany pakiet fotonów po przebyciu tego odcinka ulega kolejnemu rozproszeniu. Założono przy tym izotropowe rozpraszanie

pakietów fotonów, co oznacza, że nie ma żadnych preferencji co do wyboru nowego kierunku pakietu po rozproszeniu i dowolny kierunek może zostać wybrany z jednakowym prawdopodobieństwem. Waga pakietu fotonów w momencie emisji ma wartość równą 1. Po każdym *i*-tym rozproszeniu jego waga zmniejsza się, przyjmując wartość równą:

$$w_{i+1} = w_i * \left(\mu_a / (\mu_a + \mu_s') \right). \tag{2}$$

W sytuacji, gdy droga fotonów znajduje się w ośrodkach o różnych właściwościach optycznych, uwzględniane są także zjawiska załamania i odbicia światła na granicach ośrodków. Odpowiednio modyfikowana jest także długość pokonywanego odcinka w zależności od właściwości optycznych.



Rys. 1. Trajektorie fotonów penetrujących na różne głębokości oraz histogram czasu przelotu fotonów od punktu emisji do punktu detekcji

Symulacja ruchu pakietu fotonów jest kontynuowana dopóki „śledzony" pakiet ma wagę (2) powyżej wartości progowej (założona przez nas wartość progowa jest równa 1e-20), poniżej której uznaje się go za całkowicie zaabsorbowany. Taka procedura zabezpiecza przed długotrwałym śledzeniem pakietów fotonów, które nie wpłyną istotnie na wynik symulacji.

Dla pakietów fotonów opuszczających ośrodek zbierana jest informacja o czasie przelotu pomiędzy punktem emisji i detekcji i tworzony jest histogram czasu przelotu fotonów (rys. 1).

3. Optymalizacja kodu

W kilku ośrodkach naukowych opracowano ogólnodostępne kody Monte Carlo do symulacji transportu światła w tkankach [3], [4], [9]. Do analizy użyto kodu źródłowego programu MCML [3], który został uzupełniony o zbieranie informacji o czasie przelotu poszczególnych pakietów fotonów pomiędzy punktem emisji i detekcji. Przeprowadzone pomiary czasu wykonania poszczególnych fragmentów kodu wskazują, że najbardziej czasochłonnymi operacjami programu MCML są: procedura obliczająca długość swobodnej drogi pakietu fotonów do punktu kolejnego rozproszenia (1) (33% czasu działania całego programu) oraz funkcja obliczająca nowy kierunek pakietu fotonów po rozproszeniu (także 33% czasu działania całego programu). Najbardziej czasochłonnymi operacjami wewnątrz tych funkcji są: generowanie liczb losowych, obliczanie logarytmu naturalnego i obliczanie wartości funkcji cosinus przy ustalaniu kierunku pakietu fotonów po zajściu zjawiska rozproszenia. Funkcja cosinus została zastąpiona wielomianem (4) dla argumentów z zakresu [0;π/2). Zastosowany algorytm liczb losowych „ranmar" [6] pozwolił na aproksymację funkcji logarytm funkcją wymierną (5) dla argumentów z zakresu [5.9*e*–8;1). Zastosowana aproksymacja odbiega od dokładnych wartości nie więcej niż 1% wartości dopasowanej funkcji wymiernej w dowolnym punkcie przedziału określoności.

$$\cos(x) \approx \begin{cases} a + bx^2 + cx^4 & for \quad x \in [0; \pi/2 - 0.8) \\ \dfrac{\pi}{2} - x & for \quad x \in [\pi/2 - 0.8; \pi/2] \end{cases} \tag{4}$$

Tab. 1. Parametry aproksymacji funkcji cosinus wielomianem (4)

parameter	$y = a + bx^2 + cx^4$ $x \in [0; \pi/2 - 0.8]$
a	9.9e-1
b	-4.9e-1
c	3.7e-2

$$\ln(x) \approx \frac{a + cx^{0.5} + ex}{1 + bx^{0.5} + dx + fx^{1.5}} \tag{5}$$

Tab. 2. Parametry aproksymacji funkcji logarytmicznej funkcją wymierną (5)

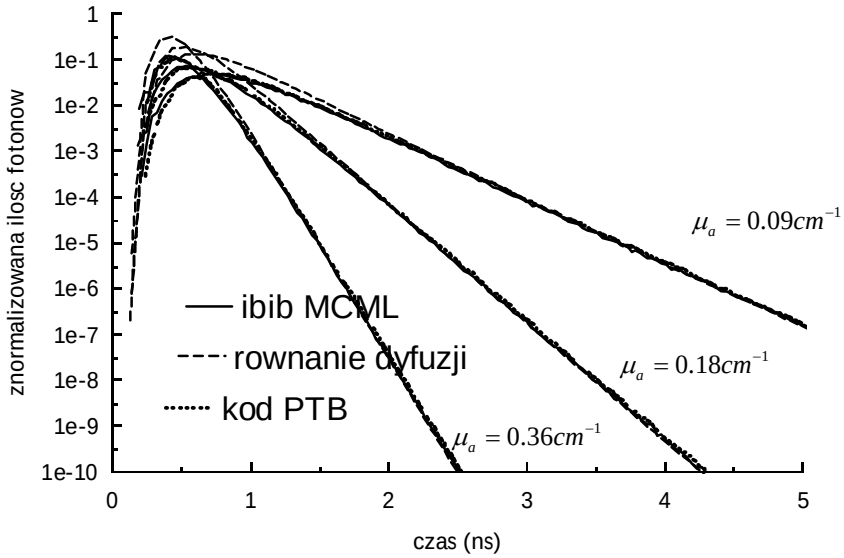
parameter	$x \in [5.5e-8; 0.01)$	$x \in [0.01; 1)$
a	-19	-8.8e0
b	1.5e3	1.5e1
c	-1.7e4	-4.2e1
d	9.3e5	1.5e1
e	-4.8e6	5.1e1
f	3.2e6	-1.9e0

Maksymalna wartość różnicy tych aproksymacji wyniosła 1% wartości funkcji w danym punkcie.

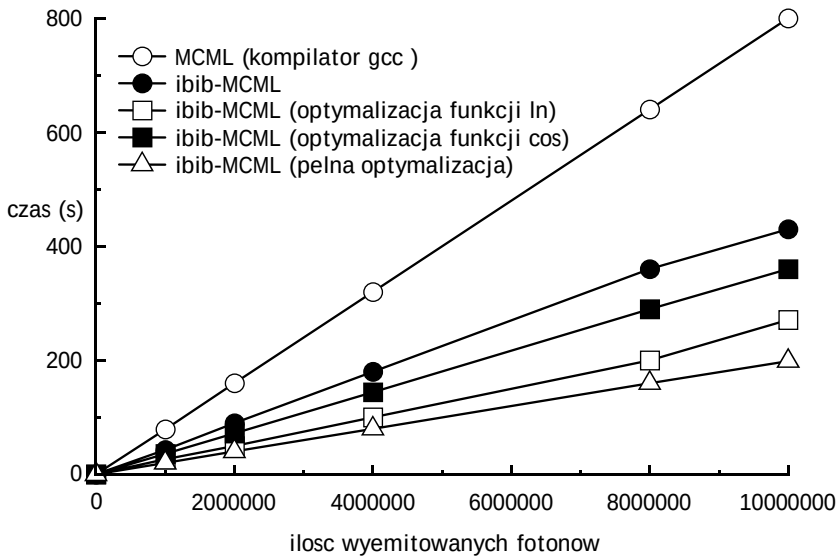
4. Wyniki

Symulacje były przeprowadzane dla obiektu o strukturze jednorodnej półprzestrzeni o parametrach optycznych zbliżonych do parametrów optycznych skóry: współczynnik absorpcji $\mu_a = 0.18cm^{-1}$, zredukowany współczynnik rozpraszania $\mu_s^{'} = 19cm^{-1}$, współczynnik załamania na granicy ośrodków n=1.4. Odległość punktu emisji i detekcji wynosiła 3 cm.

Porównanie wyników otrzymanych przy użyciu kodu MCML i kodu z aproksymacją oraz rozwiązania równania dyfuzji, zostało przedstawione na rys. 2. Mimo zastosowania aproksymacji funkcji logarytmicznej i funkcji trygonometrycznej, wyniki otrzymane opracowaną przez nas wersją kodu nie odbiegają od wyników otrzymanych zmodyfikowanym kodem MCML (rys. 2).



Rys. 2. Porównanie rozkładów czasu przelotu fotonów otrzymanych przy użyciu: równania dyfuzji (6) – linia przerywana, opracowanej zoptymalizowanej wersji kodu Monte Carlo – linia ciągła oraz wyniki symulacji przy użyciu kodu opracowanego w Physikalisch Technische Bunsesanstalt (Berlin) [1]. Parametry symulacji $\mu_s = 19cm^{-1}$, $\mu_a = 0.09, 0.18, 0.36cm^{-1}$



Rys. 3. Porównanie szybkości oryginalnego kodu MCML z szybkością kodu opracowanego w ramach zadania na różnych etapach optymalizacji

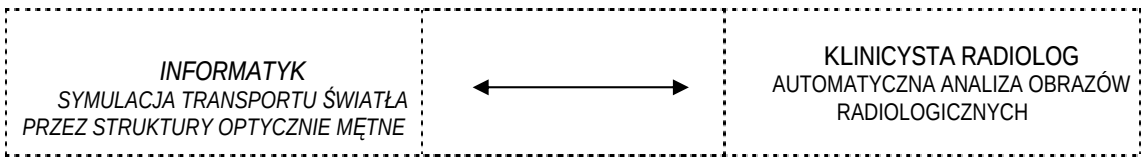
Dalsza optymalizacja polegała na stworzeniu wersji równoległej opisywanego kodu, która dzięki zastosowanemu algorytmowi generowania liczb pseudolosowych („ranmar" [6]) pozwalającemu na uzyskanie niezależnych ciągów liczb pseudolosowych dla różnych wartości inicjujących dała liniowe przyspieszenie w stosunku do kodu jednoprocesorowego. Testy były przeprowadzane na komputerze klasy PC oraz SUN Enterprise 10000.

4. Podsumowanie

Opracowany kod jest 4-krotnie szybszy od wyjściowego kodu MCML. Uwzględnienie w kodzie wykorzystania maszyn wieloprocesorowych pozwala dodatkowo przyspieszyć symulację wprost proporcjonalnie do liczby procesorów. Niedokładności pomiędzy rozwiązaniami dla symulacji z dokładnymi funkcjami logarytm i cosinus są tak niewielkie, że z punktu widzenia zastosowania w analizie rozkładów czasów przelotu fotonów nie wpływają znacząco na wynik.

Literatura cytowana

1. Liebert A. et al.: Time-resolved multidistance near-infrared spectroscopy of the adult head: intracerebral and extracerebral absorption changes from moments of distribution of times of flight of photons, Applied Optics, 43(15): p. 3037-3047, 2004.
2. Liebert A. et al.: Evaluation of optical properties of highly scattering media by moments of distributions of times of flight of photons, Applied Optics, 42(28): p. 5785-5792, 2003.
3. Wang L., Jacques S. L., Zheng l.: MCML - Monte Carlo modeling of light transport in multi-layered tissues, Computer Methods & Programs in Biomedicine, 47: p. 131-146, 1995.
4. Boas D. A. et al.: Three dimensional Monte Carlo code for photon migration through complex heterogeneous media including the adult human head, Optics Express, 10(3): p. 159-170, 2002.
5. Steinbrink J. et al.: Determining changes in NIR absorption using a layered model of the human head. Phys. Med. Biol., 46: p. 879-896, 2001.
6. Vattulainen, I., et al., A Comprehesive Study Of Some Pseudorandom Number Generators. Comput. Phys. Commun., 86: p. 209-226, 1995.
7. Okada E. et al.: Theoretical and experimental investigation of near-infrared light propagation in a model of the adult head. Appl. Opt., 36: p. 21-31, 1997.
8. Hielscher A. H., Alcouffe R. E., Barbour R. L.: Comparison of finite-difference transport and diffusion calculations for photon migration in homogeneous and heterogeneous tissues, Phys. Med. Biol., Vol. 43, 1285-1302, 1998.
9. Kienle A., Patterson M. S.: Determination of the optical properties of turbid media from single Monte Carlo simulation, Phys. Med. Biol., 41: p. 2221-2227, 1996.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 281-286.

POLA ELEKTROMAGNETYCZNE W BADANIACH BIOMEDYCZNYCH – UKŁADY EKSPOZYCYJNE I MODELOWANIE NUMERYCZNE

ELECTROMAGNETIC FIELDS EXPOSURE IN BIOMEDICAL INVESTIGATIONS

PAWEŁ BIEŃKOWSKI, TOMASZ DŁUGOSZ

Pracownia Ochrony Środowiska Elektromagnetycznego, Instytut Telekomunikacji i Akustyki, Politechnika Wrocławska, pawel.bienkowski@pwr.wroc.pl, tomasz.dlugosz@pwr.wroc.pl

Streszczenie. W pracy przedstawiono zarys rozwiązań układowych pozwalających na ekspozycję w polu elektromagnetycznym obiektów biologicznych. Zaprezentowano typowe układy stosowane w różnych zakresach częstotliwości wraz z podstawowymi metodami wyznaczania natężenia pola. Zwrócono uwagę na techniczne ograniczenia stosowalności poszczególnych układów i możliwość korelacji wyników uzyskiwanych w zamkniętych układach ekspozycyjnych z opromieniowaniem w swobodnej przestrzeni.

Słowa kluczowe: badania biomedyczne, metody numeryczne, pole elektromagnetyczne, układy ekspozycyjne

Abstract: The paper presents a review of exposure systems used in research of interaction between electromagnetic field (EMF) and living organisms. Typical systems ranging in frequency from static field to microwaves and simple methods of exposure estimation and basic limitations are described. This paper also deals with the problem of susceptibility investigation of objects on influence of electromagnetic field. The authors of numerous publications deal with the influence of the object upon the electromagnetic field distribution inside a waveguide (e.g. a TEM cell) while inverse effect: influence of line on the object has been overlooked. The problem plays essential role in chance of correlation of investigations carried out in closed space (TEM, GTEM, waveguide) and in the free space.

Keywords: biomedical investigations, numerical methods, electromagnetic fields, exposure systems

1. Wstęp

Oddziaływanie pola elektromagnetycznego (PEM) na materię ożywioną jest przedmiotem zainteresowania badaczy od kilkudziesięciu lat i badania takie są wciąż rozwijane wraz ze wzrostem nasycenia środowiska źródłami PEM. Równolegle z rozwojem badań rośnie również zapotrzebowanie na nowe układy ekspozycyjne, pozwalające odzwierciedlać specyficzne warunki ekspozycji w polach pochodzących od konkretnych urządzeń, np. telefonów komórkowych [2], [3]. Dotychczasowe badania skupiają się z jednej strony na dobroczynnych oddziaływaniach PEM na organizmy (np. fizjoterapia, diagnostyka), a z drugiej na określeniu ewentualnych szkodliwych skutków opromieniowania PEM. Badania laboratoryjne są dopełnieniem badań epidemiologicznych grup narażonych na ekspozycję na PEM czy to w środowisku (np. mieszkańcy okolicy linii wysokiego napięcia), czy na stanowiskach pracy (np. personel stacji radiolokacyjnych czy obsługa zgrzewarki w.cz.). Ponieważ są to badania interdyscyplinarne, gdzie splatają się nauki przyrodnicze (biologia, medycyna) z techniką (wytwarzanie i metrologia pola elektromagnetycznego), autorzy jako inżynierowie pragną zwrócić uwagę właśnie na aspekt techniczny tych prac. W większości przypadków badań prowadzonych na obiektach biologicznych mamy do czynienia z pracami o charakterze podstawowym i niezmiernie istotną informacją jest korelacja badań prowadzonych w warunkach laboratoryjnych z oddziaływaniem PEM na obiekt w warunkach rzeczywistych – zwykle zbliżonych do warunków swobodnej przestrzeni. Nie mniej istotna (a z punktu widzenia eksperymentu naukowego chyba najistotniejsza) jest powtarzalność eksperymentów, co z technicznego punktu widzenia przekłada się na zapewnienie identycznych warunków opromieniowania. Nie znaczy to, że konieczne jest zastosowanie identycznego układu ekspozycyjnego, ale nie do pominięcia jest wzajemne oddziaływanie źródła pola i obiektu badanego. Zjawisko to występuje nie tylko w eksperymentach biomedycznych, ale także w układach wzorców pola elektromagnetycznego i przy badaniach z zakresu kompatybilności elektromagnetycznej. Drugim ograniczeniem w stosowaniu układów ekspozycyjnych jest wielkość obszaru

jednorodnego pola. Ma to największe znaczenie w układach wielokrotnych.

2. Badania biomedyczne

Eksperymenty biomedyczne prowadzone są w dwóch zasadniczych kierunkach:

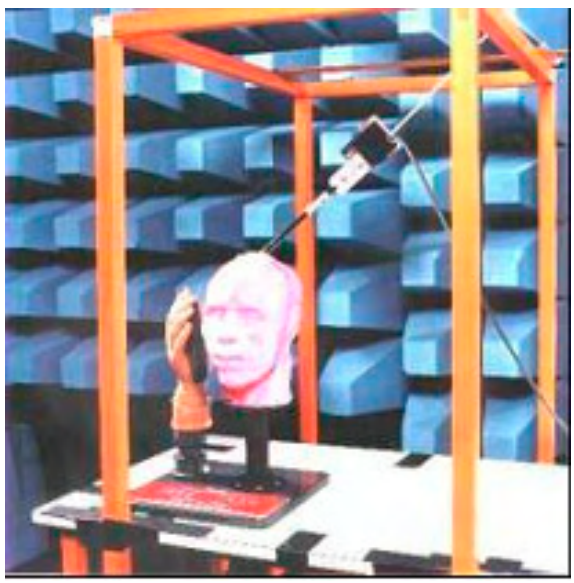
- badanie wpływu PEM na metabolizm organizmów żywych i takowe prowadzone są głównie na roślinach i zwierzętach z wykorzystaniem układów ekspozycyjnych, których budowa zależy od zakresu częstotliwości,
- wyznaczanie rozkładu PEM i pochłaniania energii elektromagnetycznej przez struktury biologiczne.

Parametrem wykorzystywanym w tych badaniach jest SAR (ang. *Specific Absorbtion Rate*). Ze względu na trudności w bezpośrednim pomiarze SAR, bardzo popularne jest wykorzystywanie „dozymetrii numerycznej", czyli wyznaczania SAR na podstawie znajomości parametrów tkanek i natężenia PEM oświetlającego obiekt (1).

$$SAR = \sigma \frac{E^2}{\rho}, \tag{1}$$

gdzie:

ρ – gęstość właściwa tkanki,
σ – przewodność tkanki,
E – wartość skuteczna natężenia pola elektrycznego.



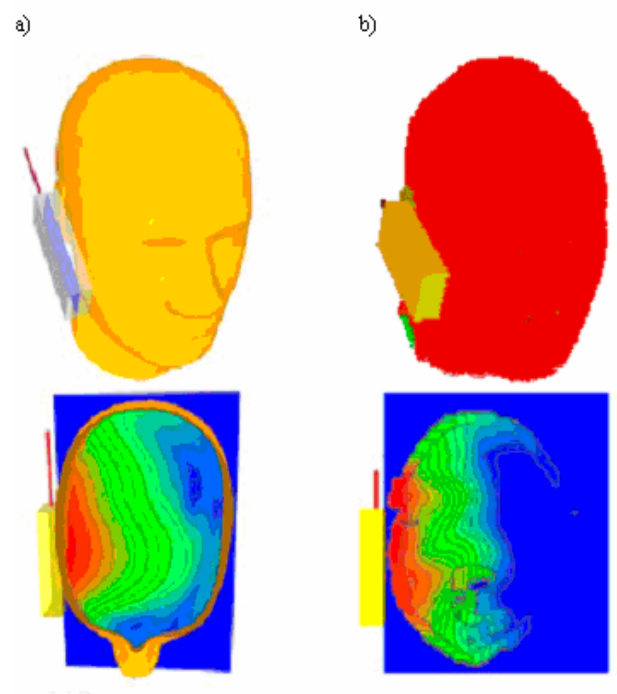
Rys. 1. Fantom do pomiaru SAR (Źródło: [8])

W badaniach rzeczywistego rozkładu PEM wewnątrz obiektu stosuje się zwykle jednorodny fantom wypełniony substancją o parametrach odpowiadających średnim parametrom badanej tkanki. Przykład stanowiska do takich badań przedstawia rys. 1 [7]. Pomiar natężenia pola wykonuje się przy użyciu miniaturowego czujnika PEM. Prowadzenie badań eksperymentalnych na fantomach coraz częściej jest ograniczane do testów sprawdzających dla symulacji komputerowych (rys. 2). Dla celów analizy numerycznej tworzone są nawet standardowe modele obiektów biologicznych. Przykładem może być anatomiczny model człowieka zamieszczony w medycznej bazie Visible Human Dataset udostępnionej przez National Library of Medicine [7]. We wspomnianej bazie zamieszczone są cyfrowe przekroje anatomiczne ciała dorosłego człowieka, a także odpowiadające im rekonstrukcje CT (ang. *Computer Tomography*) i MRI (ang. *Magnetic Resonance Imaging)*. Modele zawierają również informacje o elektrycznych właściwościach poszczególnych tkanek. Symulacje komputerowe nie zastąpią jednak badań eksperymentalnych w zakresie reakcji organizmów na PEM. Tu nieodzowne okazują się układy ekspozycyjne i żywe obiekty biologiczne.

3. Układy ekspozycyjne pola stałego i niskich częstotliwości

W zakresie niskich częstotliwości możliwe, a nawet konieczne jest niezależne rozpatrywanie składowej elektrycznej i magnetycznej PEM. Pola magnetyczne wytwarza się zwykle wewnątrz cewek w różnych konfiguracjach. Typowym źródłem pola elektrycznego są kondensatory płaskie.

Najprostszym układem ekspozycyjnym pola magnetycznego jest pętla z prądem – pojedyncza antena ramowa. Natężenie pola magnetycznego w osi prostopadłej do powierzchni anteny można wyznaczyć na podstawie zależności (2)



Rys. 2. Przykładowy rozkład SAR: a) zmierzony eksperymentalnie za pomocą fantomu, b) uzyskany drogą symulacji komputerowej (Źródło: [5])

$$H_{ANTENA\ RAMOWA} = \frac{Inr^2}{2(r^2 + x^2)^{\frac{3}{2}}} \text{ [A/m]} \tag{2}$$

gdzie:

I – prąd w antenie [A],
n – ilość zwojów anteny,
r – promień anteny [m],
x – odległość od środka anteny [m].

Znacznie większy obszar o równomiernym rozkładzie pola można wytworzyć wewnątrz powietrznego solenoidu, dla którego poprawna jest zależność (3):

$$H_{SOLENOID} = \frac{In}{\sqrt{l^2 + 4r^2}} \text{ [A/m]}, \qquad (3)$$

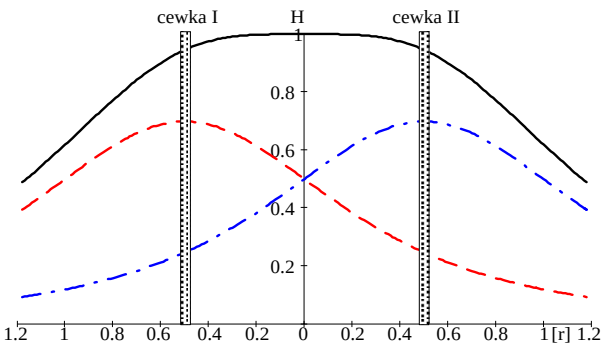
gdzie:

l - długość solenoidu w metrach.

Wadą solenoidu jest utrudniony dostęp do obszaru roboczego. Wady tej pozbawione są cewki Helmholtza (rys. 3).

Idealne cewki Helmholtza tworzą dwie identyczne anteny ramowe z jednakowym kierunkiem płynącego w nich prądu, umieszczone osiowo względem siebie w odległości odpowiadającej promieniowi anteny. Efekt równomiernego rozkładu pola wewnątrz cewek uzyskuje się dzięki sumowaniu składowych pola równoległych do osi cewek i odejmowaniu składowych prostopadłych do tej osi. Unormowane natężenie pola w osi cewek w funkcji odległości x od środka geometrycznego przedstawione jest na wykresie (rys. 4). Linie przerywane obrazują pola pojedynczych cewek, a linia ciągła pole wypadkowe.

Rys. 3. Cewki Helmholtza jako układ ekspozycyjnego pola magnetycznego



Rys. 4. Unormowane natężenie pola wewnątrz cewek Helmholtza

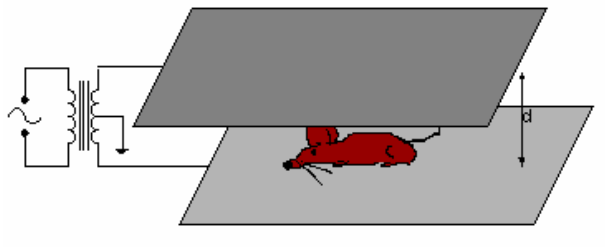
Pole wewnątrz cewek skierowane jest prostopadle do płaszczyzny anten i charakteryzuje się bardzo równomiernym rozkładem. W obszarze do 0.6 r od środka układu natężenie pola zmienia nie więcej niż 10% [6].

W kondensatorze płaskim (rys. 5) natężenie pola elektrycznego w środku przestrzeni między nieskończenie rozległymi okładkami opisane jest prostą zależnością (4):

$$E_{KONDENSATOR} = \frac{U}{d} \text{ [V/m]}, \qquad (4)$$

gdzie:

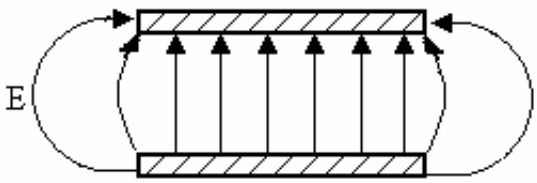
U – napięcie przyłożone go okładek,
d – ogległość między okładkami.



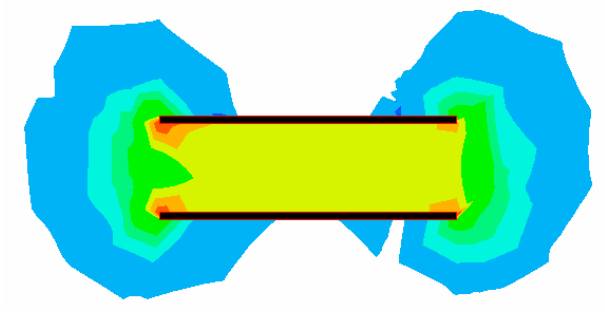
Rys. 5. Układ ekspozycyjny pola elektrycznego z kondensatorem płaskim

W przypadku rzeczywistego kondensatora w pobliżu krawędzi występują deformacje rozkładu pola. Na rys. 6 przedstawiono przebiegi teoretyczne i wyniki symulacji komputerowej.

a)



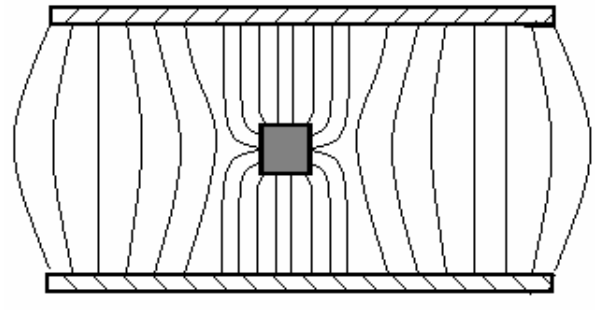
b)



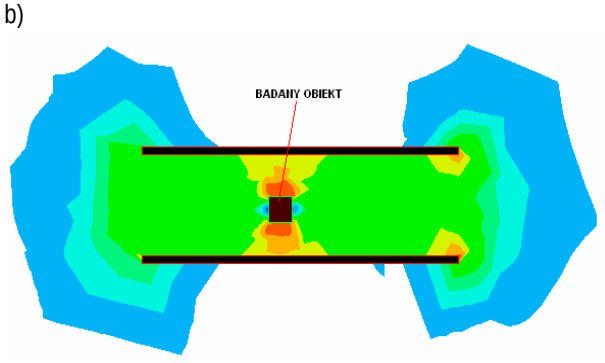
Rys. 6. Deformacja rozkładu PEM na krawędziach kondensatora płaskiego: a) teoretyczne, b) wynik symulacji metodą FEM

Umieszczenie wewnątrz kondensatora badanego obiektu powoduje powstanie dodatkowych zaburzeń PEM. Przykładowy rozkład dla obiektu w postaci przewodzącego sześcianu przedstawia rysunek 7.

a)



b)



Rys. 7. Zaburzenia PEM w kondensatorze płaskim na skutek umieszczenia w nim badanego obiektu: a) teoretyczne, b) wynik symulacji metodą FEM

Dolna okładka kondensatora może być uziemiona (szczególnie w dużych układach ekspozycyjnych) lub kondensator zasila się symetrycznie w celu eliminacji wpływu ziemi i przedmiotów otaczających kondensator. Układy z kondensatorami płaskimi stosować można do częstotliwości kilkuset kHz. Dla częstotliwości wyższych stosuje się układy z dopasowanymi liniami transmisyjnymi, np. linie TEM.

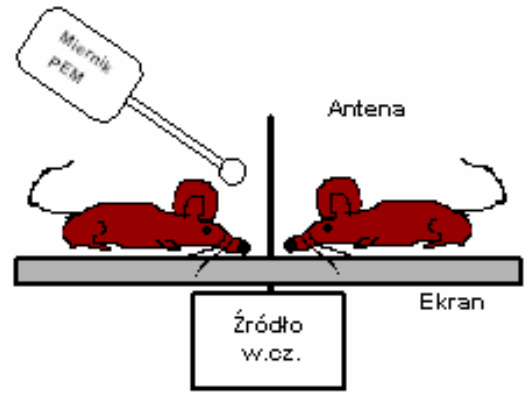
4. Radiofale

W zakresie radiofal (do kilkuset MHz) stosuje się układy ekspozycyjne z antenami liniowymi oraz odcinkami linii transmisyjnych – szczególnie chętnie komorę TEM [3].

W celu uzyskania dużych natężeń PEM obiekt badany umieszcza się w bezpośredniej bliskości anteny, w jej polu bliskim. Przykład takiego stanowiska przedstawia rysunek 8. Wyznaczenie natężenia pola w takim przypadku metodami analitycznymi jest bardzo utrudnione. Wynika to zarówno ze specyfiki pola bliskiego, jak i ze sprzężenia anteny z obiektem badań.

Pozostaje analiza numeryczna oraz jej weryfikacja z użyciem „punktowych" czujników pola.

Zaletą komory TEM jako układu ekspozycyjnego jest jej szerokopasmowość oraz niezależność natężenia pola od częstotliwości i stosunkowo prosty sposób wyznaczenia natężenia pola. Wewnątrz dopasowanej komory TEM natężenie pola wynosi (5):



Rys. 8. Układ ekspozycyjny z anteną liniową

$$E_{TEM} = \frac{U}{d} = \frac{\sqrt{P \cdot Z_0}}{d} \text{ [V/m]}, \tag{5}$$

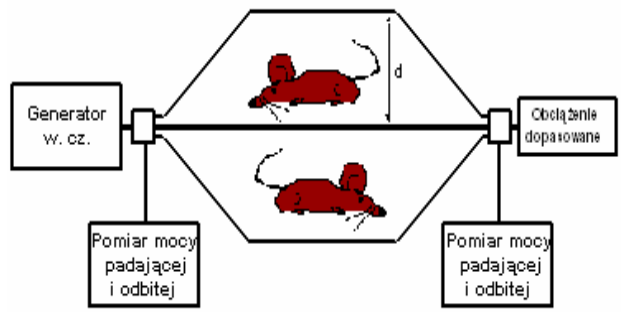
gdzie:

U i P – napięcie lub moc doprowadzona do komory,

Z_0 – dopasowana impedancja obciążenia komory.

Zastosowanie układu jak na rysunku 9 umożliwia eksperymentalne wyznaczenie SAR przez pomiar mocy na obu końcach komory pustej i po umieszczeniu w niej obiektu badanego.

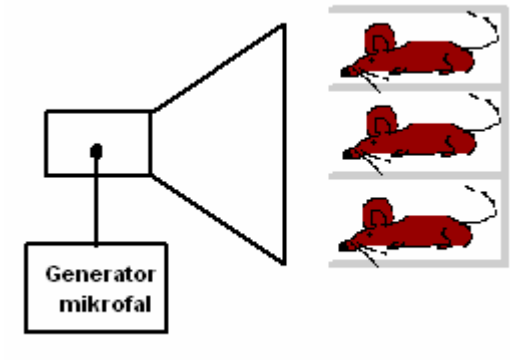
Podstawowymi ograniczeniami w stosowaniu komory TEM są ograniczenie obszaru roboczego związane z górną częstotliwością pracy oraz sprzężenie komory z obiektem badanym. W celu eliminacji wyższych modów (mody falowodowe) poprzeczne wymiary komory muszą być mniejsze od połowy długości fali dla najwyższej używanej częstotliwości, co np. dla zakresu częstotliwości stosowanych w telefonii komórkowej daje maksymalny wymiar poniżej 10 cm. W celu uniknięcia silnego oddziaływania obiektu badanego na rozkład pola, a także minimalizacji sprzężeń, wymiary obiektu badanego nie powinny przekraczać 1/3 wymiarów komory. Szerzej zagadnienie to zostanie przedstawione w dalszej części artykułu.



Rys. 9. Komora TEM – system ekspozycyjny z możliwością pomiaru SAR

5. Mikrofale

W zakresie mikrofal podstawowe źródła PEM stosowane jako układy ekspozycyjne to anteny aperturowe (rys. 10) i odcinki falowodu. W obu rozwiązaniach dużym problemem jest zapewnienie równomiernego rozkładu pola w całym obszarze roboczym oraz wyznaczenie prostymi metodami rzeczywistego poziomu opromieniowania obiektu. Jest to związane głównie ze zmianą impedancji anteny pod wpływem umieszczenia w jej bezpośrednim otoczeniu obiektu badanego, a co za tym idzie z rozstrojeniem układu.

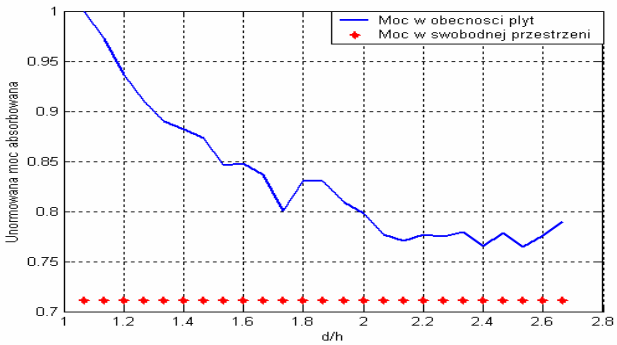


Rys. 10. Układ ekspozycji na mikrofale z anteną tubową

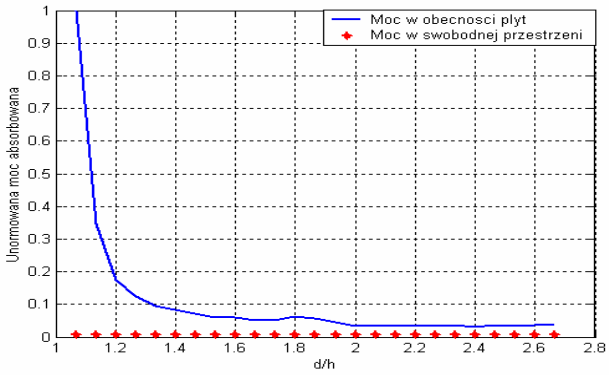
6. Wzajemne oddziaływanie układu ekspozycyjnego i badanego obiektu

Przy analizie wzajemnego oddziaływania obiektu badanego i układu ekspozycyjnego zwykle skupiano uwagę na zaburzeniu rozkładu pola w układzie ekspozycyjnym (zwłaszcza w liniach TEM i falowodach) wywoływanym umieszczeniem w obszarze roboczym obiektu, natomiast pomijano zjawisko odwrotne – wpływ linii na badany obiekt. Zagadnienie to ma istotne znaczenie w przypadku korelacji badań prowadzonych w komorach zamkniętych ze zjawiskami występującymi w swobodnej przestrzeni. Dla ilustracji istotności tego efektu wyznaczono zmianę mocy absorbowanej przez obiekt umieszczony w linii w funkcji stosunku wymiarów obiektu (h) do wymiarów komory (d) – porównaj rys. 9. Krzywe uzyskano dla stałego pobudzenia linii (stałe natężenie PEM) oraz różnych parametrów elektrycznych obiektu (zmienna konduktywność obiektu). Wyniki porównano z mocą, jaką zaabsorbowałby ten sam obiekt umieszczony w polu fali płaskiej o tym samym natężeniu w swobodnej przestrzeni.

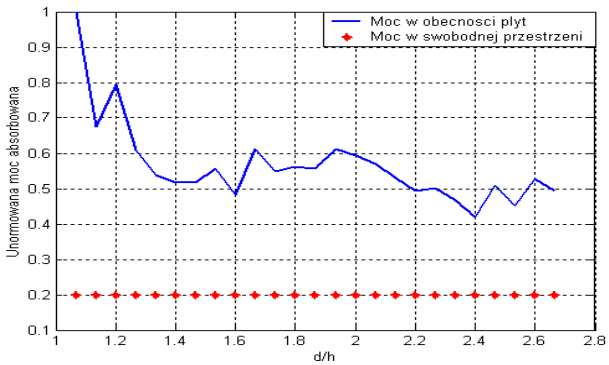
a)



b)



c)



Rys. 11. Porównanie mocy absorbowanej przez obiekt umieszczony między płytami z mocą absorbowaną przez obiekt umieszczony w polu fali płaskiej dla tego samego natężenia pola oświetlającego (f = 10 MHz, E = 1 V/m), konduktywności: a) σ = 1 mS/m, b) σ = 1 S/m, c) σ = 1000 S/m

Wyniki wskazują, że stabilizacja pochłanianej mocy występuje przy odległości między płytami równej co najmniej dwukrotnemu rozmiarowi obiektu.

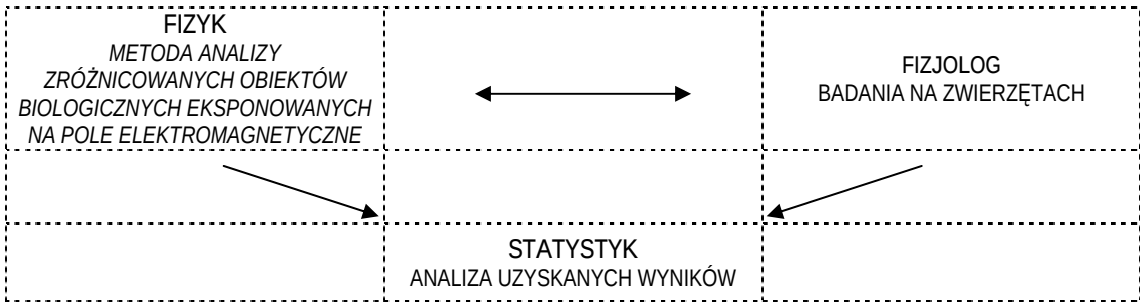
7. Podsumowanie

W pracy przedstawiono techniczny aspekt badań oddziaływania PEM na organizmy żywe. Zaprezentowano podstawowe rodzaje układów ekspozycyjnych, które pozwalają na opromieniowanie badanego obiektu jednym rodzajem pola. Wszystkie przedstawione zależności są w pełni poprawne dla ciągłych sinusoidalnych sygnałów pobudzających. Należy jednak zwrócić uwagę, że w przypadku stosowania sygnałów złożonych, zawierających wiele harmonicznych, pola generowane pod wpływem tego samego sygnału pobudzającego w różnych układach mogą mieć bardzo różne struktury widmowe, a co za tym idzie – warunki opromieniowania będą nieporównywalne. Wynika to ogólnie z różnic w transmitancji poszczególnych układów w szerokim paśmie częstotliwości, znacznie szerszym niż podstawowa częstotliwość sygnału pobudzającego. We wcześniejszych akapitach zwracano uwagę na zapewnienie równomiernego rozkładu pola w obszarze roboczym układu ekspozycyjnego i to zarówno w układzie pustym, jak i po umieszczeniu w nim obiektów badanych. Ma to kapitalne znaczenie przy jednoczesnych

badaniach grupy obiektów (np. kilka zwierząt laboratoryjnych czy też kilkadziesiąt próbek komórek). Nie spełnienie tego warunku może prowadzić do całkowitego zafałszowania wyników eksperymentu. Zaprezentowane tu zagadnienia są tylko wierzchołkiem góry lodowej problemów związanych z techniką wytwarzania i pomiarów wzorcowych pól elektromagnetycznych – technicznej strony badań oddziaływania pola elektromagnetycznego na materię ożywioną.

Literatura cytowana

1. Bieńkowski P.n.e., Trzaska H.: Sposób ekspozycji na pole elektromagnetyczne i urządzenie do ekspozycji, Zgłoszenie patentowe P 357973.
2. Bieńkowski P.: The new GSM-like exposure system, Electromagnetic Compatibility 2004, Seventeenth International Wrocław Symposium and Exhibition, s. 74-75, Wrocław 2004.
3. Crawford M. L.: Generation of Standard EM Field Using TEM Transmission Cells. IEEE Trans. On EMC, Vol. EMC – 16, No. 4, Nov. 1974.
4. Długosz T.: Badanie wrażliwości urządzeń technicznych i obiektów biologicznych, Przegląd Telekomunikacyjny i Wiad. Telekom, 12, 2004.
5. EMPIRE Software, Aplication Note: SAR/ACD Modelling.
6. O'Neill D.: Mathematical Modeling of the Magnetic Field of Helmholtz Coil. (materiał nieopublikowany).
7. http://www.nlm.nih.gov/research/visible/visible_human.html, National Library of Medicine.
8. http://www.npl.com, materiały informacyjne NPL.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 287-290.

KOMPUTEROWE METODY DETEKCJI NOWOTWORÓW PIERSI W ZDJĘCIACH MAMMOGRAFICZNYCH

COMPUTER-BASED METHODS FOR BREAST CANCER DETECTION IN MAMMOGRAMS

WITOLD DZWINEL, KRZYSZTOF BORYCZKO, TOMASZ ARODŹ, MARCIN KURDZIEL

Katedra Informatyki Akademii Górniczo-Hutniczej, Kraków, dzwinel@agh.edu.pl

Streszczenie. Rozpoznawanie obrazów medycznych stanowi jedno z istotnych zastosowań technik informatycznych w medycynie. Przykładem problemu z tej dziedziny jest komputerowe wspomaganie wykrywania nowotworów piersi na zdjęciach rentgenowskich. W niniejszym artykule przedstawiamy rezultaty prac nad komputerowymi metodami wspomagania diagnozy nowotworów piersi.
Prowadzone przez nas prace skupiają się na dwóch podejściach do wykrywania znaczników raka piersi w mammogramach. Pierwsze podejście opiera się na stworzeniu filtru wrażliwego na mikrozwapnienia, a następnie konstrukcji metody łączenia ich w skupiska niosące informację diagnostyczną. Następnie obraz poddawany jest wyostrzaniu z użyciem transformaty falkowej. Tak przetworzone zdjęcia zostały przeanalizowane przez czterech doświadczonych radiologów. Analiza wykazała użyteczność systemu dla pewnych typów piersi oraz rodzajów skupisk mikrozwapnień.
Alternatywą dla zarysowanego wyżej sposobu analizy zdjęć rentgenowskich piersi jest zastosowanie metod uczenia maszynowego do detekcji mikrozwapnień lub zacienień. Celem tego typu algorytmów jest klasyfikacja regionów mammogramu pod kątem obecności markerów raka. W niniejszym artykule prezentujemy rezultaty analizy tego podejścia z wykorzystaniem dwóch metod klasyfikacji: AdaBoost oraz SVM. Najlepsze wyniki zostały uzyskane przez klasyfikator AdaBoost. Pozwala on na wykrywanie zacienień w mammogramach z czułością i specyficznością wynoszącą około 90%.
Skuteczność metod uczenia maszynowego silnie zależy od cech użytych do reprezentacji danych wejściowych. Szczególnie istotne znaczenie ma tu redukcja liczby cech tak, by tylko te, które niosą najwięcej informacji, były używane w klasyfikacji. W niniejszym artykule przedstawiamy wyniki selekcji cech na potrzeby klasyfikacji regionów mammogramów przy użyciu nowatorskiego algorytmu klasteryzacji.

Słowa kluczowe: analiza mammogramów, filtracja, wyostrzanie zdjęć metodą falkową, klasyfikacja obrazów

Abstract. Medical image analysis is one of the most important applications of computer science in medicine. Computer-aided support for breast cancer detection in mammograms is one of the examples of such an approach. In this paper, we present the results of our efforts aimed at creating algorithmic methods for supporting the breast cancer diagnosis.
Our work focuses on two approaches to detection of cancer markers in breast images. In the first approach, we have created a dedicated filter, which accentuates the appearance of the microcalcifications. Then, we have constructed a method for partitioning the detected microcalcifications into clusters of diagnostic importance. Next, the images are subject to wavelet transform-based enhancing. Mammograms treated by such a system were analysed by four experienced radiologists. Their analysis showed that the system indeed improves the diagnosis of breast cancer.
An alternative to the approach presented above is to use a group of machine learning methods to distinguish between abnormal lesions and normal tissue. In this paper we present the evaluation of two classifiers applied to this task, the AdaBoost and the SVM. The best results, achieved by the AdaBoost method, allow for recognition of masses in mammograms with sensitivity and specificity of ca. 90%.
The success of the machine learning methods depends strongly on the features used for representing the input data. Particularly important is the reduction of the number of features, so that only the most informative ones are used in classification. In this paper we show results of features selection, for classification of mammogram regions, using a novel clustering algorithm.
Key words: mammogram analysis, filtering, wavelet-based image enhancing, image classification

1. Wstęp

Rak piersi jest jedną z głównych przyczyn śmierci kobiet na tle chorób nowotworowych. W Stanach Zjednoczonych powoduje on około 40 000 zgonów rocznie. Ponadto każdego roku diagnozowanych jest około 200 000 nowych przypadków inwazyjnego raka piersi oraz około 55 000 przypadków raka in situ. W stadium zlokalizowanym choroby współcześnie stosowana terapia pozwala osiągnąć odsetek pięcioletnich przeżyć na poziomie 97%. Jednakże odsetek ten spada do 78% w przypadku choroby regionalnie zaawansowanej i 23%

w stadium choroby rozsianej. Liczby te wyraźnie wskazują, iż wczesne rozpoznanie raka piersi ma decydujące znaczenia dla możliwości podjęcia skutecznej terapii. Amerykańskie Towarzystwo Onkologiczne (American Cancer Society) zaleca, aby każda kobieta powyżej 40 roku życia przechodziła coroczne badanie mammograficzne.

Czułość mammografii przesiewowej można zwiększyć poprzez zastosowanie Komputerowych Systemów Wsparcia Detekcji (ang. *Computer-Aided Detection* – CAD). Należy jednak podkreślić, iż automatyczna detekcja podejrzanych zmian w cyfrowym zdjęciu mammograficznym jest problemem trudnym. Obrazy mammograficzne charakteryzuje niski kontrast, zwłaszcza w przypadku piersi gęstych radiologicznie, które często spotykane są u młodych kobiet. Ponadto symptomy obecności tkanki patologicznej są w wielu przypadkach stosunkowo subtelne. I tak na przykład zacienienia o nieostrym i nieregularnym zarysie, będące częstym objawem raka, sprawiają trudności przy automatycznej detekcji. Równie problematyczna jest ocena prawdopodobieństwa złośliwości wykrytej zmiany. Z tych powodów konieczne jest opracowanie skutecznych algorytmów poprawy kontrastu zdjęć mammograficznych oraz segmentacji i detekcji zacienień i mikrozwapnień.

2. Komputerowe wspomaganie wykrywania nowotworów

W [1] zaproponowaliśmy nowatorski system wykrywania skupisk mikrozwapnień w Cyfrowej Mammografii Małego Pola. Jego celem jest wspieranie radiologa w planowaniu biopsji poprzez zaznaczanie podejrzanych regionów zdjęcia rentgenowskiego oraz zwiększanie jego kontrastu.

Stworzony przez nas system działa w trzech krokach. W pierwszej fazie wykrywane są poszczególne mikrozwapnienia. W tym celu skonstruowaliśmy filtr o skończonej odpowiedzi impulsowej. Filtr ten charakteryzuje się odpowiedzią częstotliwościową dobraną w ten sposób, by dla typowego mikrozwapnienia rezultat działania filtru pozwalał na wyróżnienie markerów raka od otaczającej tkanki. Następnie mikrozwapnienia są grupowane w skupiska przy użyciu serii operacji morfologii matematycznej. W ostatnim kroku system zwiększa kontrast zdjęcia rentgenowskiego. W tym celu wykonujemy filtrowanie medianą, pozwalające na usunięcie szumu z mammogramu oraz wykorzystujemy przekształcenie obrazu oparte o dwuwymiarową transformatę falkową.

W celu oceny skuteczności systemu zebrano zestaw 50 zdjęć rentgenowskich Małego Pola. Pochodzą one od dr med. Tadeusza J. Popieli z Regionalnej Pracowni Wczesnej Diagnostyki i Leczenia Chorób Sutka Collegium Medicum Uniwersytetu Jagiellońskiego w Krakowie. Oceny systemu dokonało niezależnie od siebie czterech doświadczonych radiologów. W pierwszym kroku przeanalizowali oni oryginalne zdjęcia pod kątem obecności skupisk mikrozwapnień. Następnie wykonali analogiczną analizę zdjęć przetworzonych przez zaproponowany system komputerowego wspomagania diagnozy. Dla każdego zdjęcia radiolodzy odnotowali liczbę skupisk mikrozwapnień wykrytą na oryginalnych zdjęciach, jak również na zdjęciach będących efektem działania systemu. Dodatkowo radiolodzy dokonali jakościowej oceny poprawy wykrywalności skupisk mikrozwapnień wynikającej z zastosowania systemu. Każdy z mammogramów został także sklasyfikowany zgodnie z taksonomią typów piersi Wolfe'a, zaś skupiska mikrozwapnień zostały podzielone na klasy odpowiadające pięciu typom podanym przez LeGal.

Ocena skuteczności systemu dokonana przez radiologów wskazuje, iż poprawia on wykrywalność nowotworów piersi. Wzrost skuteczności jest obecny zwłaszcza w sytuacji, gdy na zdjęciu występuje kilka skupisk o różnym poziomie widoczności. Ponadto ocena systemu wykazała, iż największa poprawa rozpoznania nowotworów zachodzi dla piersi typów P2 i DY według Wolfe'a, jak również dla skupisk mikrozwapnień reprezentujących typy 2 i 4, a zwłaszcza dla mikrozwapnień typu 3. Rezultaty oceny przez radiologów zostały przedstawione w tab. 1-2.

Tab. 1. Wyniki oceny systemy dla pięciu typów skupisk mikrozwapnień wg. LeGal

Typ LeGal	Średnia liczba dodatkowo wykrytych skupisk mikrozwapnień	Ocena użyteczności (1-5)
Typ 1	0,2	2,9
Typ 2	0,4	3,1
Typ 3	0,8	3,8
Typ 4	0,3	3,1
Typ 5	0,2	2,8

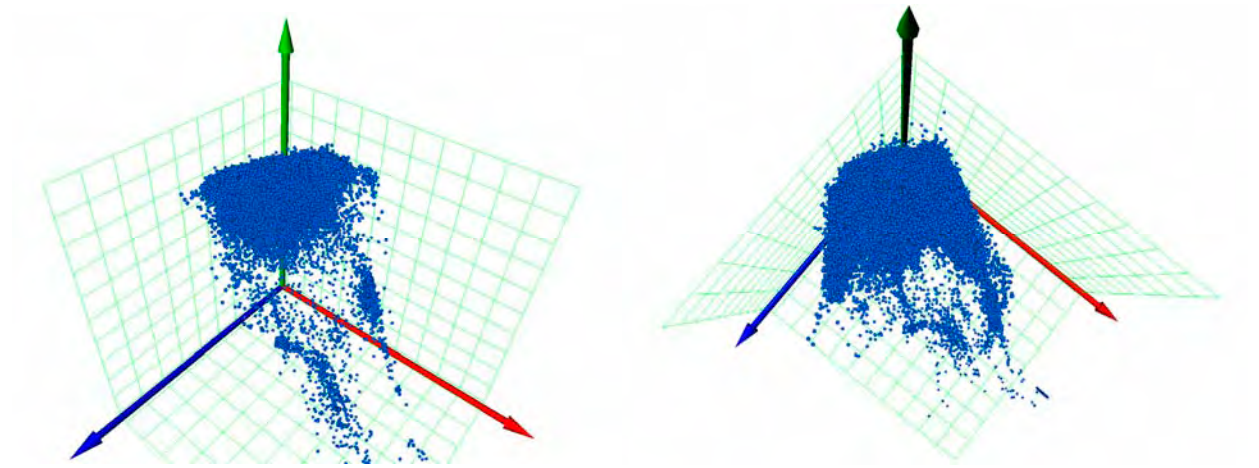
Tab. 2. Wyniki oceny systemu dla czterech typów piersi wg. Wolfe'a

Typ Wolfe'a	Średnia liczba dodatkowo wykrytych skupisk mikrozwapnień	Ocena użyteczności (1-5)
N1	0,2	1,0
P1	0,4	2,4
P2	1,0	4,2
DY	0,7	3,2

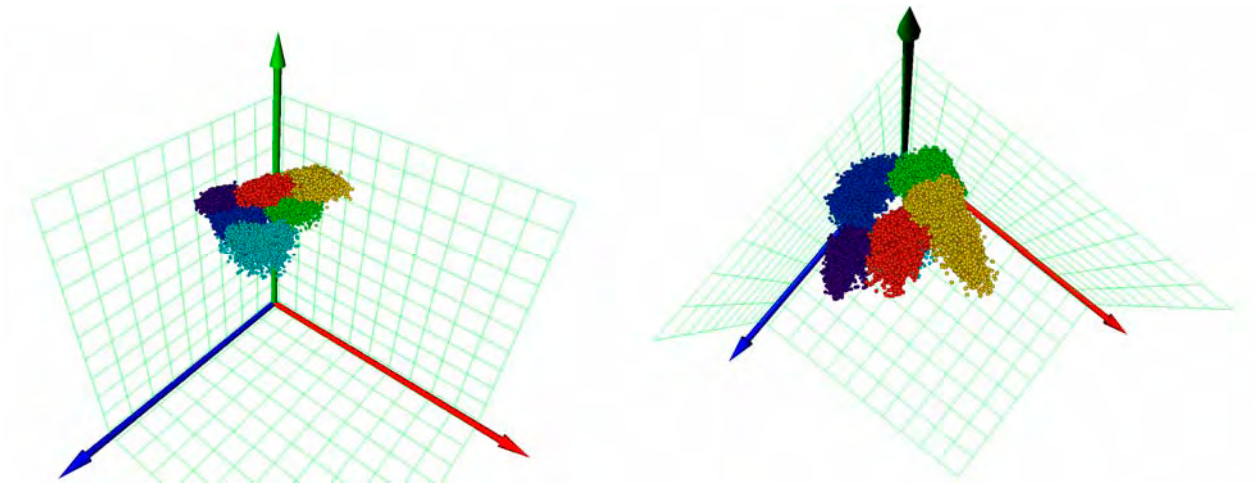
Innym podejściem do problemu wykrywania podejrzanych regionów na zdjęciach rentgenowskich piersi są metody oparte na uczeniu maszynowym. W [2] skupiliśmy się na ocenie dwóch metod klasyfikacji wyróżniających się dużą skutecznością uczenia, to jest klasyfikatora zespołowego AdaBoost oraz klasyfikatora Support Vector Machine. W metodzie AdaBoost, będącej reprezentantem szerszej metodologii występującej w literaturze pod nazwą „boosting", uczona jest duża grupa prostych klasyfikatorów. Każdy z nich podejmuje decyzję na podstawie wartości granicznej dotyczącej pojedynczej cechy danych wejściowych. Jednak aby uzyskać większą skuteczność uczenia, każdy z klasyfikatorów działa w oparciu o inną cechę. Ostateczna decyzja co do klasyfikacji przykładu jest podejmowana poprzez kombinację odpowiedzi poszczególnych klasyfikatorów prostych.

Metoda SVM wykorzystuje bank filtrów Gabora oraz próbkowanie we współrzędnych logarytmiczno-biegunowych do ekstrakcji cech obrazu rentgenowskiego. Dodatkowo, poprzez użycie funkcji jądra klasyfikator ten potrafi dopasować się do wysoce nieliniowych granic decyzyjnych.

Obydwa badane algorytmy zostały przetestowane na mammogramach o wysokiej rozdzielczości pochodzących ze zbioru Digital Database for Screening Mammography.



Rys. 1. Zbiór wielowymiarowych wektorów cech reprezentujących rejony zdjęć mammograficznych. Projekcja do przestrzeni trójwymiarowej została wykonana metodą Analizy Składowych Głównych (ang. *Principal Component Analysis*).



Rys. 2. Rezultat klasteryzacji regionów zdjęć mammograficznych z wykorzystaniem zaproponowanego przez nas algorytmu. Projekcja do przestrzeni trójwymiarowej została wykonana metodą Analizy Składowych Głównych (ang. *Principal Component Analysis*).

Najlepsze rezultaty zostały uzyskane przez klasyfikator AdaBoost. W przypadku tej metody czułość w stosunku do zacienień wynosi 92% przy specyficzności wynoszącej 88%.

3. Selekcja cech poprzez klasteryzację podejrzanych regionów mammogramu

Jednym z często stosowanych podejść do problemu weryfikacji istotności diagnostycznej podejrzanych regionów zdjęcia mammograficznego jest zastosowanie algorytmów uczenia maszynowego (ang. *machine learning algorithms*). Algorytmy te podejmują decyzję co do istotności danego regionu na podstawie wektora numerycznych cech, które opisują jego zawartość. Aby uzyskać wysoką skuteczność systemu detekcji konieczne jest więc opracowanie odpowiedniego zestawu numerycznych cech regionów. W [4] zaproponowaliśmy, aby problem ten rozwiązać stosując nowatorski algorytm klasteryzacji. Jedną z najistotniejszych cech tego algorytmu jest możliwość usuwania szumu ze zbioru danych. Jest to szczególnie przydatne w przypadku klasteryzacji regionów zdjęć mammograficznych, które nierzadko przedstawiają artefakty, takie jak brzeg zdjęcia lub fragmenty etykiety.

Selekcję cech przeprowadziliśmy, korzystając ze zbioru zawierającego około 93 000 podejrzanych regionów wykrytych na 200 wysokiej rozdzielczości cyfrowych zdjęciach mammograficznych. Dla każdego regionu wyliczyliśmy zestaw 46 statystycznych parametrów rozkładu jasności pikseli. Następnie, korzystając z klasycznego algorytmu k-środków oraz analizy histogramów cech, wybraliśmy podzbiór 27 cech. Uzyskane 27-wymiarowe wektory poklastrowaliśmy za pomocą zaproponowanego przez nas algorytmu klasteryzacji. Korzystając z rezultatu klasteryzacji, wyznaczyliśmy funkcję oceny dla każdej z 27 cech. Stosując 18 cech o najwyższej wartości funkcji oceny, uzyskaliśmy skuteczność weryfikacji istotności diagnostycznej regionów równą skuteczności uzyskiwanej dla oryginalnego zestawu 46 cech. Tak więc odrzucone 28 cech jest nieistotnych diagnostycznie.

4. Podsumowanie

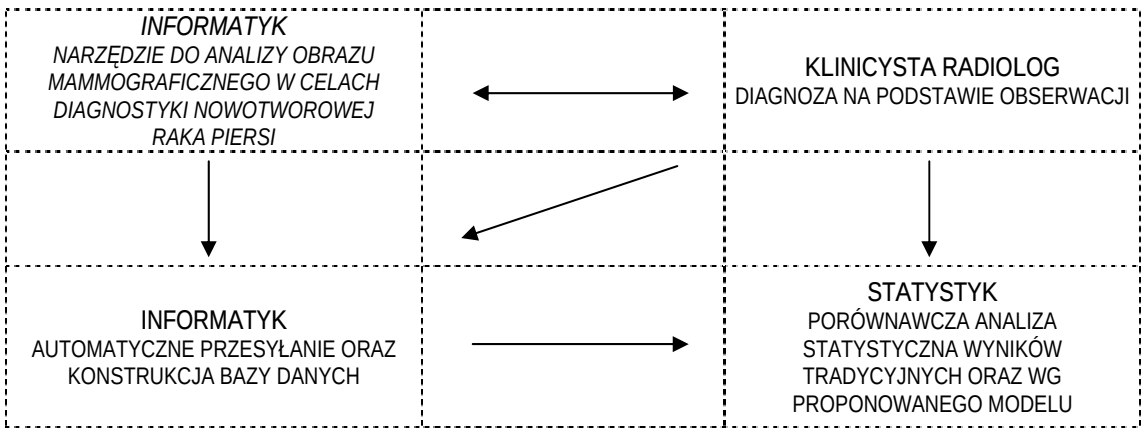
Celem prowadzonych przez nas prac jest stworzenie zintegrowanego systemu analizy mammogramów dla wsparcia diagnostyki nowotworów piersi. Przy jego tworzeniu planujemy wykorzystanie opisanego w niniejszym artykule algorytmu wykrywania mikrozwapnień opartego o filtrację oraz wyostrzanie falkowe. Do wykrywania zacienień planujemy użycie opisanych wyżej metod opartych na uczeniu maszynowym oraz selekcji cech w oparciu o klasteryzację. Uzyskane przez nas dotychczas wyniki wykazują, iż obydwie te metody charakteryzują się stosunkowo wysoką skutecznością wykrywania znaczników raka.

Podziękowania

Autorzy niniejszej pracy pragną wyrazić głębokie podziękowania dla dr med. Tadeusza J. Popieli z Regionalnej Pracowni Wczesnej Diagnostyki i Leczenia Chorób Sutka Collegium Medicum Uniwersytetu Jagiellońskiego w Krakowie za udostępnienie zbioru zdjęć rentgenowskich.

Literatura cytowana

1. Arodź T., Kurdziel M., Popiela T. J., Sevre E. O. D., Yuen D. A.: A 3D Visualization System for Computer-Aided Mammogram Analysis, University of Minnesota, Research Report UMSI 2004/181. Wysłano do Computer Methods and Programs in Biomedicine, Elsevier.
2. Arodź T., Kurdziel M., Sevre E. O. D., Yuen D. A.: Pattern Recognition Techniques for Automatic Detection of Suspicious-looking Anomalies in Mammograms. University of Minnesota, Research Report UMSI 2004/16. Computer Methods and Programs in Biomedicine, Elsevier, 2005 (w druku).
3. Boryczko K., Kurdziel M.: Recognition of Subtle Microcalcifications in High-Resolution Mammograms, [w]: Proc. IV International Conference on Computer Recognition Systems, CORES 2005, Advances in Soft Computing, Springer, 2005 (w druku).
4. Popiela T. J., Urbanik A., Arodź T., Kurdziel M.: Computeraid system for the detection of clustered microcalcifications in digital mammography, Abstract in XXXVII Polish Congress of Radiology, 2004.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 291-296.

ANALIZA OBCIĄŻEŃ PRACĄ NA PODSTAWIE WSKAŹNIKA WYKORZYSTANIA REZERWY TĘTNA I OBCIĄŻEŃ STATYCZNYCH METODĄ *OWAS*

ANALYSIS OF WORKLOADS ON THE BASIS OF THE HEART RATE INDEX AND *OWAS* METHOD

ANNA GROBORZ, TADEUSZ JULISZEWSKI, MARCIN GONCIARZ

Wydział Agroinżynierii, Akademia Rolnicza, Kraków
agroborz@tier.ar.krakow.pl, rtjulisz@cyf-kr.edu.pl, marcin@gonciarz.name

Streszczenie. W artykule przedstawiono analizę obciążeń pracą osób zatrudnionych w gospodarstwach rolnych. Na podstawie pomiarów tętna obliczono wskaźnik wykorzystania rezerwy tętna (WRT), który posłużył do oceny fizjologicznego obciążenia pracą. Do pomiarów i zapisu tętna zastosowana została aparatura firmy Polar wraz z dołączonym oprogramowaniem Polar Precision Performance. W celu analizy obciążeń statycznych wykorzystano metodę *OWAS* (Ovako Working Posture Analysis System), która została zaimplementowana jako autorski program komputerowy.

Słowa kluczowe: obciążenie pracą, wskaźnik wykorzystania rezerwy tętna (WRT), metoda *OWAS*

Abstract. Publication presents workload analysis of people employed on farms. On the basis of pulse measurement Heart Rate Reserve index was estimated, was used for the evaluation of the physiological workload. The pulse was measured and recorded by the Polar tool operating with Polar Precision Performance software. In order to analyse static loads, the *OWAS* (Ovako Working Posture Analysis System) method was implemented as computer program.

Key words: workload, Heart Rate Reserve, *OWAS* method

1. Wprowadzenie

Badaniem obciążenia fizycznego zajmuje się fizjologia pracy, która jest nauką odgrywającą istotną rolę w badaniach ergonomicznych. Wyniki badań fizjologicznych mają na celu opracowanie praktycznych zaleceń dotyczących sposobu wykonywania pracy i warunków, w jakich powinna ona przebiegać. Obciążenie pracą fizyczną na stanowisku pracy dzieli się na dynamiczne i statyczne. Podział ten jest związany z charakterem pracy mięśni – ich skurczem. Wysiłek dynamiczny przebiega w warunkach ruchu w związku z przemieszczeniem ciała ludzkiego lub jego części w przestrzeni i zachodzi przy udziale izotonicznych skurczów mięśni, podczas których następuje skrócenie włókien – okresy skurczu i rozkurczu mięśni [19], [9], [15]. Wysiłek statyczny jest wykonywany przy udziale skurczów izometrycznych, które powodują wzrost napięcia mięśni bez wyraźnego ich skrócenia. Praca statyczna przebiega w warunkach bezruchu, wiąże się przede wszystkim z koniecznością utrzymania niewygodnej i wymuszonej pozycji ciała [15].

Nadmierne obciążenie organizmu związane z pracą zawodową jest przyczyną zmęczenia obwodowego, obejmującego układ mięśniowo-szkieletowy i ogólne zmiany czynnościowe oraz zmęczenia ośrodkowego układu nerwowego. Zmiany składające się na zmęczenie obwodowe i ośrodkowe prowadzą do dolegliwości, a nawet urazów układu mięśniowo-szkieletowego, wyczerpania zasobów energetycznych i metabolicznych, zaburzenia czynności percepcyjnych, w tym i koordynacji wzrokowo-ruchowej i obniżenia sprawności umysłowej [7].

Rozwój technologii, w tym mechanizacja i automatyzacja, znacznie zmniejsza liczbę osób zatrudnionych przy pracach wykonywanych ręcznie: podnoszeniu, przenoszeniu, ciągnięciu, pchaniu dużych i ciężkich przedmiotów i narzędzi. Jednak nadal istnieją zawody, w których wielu spośród wykonywanych czynności nie da się zastąpić rozwiązaniami technicznymi, przykładowo prace naprawcze, wykonawcze w budownictwie czy rolnictwie.

2. Cel i zakres

Celem badań była analiza obciążeń pracą osób pracujących w rolnictwie, w fermie drobiu (kur niosek), na podstawie wskaźnika wykorzystania rezerwy tętna i metody *OWAS*.

Zakres badań obejmował siedem osób (dwie kobiety i pięciu mężczyzn) pracujących przy produkcji jaj konsumpcyjnych w systemie klatkowym w obrębie jednej fermy z woje-

województwa śląskiego. Charakterystykę badanych osób przedstawia tab. 1.

Tab. 1. Charakterystyka badanych osób

Osoba	Płeć*	Wiek [lat]	Masa ciała [kg]	Wzrost [m]	BMI ** [kg/m²]	Uwagi
A	K	37	72	1,66	26,1	pali papierosy
B	K	48	65	1,65	23,9	pali papierosy
C	M	21	71	1,80	21,9	brak
D	M	25	75	1,82	22,6	pali papierosy
E	M	27	73	1,70	25,3	pali papierosy
F	M	40	65	1,70	22,5	brak
G	M	41	67	1,69	23,5	pali papierosy

*K – kobieta
M – mężczyzna,
**Wskaźnik BMI (1).

$$BMI\left[\frac{kg}{m^2}\right] = \frac{masa[kg]}{wzrost^2[m^2]} \quad (1)$$

Wartość wskaźnika BMI [kg/m²] wg WHO (World Health Organization):

BMI <18,5 – niedowaga,
BMI 18,5-24,9 – normalna waga,
BMI 25-29,9 – nadwaga,
BMI 30 –34,9 – I klasa otyłości,
BMI 35-39,9 – II klasa otyłości,
BMI >40 – III klasa otyłości.

3. Metodyka

Analizę obciążeń pracą badanych osób przeprowadzono przy pomocy:
- wskaźnika wykorzystania rezerwy tętna obliczonego na podstawie pomiarów tętna (obciążenie dynamiczne),
- metody *OWAS* – Ovako Working Posture Analysis System (obciążenie statyczne).

Tętno (HR) jest najłatwiejszym do zmierzenia i zarejestrowania wskaźnikiem reakcji układu krążenia na wysiłek [20]. Maksymalne tętno (HR_{max}) zależne jest przede wszystkim od wieku człowieka i można je obliczyć z zależności [1]:

HR_{max} = 220 – wiek osoby (w latach).

Minimalne tętno, zwane też *tętnem bazowym*, (HR_{min}) występujące podczas snu jest zróżnicowane indywidualnie w zakresie od 45 do 70 ud/min. Różnica pomiędzy maksymalnym tętnem a tętnem minimalnym jest nazywana *rezerwą tętna*, tj. swego rodzaju „zapasem" częstości skurczów serca jako reakcji na wysiłek. Wykorzystanie rezerwy tętna można wyrazić w wartościach bezwzględnych (ud/min), tj. wzrostem częstości skurczów serca w porównaniu do HR_{min} lub w jednostkach względnych (%) jako iloraz wzrostu częstości skurczów serca (w porównaniu do HR_{min}) do rezerwy tętna [2].

Pomiary wykonywano miernikiem tętna POLAR VANTAGE, który umożliwia także bezprzewodową rejestrację wyników. Przy pomocy interfejsu zarejestrowane dane, tj. tętno, przenoszone są do komputera, gdzie mogą być analizowane programem dostarczanym przez producenta miernika. Program komputerowy Polar Precision Performance umożliwia obliczanie między innymi średniej wartości tętna badanej osoby podczas pracy w wybranym przedziale czasowym [10]. Metoda bezprzewodowego pomiaru tętna została opracowana w 1983 r. na Uniwersytecie w Oulu, w Katedrze Elektroniki (Finlandia). Spośród metod pomiaru i rejestracji tętna powszechnie stosuje się obecnie bezprzewodowe mierniki. Metoda ta wykorzystywana jest do ergonomicznych badań obciążenia pracą personelu zatrudnionego w przemyśle [15], usługach transportowych [11], a także w rolnictwie [12]. Na Wydziale Agroinżynierii Akademii Rolniczej w Krakowie od kilku lat prowadzone są badania obciążenia pracą osób zatrudnionych w rolnictwie, m.in. traktorzystów, osób zatrudnionych przy pracach polowych, w ogrodnictwie i w fermach drobiu czy krów mlecznych.

Wskaźnik wykorzystania rezerwy tętna [4], [16] został obliczony według następującego wzoru (2):

$$WRT = \frac{HR_{pracy} - HR_{min}}{HR_{max} - HR_{min}} \cdot 100\,[\%] \quad (2)$$

gdzie:
HR_{pracy} – średnie tętno danej czynności [ud/min],
HR_{min} – tętno bazowe [ud/min],
HR_{max} – tętno maksymalne dla danej osoby [ud/min].

Klasyfikacja wysiłku na podstawie wykorzystania rezerwy tętna przy długotrwałym obciążeniu jest następująca [2]:

do 25% – wysiłek bardzo mały,
25 – 35% – wysiłek relatywnie mały,
35 – 50% – wysiłek średni (niem. *mittelstark*),
50 – 65% – wysiłek dopuszczalny tylko dla osób ze zdrowym układem krążenia,
65 – 80 % – wysiłek dopuszczalny tylko warunkowo,
powyżej 80 % – wysiłek niedopuszczalny.

Analiza obciążenia metodą *OWAS* służy do oceny wielkości obciążenia statycznego na stanowiskach pracy. Jest powszechnie stosowanym systemem identyfikacji i oceny postawy całego ciała ludzkiego podczas pracy. Metoda ta została opracowana w Finlandii przez firmę Ovako [13], [14] w odpowiedzi na potrzebę prostej, jasnej i uzasadnionej metody analitycznej. Ich celem była analiza obciążeń postawy ciała ludzkiego i poprawa czynników ergonomicznych w miejscu pracy. Stworzenie stanowiska pracy, które nie powodowałoby nadmiernego obciążenia i zmęczenia układu mięśniowo-szkieletowego pracownika, wymaga przeanalizowania parametrów związanych z pozycją przy pracy, siłą wywieraną podczas pracy oraz czasem oddziaływania siły przy określonej pozycji. Parametry te powinny być rozpatrywane łącznie. Metoda *OWAS* bierze pod uwagę obciążenie pochodzące od czterech czynników:
- pozycji pleców (cztery zakodowane pozycje: 1 – wyprostowane, 2 – zgięte do przodu, 3 – skręcone, 4 – zgięte i skręcone),
- położenia przedramion (trzy pozycje: 1 – obydwa poniżej stawu łokciowego, 2 – jedno powyżej stawu łokciowego, 3 – obydwa powyżej stawu łokciowego),
- pracy nóg (siedem pozycji: 1 – pozycja siedząca, 2 – stojąca z nogami wyprostowanymi, 3 – stojąca z jedną nogą wyprostowaną, 4 – stojąca z nogami zgiętymi, 5 – stojąca z jedną nogą zgiętą, 6 – klęczenie na jednym lub obu kolanach, 7 - chodzenie) [18],

- wielkości obciążenia zewnętrznego (trzy wielkości, zróżnicowane według obciążonych osób, oparte na aktach prawnych - tab. 2) [6], [7].

Tab. 2. Kody obciążenia zewnętrznego w zależności od populacji [7]

Kod obciążenia zewnętrznego	Mężczyźni	Kobiety i młodociani chłopcy	Dziewczęta
	[kg]		
1	< 10	< 5	< 2
2	10 - 20	5 - 10	2 - 6
3	> 20	> 10	> 6

Na podstawie pełnego kodu pozycji ciała ludzkiego składającego się z czterech cyfr (po jednej od każdego czynnika obciążającego) należy odczytać typ kategorii z diagramu (rys. 1).

PLECY																							
	PRZEDRAMIONA																						
NOGI		1			2			3			4			5			6			7		
OBC.		1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	1	1	1	1	1	1
	2	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	1	1	1	1	1	1
	3	1	1	1	1	1	1	1	1	1	2	2	3	2	2	3	1	1	1	1	1	2
2	1	2	2	3	2	2	3	2	2	3	3	3	3	3	3	3	2	2	2	2	3	3
	2	2	2	3	2	2	3	2	3	3	3	4	4	3	4	4	3	3	4	2	3	4
	3	3	3	4	2	2	3	3	3	3	3	4	4	4	4	4	4	4	4	2	3	4
3	1	1	1	1	1	1	1	1	1	2	3	3	3	4	4	4	1	1	1	1	1	1
	2	2	2	3	1	1	1	1	1	2	4	4	4	4	4	4	3	3	3	1	1	1
	3	2	2	3	1	1	1	2	3	3	4	4	4	4	4	4	4	4	4	1	1	1
4	1	2	3	3	2	2	3	2	2	3	4	4	4	4	4	4	4	4	4	2	3	4
	2	3	3	4	2	3	4	3	3	4	4	4	4	4	4	4	4	4	4	2	3	4
	3	4	4	4	2	3	4	3	3	4	4	4	4	4	4	4	4	4	4	2	3	4

Rys. 1. Diagram do określania kategorii według kodu pozycji

Dokładny opis kategorii oraz zalecenia odnośnie ewentualnych zmian na stanowisku pracy przedstawiono w tabeli 3. Do celów analizy obciążenia układu mięśniowo-szkieletowego dla każdej z utrzymywanych pozycji istotny jest czynnik czasu trwania określonej czynności opisanej przez kategorię. W zależności od czasu utrzymywania pozycji ciała określonej zgodnie z kategoriami metody *OWAS* dokonano oceny obciążenia statycznego jako duże, średnie lub małe (tab. 4). Pozycje ciała zostały zróżnicowane na wymuszoną i niewymuszoną. Pozycja wymuszona to taka, która jest narzucona konstrukcją stanowiska pracy lub rodzajem wykonywanych czynności, a jej modyfikacja pod wpływem subiektywnego odczucia pracownika i zgodnie z jego preferencjami nie jest możliwa. Natomiast pozycja niewymuszona to taka, która może być zmieniana zgodnie z poczuciem wygody pracownika, bez wpływu na jakość wykonywanej pracy.

Ze względu na adaptowalność do analizy codziennego miejsca pracy i możliwości oceny różnych postaw ciała metoda ta była stosowana w wielu badaniach pozycji ciała w różnych zawodach: konstruktorzy budowlani [21], praca przy uprawie roślin leczniczych i włókienniczych [5], kobiety pracujące w fermach krów mlecznych [22], osoby zatrudnione przy produkcji owoców [3] czy przy sadzeniu roślin w leśnictwie [8].

Tab. 3. Opis kategorii

Kategoria	Opis kategorii
Kategoria 1	- pozycja lub pozycje przyjmowane podczas pracy są naturalne, - obciążenie jest optymalne lub akceptowalne, - nie ma potrzeby dokonywania zmian na stanowisku;
Kategoria 2	- pozycja lub pozycje przyjmowane podczas pracy mogą mieć negatywny wpływ na układ mięśniowo-szkieletowy, - obciążenie jest prawie akceptowalne, - nie ma potrzeby dokonywania natychmiastowych zmian na stanowisku, ale należy wziąć pod uwagę konieczność przeprowadzania takich zmian w bliskiej przyszłości;
Kategoria 3	- pozycja lub pozycje przyjmowane podczas pracy mają negatywny wpływ na układ mięśniowo-szkieletowy, - obciążenie jest duże, - zmiany na stanowisku należy przeprowadzić tak szybko, jak to możliwe;
Kategoria 4	- pozycja lub pozycje przyjmowane podczas pracy mają bardzo negatywny wpływ na układ mięśniowo-szkieletowy, - obciążenie jest bardzo duże, - zmiany na stanowisku należy przeprowadzić natychmiast.

Tab. 4. Ocena obciążenia statycznego w zależności od czasu utrzymywania pozycji ciała

Ocena obciążenia	Pozycja ciała przy pracy (kategorie OWAS)	Czas utrzymywania jednej pozycji [% zmiany roboczej]
małe	Pozycja niewymuszona kategorii 1	do 70
	Pozycja wymuszona kategorii 1 lub niewymuszona kategorii 2	do 50
	Pozycja wymuszona kategorii 2	do 30
średnie	Pozycja niewymuszona kategorii 1	powyżej 70
	Pozycja wymuszona kategorii 1 lub niewymuszona kategorii 2	50 - 70
	Pozycja wymuszona kategorii 2	30 - 50
	Pozycja wymuszona kategorii 3 lub 4	do 30
duże	Pozycja wymuszona kategorii 1 lub niewymuszona kategorii 2	powyżej 70
	Pozycja wymuszona kategorii 2	powyżej 50
	Pozycja wymuszona kategorii 3 lub 4	powyżej 30

Metoda *OWAS* została zaimplementowana jako autorski program komputerowy wspomagający analizę obciążeń statycznych.

4. Wyniki

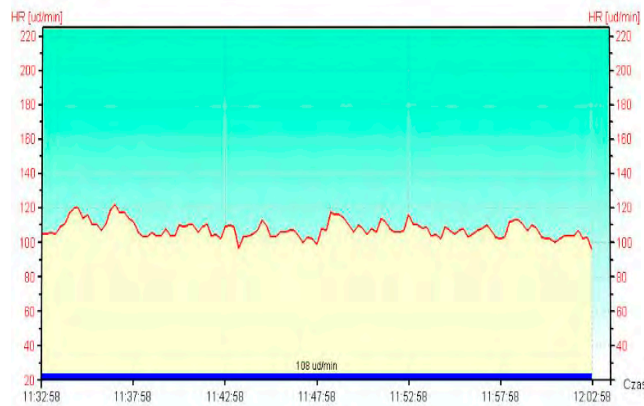
Analiza wskaźników wykorzystania rezerwy tętna podczas czynności wykonywanych przy produkcji jaj konsumpcyjnych pozwoliła na klasyfikację stopnia ciężkości wysiłku fizycznego według skali Buchbergera. Wyniki badań przeprowadzonych

na podstawie pomiarów tętna oraz obliczonego wskaźnika wykorzystania rezerwy tętna przedstawiono w tab. 5, w której dokonano także oceny wysiłku fizycznego badanych osób. Obciążenie pracą podczas poszczególnych czynności oceniono jako relatywnie małe (11 zapisów czynności), bardzo małe (10 zapisów), średnie (7 zapisów – np. załadunek jaj do samochodu, przenoszenie jaj posortowanych według klas wagowych na półki wózków transportowych) oraz dopuszczalne tylko dla osób ze zdrowym układem krążenia (1 zapis – w sytuacji awaryjnej ręczny zbiór jaj).

Na rys. 2 przedstawiono wykres tętna w czasie jednej z badanych osób (osoba B) podczas sortowania i przenoszenia posortowanych jaj na wózki transportowe. Czynność ta została zaklasyfikowana według wskaźnika WRT [%] jako wysiłek średni.

Tab. 5. Tętno, wskaźnik wykorzystania rezerwy tętna badanych osób podczas pracy w fermie drobiu (kur niosek) oraz ocena wysiłku fizycznego

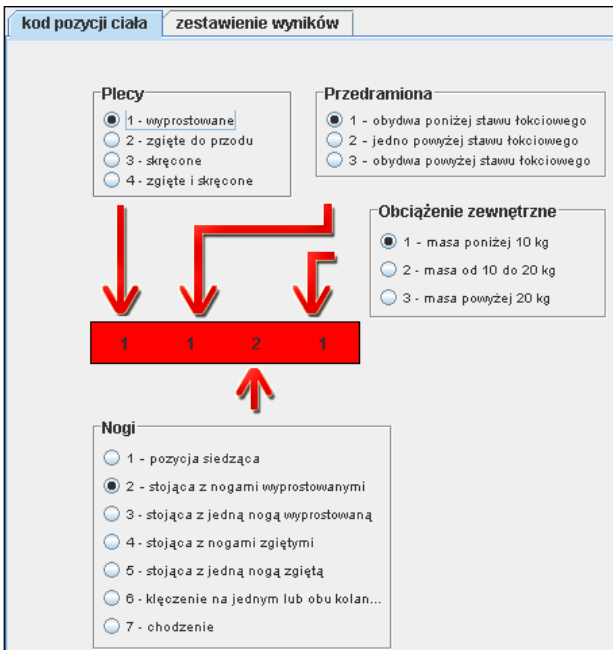
Osoba	Czynność	HR_{pracy}	HR_{max}	HR_{min}	WRT	Ocena wysiłku fizycznego
		[ud/min]			[%]	
A	zamiatanie hal	91	183	44	33,8	relatywnie mały
A	odpoczynek w pozycji siedzącej	68	183	44	17,3	b. mały
A	sortownia, wybieranie stłuczek	76	183	44	23,0	b. mały
A	zamiatanie hal	79	183	44	25,2	relatywnie mały
B	sortownia, przenoszenie jaj	111	172	60	45,5	średni
B	administracja	92	172	60	28,6	relatywnie mały
B	sortownia, przenoszenie jaj	108	172	60	42,9	średni
B	wydawanie i sprzedaż jajek	94	172	60	30,4	relatywnie mały
C	naprawa silnika elektrycznego w pozycji stojącej	76	199	44	20,6	b. mały
C	odpoczynek w pozycji siedzącej	80	199	44	23,2	b. mały
C	ręczny zbiór jaj	108	199	44	41,3	średni
C	zamiatanie hal	117	199	44	47,1	średni
C	odpoczynek w pozycji siedzącej	88	199	44	28,4	relatywnie mały
C	wnoszenie do klatek uwolnionych przypadkowo kur	96	199	44	33,5	relatywnie mały
C	prace remontowo-porządkowe, naprawa taśmy pomiotowej	82	199	44	24,5	b. mały
C	prace domowe: zmywanie naczyń, wycieranie kurzu	84	199	44	25,8	relatywnie mały
C	rozłupywanie drewna siekierką w pozycji pochylonej	93	199	44	31,6	relatywnie mały
C	naprawa łopaty w pozycji pochylonej	82	199	44	24,5	b. mały
D	kontrola produkcji, chodzenie między bateriami klatek i wzdłuż linii technologicznej	87	195	50	25,5	relatywnie mały
E	sprzątanie samochodu dostawczego	102	193	57	33,1	relatywnie mały
E	telefonowanie, administracja	85	193	57	20,6	b. mały
E	załadunek jaj	109	193	57	38,2	średni
F	sprzątanie hal, zamiatanie	80	180	53	21,3	b. mały
G	naprawa silnika elektrycznego w pozycji stojącej	75	179	43	23,5	b. mały
G	sortownia, wybieranie stłuczek	76	179	43	24,3	b.mały
G	usuwanie awarii maszyny w linii technologicznej (przekładnia), zatrzymanie produkcji	104	179	43	44,9	średni
G	ręczny zbiór jajek	118	179	43	55,1	dop. dla osób ze zdr. ukł. krąż.
G	sortownia, wybieranie stłuczek	95	179	43	38,2	średni
G	kontrola maszyn, chodzenie wzdłuż linii technologicznej	89	179	43	33,8	relatywnie mały



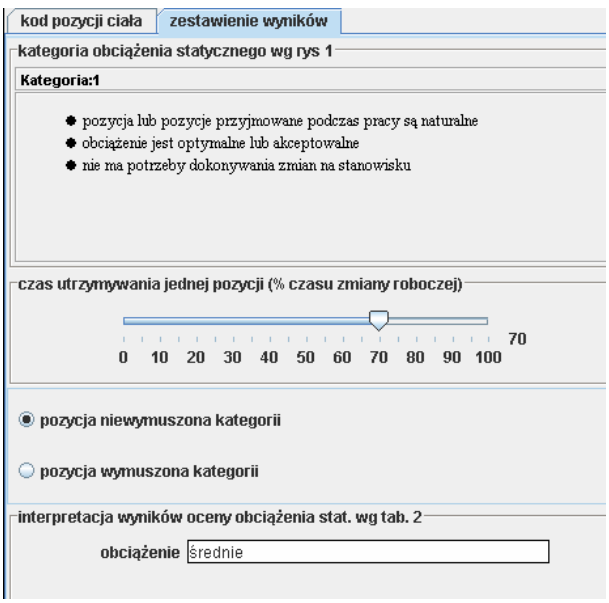
Rys. 2. Wykres tętna w czasie (w programie Polar Precision Performance SW 3) podczas sortowania jaj konsumpcyjnych (układanie i przenoszenie wytłaczanek z jajkami) przy sortowni MOBA 1500

Do analizy obciążenia statycznego metodą *OWAS* wzięto pod uwagę osobę B podczas sortowania jaj. Pomiar polegał na obserwacji osoby sortującej jaja konsumpcyjne, przenoszącej posortowane wytłaczanki z jajami na wózki transportowe. Masa przenoszonych jajek wynosiła 10-11 kg. Czynność zarejestrowano metodą filmową (kamera cyfrowa). Następnie analizowano poszczególne pozycje ciała w interwałach czasowych, co 10 sekund i zapisywano kod w programie komputerowym. Na rys. 3 przedstawiono określanie kodu po-

zycji ciała przy wykorzystaniu programu komputerowego. Ocena obciążenia statycznego zależy od kategorii oraz czasu trwania poszczególnej czynności (pozycji ciała). Przykładową ocenę pokazano na rys. 4.



Rys. 3. Określenie kodu pozycji ciała w programie komputerowym *OWAS*



Rys. 4. Zestawienie wyników oceny obciążenia statycznego

Wyniki pomiarów obciążenia statycznego metodą *OWAS* przedstawiono w tab. 6, w której zapisano występujące kody pozycji ciała (plecy, przedramiona, nogi, obciążenie zewnętrzne) oraz klasyfikację danego kodu do odpowiedniej kategorii. Następnie w tab. 7 zamieszczono sumaryczny i procentowy udział kategorii poszczególnych pozycji ciała w stosunku do całkowitego czasu badania. Najczęściej przyjmowaną pozycję ciała przez badaną osobę zaliczono do kategorii pierwszej (54% całego czasu badania). Nieco mniejszy procent czasu pracy (32%) zawierał się w kategorii drugiej oraz 14% do kategorii trzeciej. Do kategorii czwartej nie zakwalifikowano żadnej postawy ciała badanej osoby. Na podstawie czasu pracy i kategorii poszczególnych pozycji ciała obciążenie statyczne oceniono jako średnie (tab. 7).

Tab. 6. Kody pozycji ciała badanej osoby metodą *OWAS* (plecy, przedramiona, nogi, obciążenie zewnętrzne) wraz z klasyfikacją do odpowiedniej kategorii

Plecy	Przedramiona	Nogi	Obciążenie zewnętrzne	Kategoria
2	3	2	1	2
1	1	7	1	1
1	3	2	1	1
1	3	2	1	1
1	3	2	1	1
2	1	3	3	3
1	3	2	1	1
2	1	4	3	3
2	2	2	1	2
1	2	2	1	1
1	1	7	1	1
1	3	2	1	1
1	3	2	1	1
2	3	2	1	2
2	3	2	1	2
2	3	2	1	2
2	3	2	1	2
1	3	2	1	1
2	3	2	1	2
2	2	2	1	2
2	1	4	3	3
1	2	2	1	1
2	3	3	1	3
2	1	2	1	2
1	1	2	1	1
1	1	2	1	1
1	1	2	1	1
1	1	2	1	1

Tab. 7. Ocena obciążenia statycznego

Kategoria	Ilość pozycji w danej kategorii	Udział procentowy [%]	Ocena obciążenia statycznego
1	15	54	średnie
2	9	32	średnie
3	4	14	średnie
4	0	0	brak
Suma	28	100	**średnie**

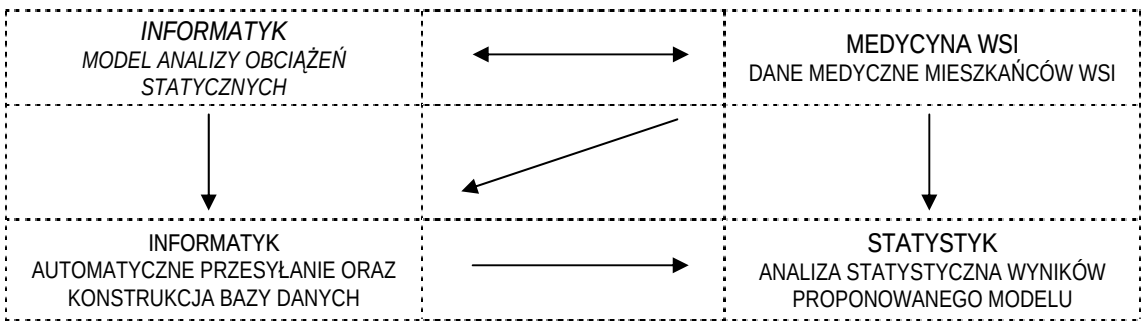
5. Wnioski

1. Wskaźnik wykorzystania rezerwy tętna i metoda *OWAS* mogą być zastosowane do oceny obciążenia pracą personelu w fermach drobiu.
2. Większość czynności w analizowanej technologii produkcji w fermie drobiu powoduje relatywnie małe i małe obciążenia pracą personelu. Tylko wyjątkowo występowały obciążenia określane jako dopuszczalne tylko dla osób ze zdrowym układem krążenia.
3. Obciążenia statyczne ocenione przy pomocy metody *OWAS* klasyfikują się w większości w kategorii 1, tj. kategorii obciążeń statycznych średnich.

Literatura cytowana

1. American College of Sports Medicine, 1990, (http://support.polar.fi – z dnia 22.04.2005).
2. Buchberger J.: Die Beurteilung von Arbeitsbeanspruchungen aufgrund der kontinuierlich registrierten Herzschlagfrequenz. Arbeitsärztlicher Dienst des BIGA, Arbeitsmedizinische Informationen, Nr.12, Bern 1984.
3. Calisto C.: Ergonomic investigations in fruit growing. Musculoskeletal disorders and their risk factors, Verlag Grauer, Stuttgart 1999.
4. Center for Lifetime Wellness, University of Rochester Medical Center, 2003. (http://www.stronghealth.com/services/primarycare/toolkitfiles/heartrate.cfm – z dnia 22.04.2005).
5. Chavalitsakulchai P., Shahnavaz H.: Ergonomics method for preventions of musculoskeletal discomforts among female industrial workers: physical characteristics and work factors, Journal of Human Ergology, 22:95-113, 1993.
6. Corlett E. N.: The evaluation of posture and its effects. In J. R. Wilson and E. N. Corlett (Eds.), Evaluation of human work: a practical ergonomics methodology (2nd ed.), Taylor & Francis, London 1995.
7. European Agency for Safety and Health at Work, 2003, (http://www.bp.edu.pl/good_practice/oums.stm – z dnia 22.04. 2005).
8. Giefing D. F., Grzywiński W., Kosak J.: Physical work load of workers during planting using different types of dibbles, Electronic Journal of Polish Agricultural Universities, Vol.6, Issue 1, Series Forestry, 2003.
9. Grandjean E.: Fizjologia pracy. Zarys ergonomii, PZWL 1971.
10. Groborz A., Juliszewski T.: Wstępna ocena obciążenia pracą kobiet podczas typowych prac w gospodarstwie domowym na wsi, Inżynieria Rolnicza, 11(53): 91-96, Kraków 2003.
11. Honaker R. E.: Assessing trailer material handling tasks: biomechanical modeling, posture categorization, physiological measure, and subjective rating. Thesis submitted to the Faculty of the Virginia Polytechnic Institute and State University for the degree of Master of Science in Industrial and System Engineering, Blacksburg, Virgina1996.
12. Juliszewski T.: Udoskonalenia techniczne w gospodarstwach rolnych a obciążenie pracą. W: Obciążenie fizyczne i psychiczne pracą w rolnictwie. Praca zbiorowa pod red. F. Bujaka, J. Zagórskiego. Instytut Medycyny Wsi, 151-156, Lublin 2001.
13. Kahru O., Kansi P., Kuorinka I.: Current working postures in industry: A practical method for analysis, Applied Ergonomics, 8: 199-201, 1977.
14. Kahru O., Härkönen R., Sorvali P., Vepsäläinen P.: Observing working postures in industry: Examples of OWAS application, Applied Ergonomics, 12(1): 13-17, 1981.
15. Kamieńska-Żyła M.: Ocena obciążenia fizycznego człowieka w procesie pracy, W: Ergonomia i ochrona pracy, pod red. S. Knapika, Wydawnictwa AGH, Kraków 1996.
16. Karvonen M. et al.: The effect of training on heart rate. A longitudinal study, Ann. Med. Exp. Biol. Fenn, 35: 307-315, 1957.
17. Kaukiainen A. et al.: New equipment to lighten the work load of construction workers, International Journal of Occupational Safety and Ergonomics (JOSE), Vol. 8, No. 2: 209-224, 2002.
18. Keyserling W. M.: Postural analysis of the trunk and shoulders in simulated real time, Human Factors, 29: 569-583, 1986.
19. Koradecka D., Łatach H.: Ocena wielkości obciążeń pracą fizyczną na stanowiskach różnych, CIOP, Warszawa 1981.
20. Koradecka D. et. al.: Bezpieczeństwo pracy i ergonomia, Centralny Instytut Ochrony Pracy, Warszawa1997.
21. Mattila M., Karwowski W., Vilkki M.: Analysis of working postures in hammering tasks on building construction sites using the computerized OWAS method, Applied Ergonomics, 24(6): 405-412, 1993.
22. Nevala-Puranen N.: Reduction of Farmers' postural load during occupationally oriented medical rehabilitation, Applied Ergonomics, 26(6): 411-415, 1995.

Praca naukowa finansowana ze środków budżetowych na naukę w latach 2005/2006 jako projekt badawczy.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No.1/2, 2005, pp. 297-300.

NOWE METODY POMIARU I ANALIZY PATOLOGICZNEGO DRŻENIA KOŃCZYN GÓRNYCH

INNOVATIVE WAYS OF MEASURMENT AND ANALYSIS OF PATHOLOGICAL TREMOR OF UPPER LIMBS

ANDRZEJ IZWORSKI[*], MARCIN MICHAŁEK[*], MONIKA RUDZIŃSKA[***]

[*]*Katedra Automatyki Akademii Górniczo-Hutniczej, Kraków, izwa@agh.edu.pl,*
[**]*michal@biocyb.ia.agh.edu.pl*
[***] *Klinika Neurologii Collegium Medicum Uniwersytetu Jagiellońskiego, Kraków, rudzinsk@neuro.cm-uj.krakow.pl*

Streszczenie. Patologiczne drżenia ludzkiego ciała objawiają się rytmicznym i naprzemiennym skurczem mięśni, co w głównej mierze jest następstwem tzw. chorób pozapiramidowych, do których należy choroba Parkinsona. Efektem zaburzeń w PD jest drżenie niektórych części ciała, w dominującej zaś części okolic twarzy, barku i kończyn górnych. Obecne metody pomiarowe tego typu drżenia, wykorzystujące głównie tablety cyfrowe jako urządzenia peryferyjne komputera, umożliwiając akwizycję danych z dwuwymiarowej powierzchni i dodatkowej informacji w postaci nacisku, nie potrafią wiernie zarejestrować swobodnych drgań parkinsonowskich. Biorąc pod uwagę konieczność pomiaru wszystkich cech choroby, zdecydowano się na projekt nowej, całkowicie nieinwazyjnej metody z wykorzystaniem kamer cyfrowych i punktowych źródeł światła umieszczonych bezpośrednio na badanym i będącym w ruchu obiekcie.

Słowa kluczowe: drżenie, choroba Parkinsona, tablet cyfrowy, analiza 3D

Abstract. Pathological tremor of human body manifests itself by periodic, alternating muscle contractions, and is mainly caused by the so-called extra-pyramidal disorders, including Parkinson's disease. The paper describes a 3D method for measurement and visualization of upper limbs tremor, making use of digital camcorders for registration of moving point light sources. The measurement is performed separately for the three principal planes of the Cartesian coordinate system. The present work contains a description of the measurement stand and measurement method, the obtained results, and the prospects for application of the method helpful in medical diagnosis. In the long term prospect the described method will assist in examinations and evaluation of the upper limbs tremor in patients with Parkinson's disease. The presented method of tremor measurement considerably enhances the currently applied acquisition techniques, which usually reduce the measurement space to 2D.

Key words: tremor, Parkinson's disease, 3D analysis, biomedical signal processing

1. Pomiar i analiza drżenia kończyn z wykorzystaniem tabletów cyfrowych

Podstawowymi parametrami do obiektywnej oceny postępu choroby są amplituda i częstotliwość drżenia [2]. Obie te cechy mogą zostać uzyskane metodą wykorzystującą tablice elektroniczne (zwane również tabletami cyfrowymi lub graficznymi). Najczęściej wykonywanym testem jest rysowanie spirali Archimedesa według przyjętego wzorca [3], [5], [6]. Dane z tabletu w trakcie badania przekazywane są portem RS232 lub USB do komputera i analizowane przez aplikację pracującą pod kontrolą systemu Windows, stworzoną w Laboratorium Biocybernetyki AGH. Program pobiera z obszaru roboczego tabletu następujące parametry badania: położenie wzdłuż osi X, położenie wzdłuż osi Y i nacisk pióra. Bazując na otrzymanych danych, można wyznaczyć charakterystyki czasowe i częstotliwościowe drżenia w płaszczyźnie X-Y. Siła nacisku jest tutaj dodatkowym wskaźnikiem, nie może być jednak rozpatrywana z punktu widzenia drgań wzdłuż trzeciej osi układu kartezjańskiego, osi Z.

Dodatkowo wskaźnik (pióro), które pacjent trzyma w dłoni, musi znajdować się podczas rejestracji w odległości do kilku milimetrów od powierzchni roboczej i nie może być odchylone od pionu o więcej niż zadany maksymalny kąt.

2. Rejestracja i analiza drżenia za pomocą kamer cyfrowych

Cechą choroby Parkinsona jest występowanie drżenia spoczynkowego, które obejmuje najczęściej kończyny górne, głowę, a także dolne okolice twarzy [1], [4]. Podane wyżej objawy drżenia spoczynkowego ulegają redukcji w trakcie

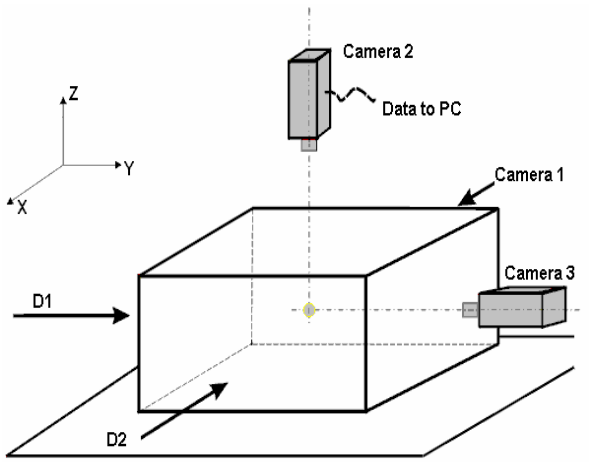
wykonywania czynności ruchowych [7], [8]. Dlatego istotne jest, aby pomiar nie był zakłócony i wykonywany całkowicie swobodnie.

Opisywana metoda polega na zastosowaniu kamer cyfrowych do rejestracji drżenia kończyn górnych, która pozwala dokonywać pomiarów przesunięcia wzdłuż wszystkich trzech osi kartezjańskiego układu współrzędnych X-Y-Z. Rejestracja drżenia kończyny dokonywana jest w oparciu o ruch punktowego źródła światła, znajdującego się na specjalnej opasce umieszczanej na ciele pacjenta. Możliwe jest zastosowanie jednego źródła światła lub dwóch źródeł jednocześnie. Rejestracji drżenia dokonują jednocześnie trzy kamery.

2.1. Opis stanowiska pomiarowego

Akwizycja danych odbywa się w założonej przestrzeni o strukturze sześcianu, gdzie długość każdego boku wynosi 0,3 m. Kubatura taka pozwala zarejestrować drżenia kończyny z dość dużym zapasem przestrzeni oraz wykorzystać stanowisko do innych pomiarów drgań. Schemat całego stanowiska zaprezentowano na rys. 1.

Ponieważ obraz uzyskany z każdej kamery jest dwuwymiarowy, zachodzi tu redundancja pomiaru, tzn. każda z kamer może zarejestrować drgania wzdłuż dwóch osi układu, ale głównie wykorzystuje się tylko jedną. Rozwiązanie to pozwala na uzyskanie pewności, że wektory pomiarów drgań są kompletne i nie nastąpiła utrata danych w wyniku np. zasłonięcia źródła światła badanym obiektem.



Rys. 1. Widok stanowiska do pomiaru drgań za pomocą kamer w przestrzeni X-Y-Z

Ponadto należy zauważyć, że dokładność kamery zależy od odległości od jej obiektywu (sensora), a badany obiekt może przemieszczać się dowolnie w założonym obszarze. W tej sytuacji każda kamera bazuje na wskazaniach kamery sąsiedniej w celu dobrania odpowiednich współczynników skalowania.

Kamery wyposażono w sensor KAC-1310 o rozdzielczości XVGA, czyli 1280x1024 oraz programowalny układ FPGA z modułem dokonującym akwizycji obrazu oraz jego przetworzenia na postać binarną i z detekcją punktowego źródła światła. Możliwy jest również podgląd *on-line* obrazu zarówno w skali szarości, jak i z detekcją plamki świetlnej. Każda z kamer pracuje z częstotliwością próbkowania plamki na poziomie 56 Hz, co przy drżeniach charakteryzujących PD na poziomie 4-7 Hz jest w zupełności wystarczające [2]. Praca kamer jest programowo zsynchronizowana w czasie tak, aby dokonywać odczytów wartości współrzędnych każdej kamery w tej samej chwili czasowej, co umożliwi rekonstrukcję trajektorii ruchu. Do pomiarów drżenia kończyny wykonano opaskę wyposażoną w diodę półprzewodnikową LED o podwyższonej intensywności emisji światła, z możliwością umieszczenia jej w różnych częściach kończyny. Masa źródła światła w odniesieniu do masy ciała jest pomijalnie mała, a możliwość umieszczania go w dowolnym miejscu na kończynie czyni zastaw uniwersalnym w zastosowaniu. Położenie kamer pozwala na możliwość bezkolizyjnego wprowadzania obiektu i dokonania pomiarów z dwóch kierunków, oznaczonych na rys. 1 odpowiednio jako D1 i D2. Na rys. 2 zaprezentowano zdjęcie stanowiska pomiarowego w trakcie przeprowadzania badania.

Rys. 2. Pacjent podczas pomiaru drżenia za pomocą kamer w przestrzeni pomiarowej

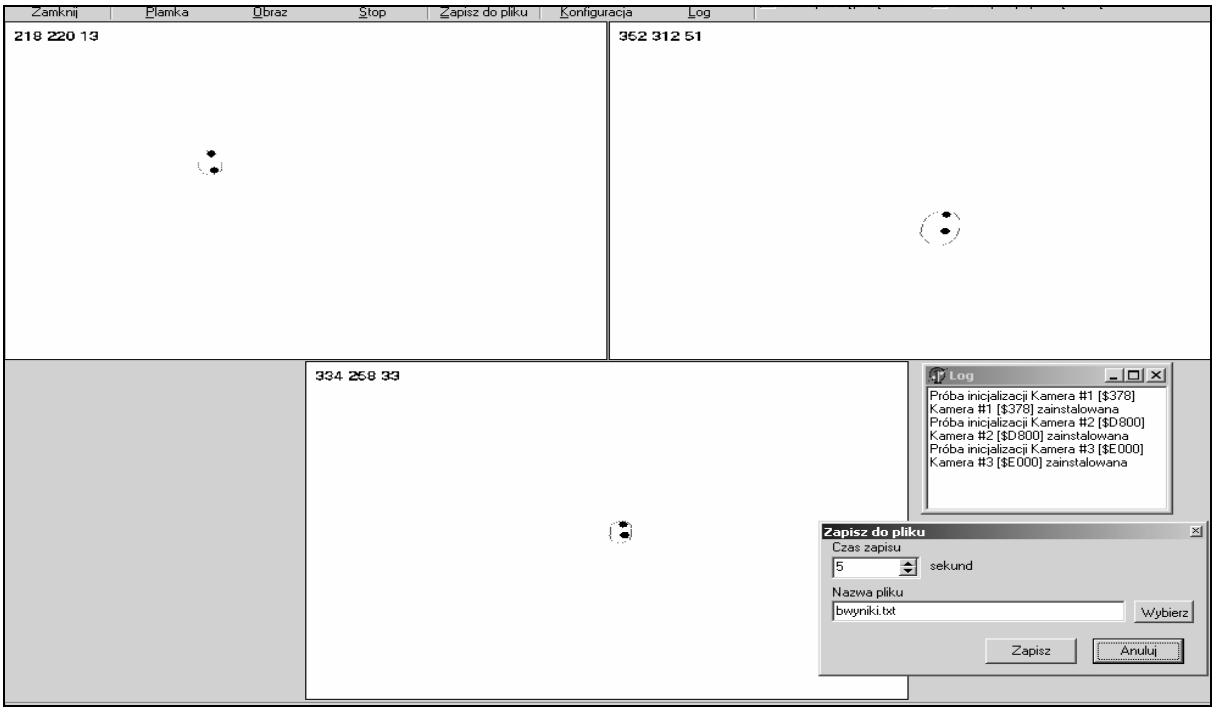
3. Wspomagana komputerowo wizualizacja i analiza danych z pomiarów drżenia w przestrzeni

Do obsługi akwizycji danych oraz konfiguracji parametrów kamer służy oprogramowanie sterujące, którego widok na panel główny przedstawiono na rys. 3. Za jego pomocą możliwa jest wizualizacja źródła światła w obszarze pomiarowym dla wszystkich trzech kamer jednocześnie, odczyt współrzędnych plamki i ich zapis oraz korekcja następujących parametrów kamer: wzmocnienie (zgrubne i dokładne), kontrast, czas ekspozycji plamki światła, próg jej detekcji, okres próbkowania oraz tryb pracy z obrazem dzielonym, w którym możliwa jest rejestracja drgania dwóch źródeł światła jednocześnie.

Dane uzyskane w procesie akwizycji zapisywane są następnie do pliku tekstowego, co umożliwia import danych i ich analizę w innych aplikacjach. Przykład zarejestrowanych danych z wykorzystaniem jednego źródła światła dla trzech kamer przedstawiono na rys. 4. Kolumny od lewej oznaczają odpowiednio: numer kolejnej próbki, wartość współrzędnej odciętych, wartość współrzędnej rzędnych, średnia, z której liczony jest środek ciężkości plamki, wskazanie zegara.

Kolumny od drugiej do piątej powtarzają się trzy razy, osobno dla każdej kamery.

Analizy uzyskanych w ten sposób danych dokonano w programie MATLAB. Aplikacja stworzona w tym pakiecie dokonuje wstępnej obróbki i filtracji górno-przepustowej sygnałów oraz wyznacza charakterystyki czasowe, częstotliwościowe, spektrogramy i gęstości widmowe mocy PSD dla każdej z osi oraz ich wypadkowej dla przemieszczenia, prędkości i przyspieszenia. W wyniku konsultacji z lekarzami aplikacja podaje szereg przydatnych przy analizie danych parametrów, takich jak: maksymalna, średnia i minimalna amplituda drżenia, dominująca częstotliwość i maksymalna wartość PSD, zawartość procentowa maksimum PSD w stosunku do całości PSD w otoczeniu 1 Hz oraz różnice i stosunki wyznaczonych amplitud. Wszystkie przedstawione parametry obliczane są dla każdej osi oraz dla ich wypadkowej.



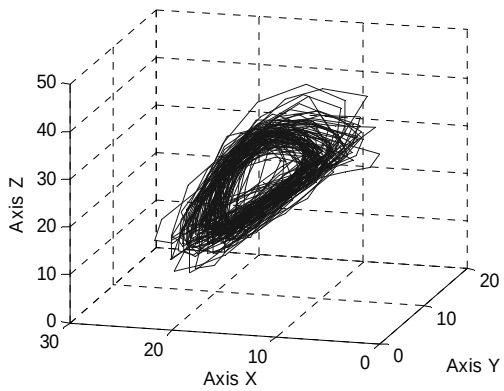
Rys. 3. Panel wizualizacji, kontroli i zapisu danych uzyskanych z trzech kamer

19	331	140	24	5543685	317	333	16	5543700	260	262	35	5543715
20	180	159	12	5543756	314	351	20	5543790	248	278	31	5543820
21	169	174	17	5543845	314	344	17	5543860	257	268	36	5543875
22	180	160	14	5543896	317	324	14	5543910	260	261	34	5543925
23	174	166	15	5543967	305	349	14	5543985	250	278	33	5543995
24	337	133	23	5544055	319	329	16	5544070	261	261	35	5544085
25	179	161	19	5544127	316	350	20	5544140	252	276	35	5544155
26	175	170	19	5544215	166	182	19	5544197	256	272	35	5544210
27	345	126	20	5544250	319	326	14	5544265	262	260	35	5544280
28	175	169	19	5544322	307	354	14	5544335	251	277	33	5544350
29	174	171	18	5544392	318	339	16	5544410	261	264	33	5544420
30	335	140	22	5544480	312	354	17	5544495	252	277	33	5544510
31	171	74	19	5544551	318	344	18	5544565	260	266	35	5544580
32	331	150	25	5544675	307	352	13	5544690	254	273	33	5544705

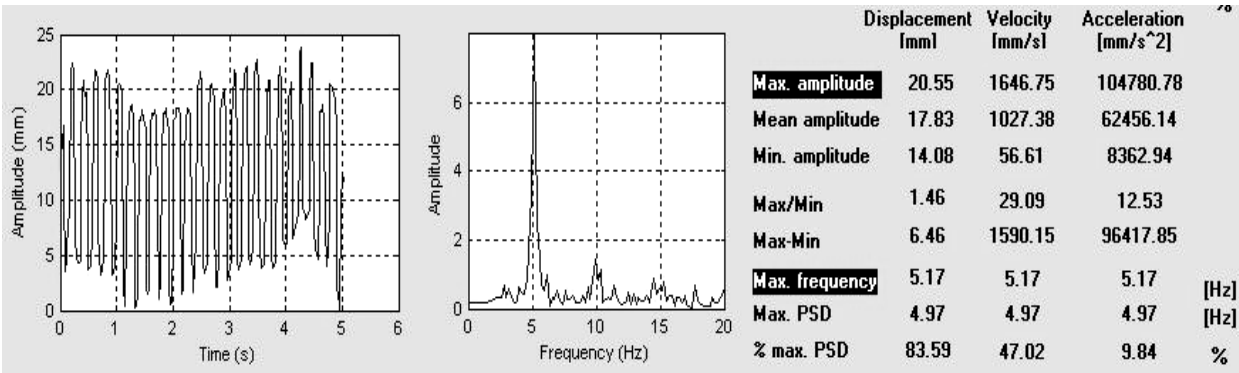
Rys. 4. Dane w postaci pliku tekstowego, pozyskane w trakcie zapisu

Rysunek 5 przedstawia odtworzenie trajektorii drżenia pacjenta w przestrzeni pomiarowej. Aplikacja posiada również bardzo intuicyjny interfejs graficzny, co czyni ją przyjazną dla użytkownika. Fragment interfejsu wraz z charakterystykami dla przesunięcia i wyznaczonymi współczynnikami dla jednej osi przedstawia rys. 6.

Zastosowania opisanego stanowiska badawczego nie ograniczają się jedynie do pomiarów drżenia w sensie medycznym. W przestrzeni pomiarowej można dokonywać akwizycji drgań dowolnego punktowego źródła światła dla obiektu, którego gabaryty pozwalają na całkowite bądź tylko częściowe umiejscowienie go w przestrzeni roboczej stanowiska. Dobrym przykładem może być tutaj ocena jakościowa drgań ramienia i/lub kiści manipulatora oraz jego prędkości liniowej i kątowej.



Rys. 5. Odtworzenie trajektorii drżenia kończyny pacjenta w przestrzeni



Rys. 6. Charakterystyka czasowa i częstotliwościowa drżenia oraz wyznaczone wyniki dla danego pacjenta

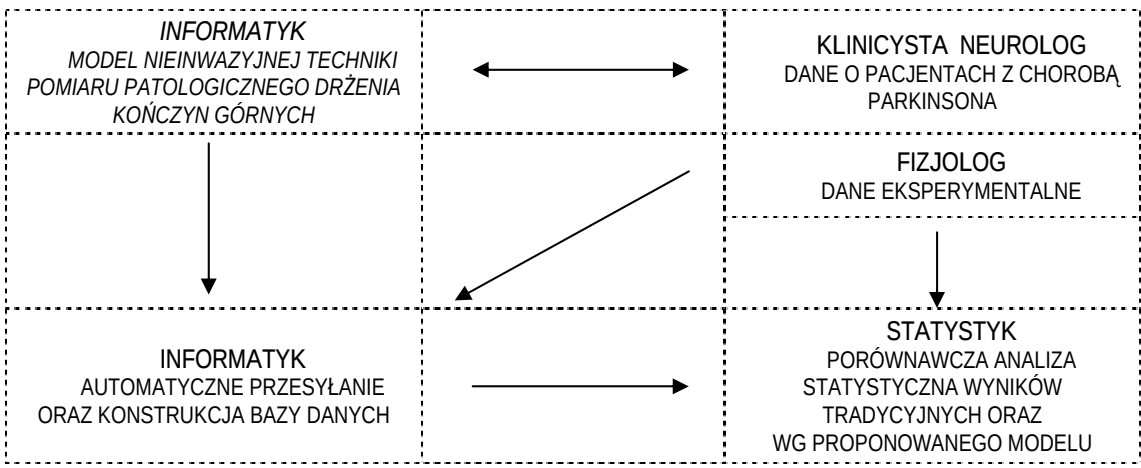
4. Wnioski

Wspomaganie komputerowe pomiarów drgań oraz możliwość ich współpracy z nowoczesnymi i specjalizowanymi urządzeniami peryferyjnymi znacząco poszerza możliwości akwizycji samych danych oraz ich wizualizacji i dokładnej analizy.

Stosowane obecnie metody pomiarowe drżenia w postaci przykładów pisma odręcznego czy rysowania spiral Archimedesa umożliwiają zgrubną i bardzo subiektywną ocenę drżenia. Zastosowanie tablic cyfrowych jest rozwiązaniem pozwalającym określać wielkość amplitudy drżenia oraz jej częstotliwości, co umożliwia obiektywną ocenę i umożliwia przeprowadzenie dokładniejszej klasyfikacji. Choroba Parkinsona cechuje się drżeniem spoczynkowym, które ulega redukcji podczas wykonywania określonych czynności ruchowych. Opisana metoda wykorzystująca kamery cyfrowe pozwala na pomiar drżenia spoczynkowego bez zakłóceń. Dodatkowym atutem jest tutaj możliwość rejestracji w przestrzeni X-Y-Z, co pozwala na analizę drgań wzdłuż każdej z osi kartezjańskiego układu współrzędnych. Częstotliwości drgań w przypadku chorób objawiających się drżeniem kończyn górnych z reguły nie przekraczają wartości kilkunastu herców, co umożliwia swobodną rejestrację i analizę różnego rodzaju schorzeń za pomocą przedstawionych powyżej metod.

Literatura cytowana

1. Bain P. G.: Clinical Measurement of Tremor, Movement Disorders, vol. 13, 1998.
2. Elble R. J., Brilliant M., Leffler K., Higgins, C.: Quantification of Essential Tremor in Writing and Drawing, Movement Disorders, vol. 11, 1996.
3. Elble R. J., Sinha R., Higgins Ch.: Quantification of tremor with a digitizing tablet, Journal of Neuroscience Methods, vol. 32, 1990
4. Friedman A.: Choroba Parkinsona, α-Medica Press, Bielsko Biała, 1999.
5. Hollerbach J.: M. An Oscillation Theory of Handwriting, Biological Cybernetics, vol. 39, 1981.
6. Pullman S. L.: Spiral analysis: A new technique for measuring tremor with a digitizing tablet, Movement Disorders, vol. 13, 1998
7. Rothwell J. C.: Physiology and Anatomy of Possible Oscillators in the Central Nervous System, Movement Disorders, vol. 13, 1998.
8. Szczudlik A.: Farmakologiczne i chirurgiczne metody leczenia choroby Parkinsona, Choroba Parkinsona, XVI Zimowa Szkoła Instytutu Farmakologii PAN, Instytut farmakologii PAN, pp. 35-44, 1999.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 301-306.

ARE THE EARTH MAGNETIC FIELD AND SCHUMANN RESONANCE RELATED TO GLOBAL HUMAN ACTIVITY?

CZY POLE MAGNETYCZNE ZIEMI I REZONANSE SCHUMANNA MAJĄ ZWIĄZEK Z GLOBALNĄ AKTYWNOŚCIĄ CZŁOWIEKA?

STANISŁAW MICEK*, GRZEGORZ MICEK**

**Zakład Doświadczalnej Fizyki Komputerowej, Instytut Fizyki, Uniwersytet Jagielloński, ul. Reymonta 4, 30-059 Kraków, ufmicek@cyf-kr.edu.pl*
***Zakład Rozwoju Regionalnego, Instytut Geografii i Gospodarki Przestrzennej, Uniwersytet Jagielloński, ul. Grodzka 64, 31-044 Kraków, g.micek@geo.uj.edu.pl*

Abstract. Weak oscillating electromagnetic fields at extremely low frequency (ELF) induced by Schumann (AC) resonance and static (DC) magnetic fields are investigated in exerting a possible effect on biological systems on the Earth. An attempt is made to explain higher human activity on the Earth as the result of ion cyclotron resonance (ICR) conditions. The effect of cyclotron resonance on heart rate variability (HRV) and ion solution of NaCl is also shown. From IGRF model of DC magnetic field on the Earth and from frequency of Schumann resonance AC, regions of cyclotron resonance were estimated. These regions can be correlated with birth rate, population structure and armed conflicts. The result of the analysis indicates that a global human activity could be associated with influence of magnetic fields.

Key words: HRV, ICR, Schumann resonance, IGRF model, geomagnetic field, birth rate, death rate, HIV/AIDS, armed conflict

Streszczenie. W pracy badany jest wpływ na układy biologiczne Ziemi słabych pól elektromagnetycznych o ekstremalnie niskich częstotliwościach (ELF) wytworzonych przez rezonanse Schumanna (AC) w obecności stałych (DC) pól magnetycznych. Przedstawiono próbę wyjaśnienia większej aktywności ludzi na Ziemi przy pomocy efektu jonowego rezonansu cyklotronowego (ICR). Pokazano też wpływ rezonansu cyklotronowego na zmienność rytmu serca i na związek jonowy NaCl. Przyjmując stałe pole magnetyczne na Ziemi zgodnie z modelem IGRF oraz naturalne zmienne pole magnetyczne pochodzące z rezonansów Schumanna, znaleziono obszary występowania rezonansu cyklotronowego. Obszary te można korelować z miejscami o większym współczynniku urodzeń, strukturą wiekową ludności i z konfliktami zbrojnymi. Wynik analizy wskazuje na to, że globalna aktywność człowieka może być połączona z wpływem pól magnetycznych.

Słowa kluczowe: rytm serca, jonowy rezonans cyklotronowy, rezonans Schumanna, współczynnik urodzeń, konflikty zbrojne

1. Introduction

Recent studies in the solar-terrestrial physics led to the recognition of the role of the geomagnetic field as one of the most important characteristics of human environment. The role of magnetic fields in the dynamics of complex biological process is still far from being understood. Some attempts have been made to test different mechanisms of ELF (Extremely Low Frequency 1-50 Hz) magnetic field interactions experimentally. Several mechanisms have been proposed: Hall effect, ion cyclotron resonance ICR, ion parametric resonance, biological electron transfer, forces on endogenous magnetite particles, magnetochemistry (effect on free radical reactions due to electron spin reversals brought about by the magnetic field).

Ions travelling in the external DC magnetic field are deflected by the Lorentz force and generate an electric field. This phenomenon is known as the Hall effect [1], [2], [3].

Lednev [4] has proposed a mechanism of parametric resonance that would allow very weak magnetic fields, at the cyclotron resonance frequency for Ca^{2+} ions in the Earth's field, to induce biological effects. The ion parametric resonance is based on the idea that an ion weakly bound within a protein, notably Ca^{2+} within calmodulin, can be modelled as charge oscillator. An ambient DC magnetic field, such as that of the Earth, would cause Zeeman splitting of each vibrational level into two levels separated by the ion cyclotron frequency. When an alternating magnetic field is applied in parallel with a DC magnetic field, the resulting frequency modulation of levels will change the probability of ion transition from the ground state to the excited state.

Direct effects of applied magnetic fields on DNA are also possible [5]. The electron mobility along the π bonds in the central core of DNA strands can be surprisingly large and it is not known whether it is large enough to make transverse deflection of electron currents due to applied magnetic fields.

Modulators of K^+, Ca^{2+}, and Na^+ currents can influence the intrinsic rate of automatic cells. The passage of Na^+, K^+, Ca^{2+} and Cl^- ions across the cardiac plasma membrane (sarcolemma) through specialised voltage and ligant operated ion channels is the basis of normal excitability and conduction. The process that accounts for opening and closing of ion channels is called gating. Voltage gated ion channels are the largest group of cardiac ion channels. Gating currents have provided considerable insight into the function of ion channels [6]. Voltage gated ion channels may occupy one of two conductance states, open or closed. The direction of movement of a given ion through the open channel is determined not only by its electrochemical gradient but also by magnetic field [7]. Ion movement can be controlled by creating a specific relationship between the strength of the stable magnetic field B_{DC} (B_x,B_y,B_z) and the frequency of the AC field oscillation B_{AC} (B_x,B_y,B_z). The ion trajectory can be predicted by calculations. As it will be shown later at the cyclotron resonance, ion pathway is different than out of the resonance.

The purpose of this paper is to show experimental evidence for the Earth's cyclotron resonance effect on entire human population.

2. Cyclotron resonance

A charge q of mass m, which moves at velocity V in a constant (static DC) magnetic field B_{DC}, will follow in vacuum a circular trajectory of the radius R:

$$R = \frac{V}{B_{DC}} * \frac{m}{q} \tag{1}$$

The radius R for ions travelling with small velocity can be very small. The cyclotron frequency f_c will be given by $f_c = (q/m) * B_{DC}/(2\pi)$. It has been noted that cyclotron frequencies of many physiologically important ions fall below 100 Hz in the geomagnetic field (50 μT). For example, $f_c = 38.4$ Hz for $^{40}Ca^{2+}$ ion is seen at B_{CD}=50 μT magnetic field.

Such movement of ions seems to be not possible in the dense, collision-dominated fluids of biological materials. Furthermore, ions in biological fluids are normally hydrated, and the cyclotron frequency depends on the number of water molecules (and their total mass) in the sheath of solvent molecules of each ion. Despite the apparent implausibility of cyclotron resonance as a mechanism of interaction between magnetic field and biological system, many experiments have shown a response to combinations of static and alternating magnetic fields. This response reaches, in different systems, either a maximum or minimum at the cyclotron frequency for various physiologically important ions.

According to the Lorentz force the ion trajectory can be calculated for a given constant and oscillating magnetic field B_{DC} (B_x,B_y,B_z) and B_{AC} (B_x,B_y,B_z). Calculated trajectory of the Cl^- ion presented in Fig. 2 depends on the frequency of AC field. For given B_{DC}(15,0,47) μT and B_{AC}(5,0,0) μT we can see two trajectories for 15 Hz and 21 Hz. The ion path for 21 Hz, close to the cyclotron resonance, is concentrated at the origin. For 15 Hz frequency and the initial velocity of V(0.1,0,0)m/s, Cl^- ion is reaching the distance above 150 μm in 5 seconds. Ions collisions change this value but the difference in the path length between 15 Hz and 21 Hz remains.

The length of the ion path depends on its time of life. After time of life, an ion reaches the distance:

$$d = \sqrt{X_o^2 + Y_o^2 + Z_o^2} \tag{2}$$

Trajectory calculations made in 3D space shown in Fig. 1 indicate the pronounced minimum in d at the cyclotron resonance frequency.

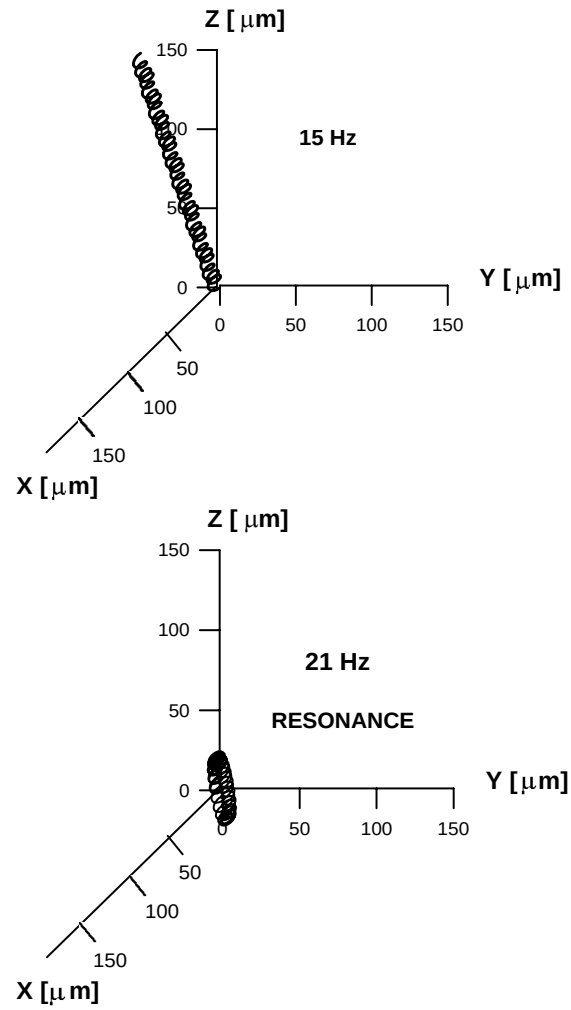


Fig. 1. Trajectory calculation of a free Cl^- ion moving with an initial velocity V(0.1,0,0) m/s in constant B_{DC}(15,0,47) μT and alternating B_{AC}(5,0,0) μT magnetic field at the frequency 15 Hz (upper trajectory) and at 21 Hz (resonance). At the resonance Cl^- ion remains closer to the centre of XYZ axis.

3. Heart rate variability (HRV)

The sinus node has the highest intrinsic rate and acts as the dominant pacemaker and it is the central control element of the autonomic regulation. It generates regular heartbeat rate seen in electrocardiogram. Intervals between two consecutive R waves correspond to one cardiac interbeat interval RR. The change of RR interval in ECG describes the heart rate variability (HRV). The HRV is a very sensitive indicator of living organisms. The analysis of RR time series exhibits different oscillating sources of the variability of heart beat generation. The different regions in the power spectrum are related to special physiological phenomena. The high frequency band (HF) from 0.15 to 0.45 Hz [8] represents the modulation of the vagal activity especially influenced by respiration and low frequency band (LF) from 0.03 to 0.15 Hz

reflects modulation of sympathetic or parasympathetic tone by baroflex activity (blood pressure regulation) [9].

The correlation dimension D2 can be derived from methods of non-linear dynamics, also called chaos theory, which describe complex processes and complicated interactions. D2 should be able to record additional information about the state and temporal changes of the autonomic system. Therefore, we have applied the correlation dimension method to distinguish heart rate dynamics in ELF magnetic field. Correlation dimension of RR intervals is a better predictor of risk of sudden cardiac death than a stochastic measure, such as the standard deviation [10]. It was found that reduction of the D2 value in HRV below 1.2 precedes lethal arrhythmia by hours [11]. Calculation of D2 for time series similar to sine wave gives D2=1. For more complicated series D2 > 1.

4. Measurements of HRV

Standard ECG was measured in 8 channels by 12-bit converters with the frequency 200 Hz and stored in 4-minute time series. The magnetic field of a randomized frequency was generated by Helmholtz coils (1 m^2). Subject and investigators were not informed about irradiation frequency and whether the field was switched on. Two independent computer systems: for control and data acquisition were used. The pulse rate of subject was controlled during measurements.

5. Correlation dimension D2 of HRV

For a given ECG time series, RR intervals were calculated. The point defined as R was estimated by looking at the first derivative in ECG time series in comparison to the given pre-set level determined in the preliminary experiment. As a result of measurements the RR data in time steps given by the time difference between consecutive QRS events was received. Next a reference multidimensional vector with dimension m was selected. Similarly the second vector of the same dimension m was constructed starting at a different point of the time series. This operation was performed sequentially for all other starting points. The difference between the reference vector and the second vector was determined and stored. Then reference vector was moved to the next point of the data until all possible vector differences were calculated and stored. The number of vectors whose lengths were lower than r given by C(r) was plotted on log-log scale (log C(r) vs. log r).

The correlation dimension D2 of the time series was calculated as a logC(r)/log r ratio. It was done by linear regression in the linear region of a log-log plot. Then calculations were repeated for the next m dimension. For some values of m, D2 converges if HRV follows deterministic chaotic dynamics. It means that increment of m was no longer correlated with an increase in D2 [12]. Algorithm rejected D2 where there was no linear part of log-log dependence or D2 was not converging. In one calculation 250 RR data points were used.

Measurements and calculations of D2 vs. frequency of magnetic field (Fig. 2) were made for ECG time series with average pulse rate from 61 to 72 bpm. They indicate that D2 has a maximum at frequency of about 3 Hz, 7 Hz and 10 Hz in ELF region. Minimum at 1.6 Hz, 4 Hz and 25 Hz point out that heart pulse rates behave more stable in the ELF magnetic field. Experimental data support the idea that both harmonics and subharmonics of Cl^- and Na^+ ions are involved in HRV.

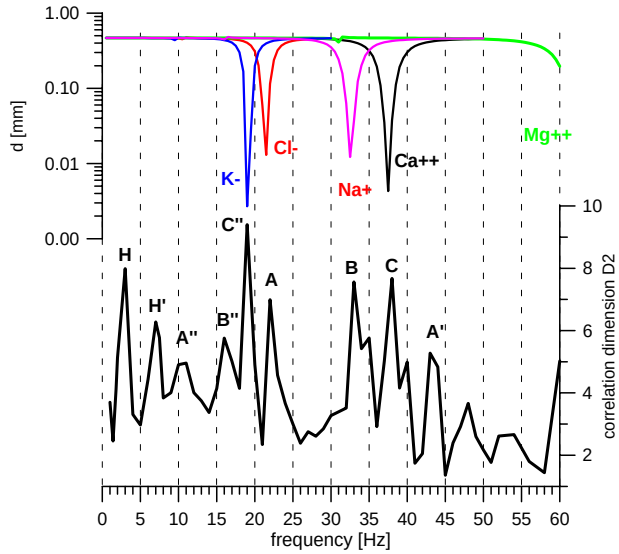


Fig. 2. Correlation dimensions D2 (lower curve) calculated from heart rate variability and model calculations (upper curves) of path length d versus magnetic field frequency. Initial ion velocity in calculations is V(0,0,0.1) m/s and B_{DC}(17,0,46) μT, B_{AC}(5,0,0) μT is a constant and alternating magnetic field. Three peaks A, B, C can be identified as Cl, Na, Ca and A'', B'', C'' as subharmonics (A' as harmonics of A).

A presented phenomenon is similar to the Hall effect. The Hall voltage is generated across the conductor perpendicular to both the direction of bulk current flow and to the magnetic field. A similar effect is observed for AC magnetic field. Ions are moving in accordance to current flow. The ion trajectory is very sensitive at the resonance to the relative direction DC and AC magnetic fields.

The comparison of trajectory calculations with the correlation dimension D2 is shown in Fig. 4. The ion identification was made according to m/q ratio. No identification was shown for frequency below 10 Hz. The mass of the ion in this frequency region is higher than 70 atomic mass units.

The haemoglobin with the mass 523 and charge 2^- can be identified as m/q = 523/2= 261.5 and for B_{DC} = 50 μT. The cyclotron resonance can be seen at about 3 Hz. It is described in Fig. 2 as H with a harmonics H'.

6. ICR mechanism in NaCl

The mechanism of Hall effect [13] at the ELF cyclotron resonance was tested in the solution of NaCl (p.a., 3 mole/L). Ions were moving in the NaCl solution according to the direction of X axis (current 2 mA) irradiated by alternating magnetic fields B_{AC}(0,0,5) μT or B_{AC}(0,0,45) μT and in the stable magnetic field B_{DC}(10,0,37) μT.

During the time of 40 s the voltage V(t) was measured in Y direction perpendicular to X. For each frequency of magnetic field two runs were made (20s+20s). In the first run only oscillating magnetic field B_{AC} was present, the stable magnetic field was zero (± 0.5 μT). In the second run the constant magnetic field B_{DC}(10,0,37) was turned on in the given phase of AC current. During the first period the voltage difference

V(1)-V(19) was calculated and in the second, the difference V(21)-V(40) was calculated (time (s) is in brackets). For the frequency of cyclotron resonance some difference between (V(1)-V(19)) and (V(21)-V(40)) was observed. This difference was measured in the random order of frequency, over 20 times for one frequency. Average values were plotted in Fig.3.

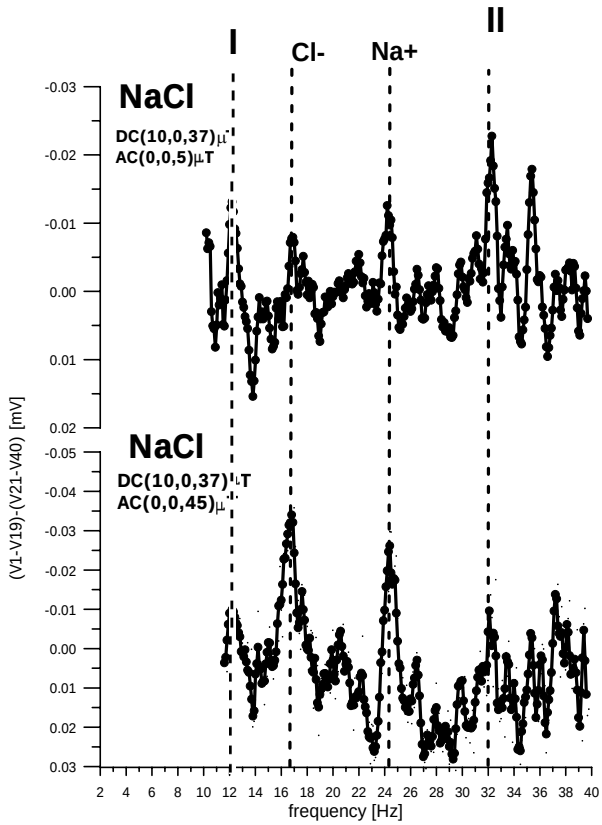


Fig. 3. ELF cyclotron resonance in NaCl for alternating magnetic field 5 μT (upper curve) and 45 μT (lower curve). Resonances for two ions: $^{35}Cl^-$, $^{23}Na^+$ are indicated by dotted lines. Lines I and II can be explained as ICR subharmonics of Na^+ and harmonics of Cl^-.

Two curves were made for different AC magnetic fields: 5 μT and 45 μT. The amplitude of peaks depends on the AC magnetic field. In measurements for high AC magnetic field (45 μT, lower curve in Fig. 3), peaks at lower frequency have higher amplitude whereas for low magnetic field (5 μT, Fig. 3 upper curve) is reversed. For each ion, there exists a maximum in Hall voltage for different strength of AC magnetic field. In the mass range under consideration the maximum is between 0.5 and 50 μT.

7. The Earth DC and AC magnetic field

The Earth magnetic field model (IGRF) was used for calculating the geomagnetic field DC components at any point of space from the Earth's surface up to the Moon's orbit [14]. Upon specifying the year, day of the year, and universal time as input parameters, the model calculates elements of the Earth magnetic field (DC). It also updates the coefficients of spherical harmonic expansions, approximating the Earth's internal magnetic field. That field was computed in a dipole approximation with the length of expansions automatically controlled to maintain the assumed precision. In our case, every point on the Earth was estimated with a 5-degree precision.

One of the biggest electromagnetic resonance systems generating oscillating magnetic field AC on the Earth is the cavity formed between the Earth and its ionosphere. These phenomena are known as Schumann resonances [15]. Lightning discharges in billions of coulombs provide a broad frequency spectrum modified by the resonance properties of the earth-ionosphere cavity system. The Schumann spectrum contains broad maximum at frequency about 8 Hz, 14 Hz, 21 Hz etc. The spectrum of Schumann resonance shown in Fig. 4 depends on the relative position of discharge source and receiver. It can be calculated by electrical model of cavity designed by A. Kułak [15]. In calculations of ICR area only Schumann peak frequency and amplitude of AC magnetic field derived by model was taken with a resonance width of 0.4 Hz. Three zones of thunderstorm with similar activity (in Africa, Asia and S. America) were used in calculations.

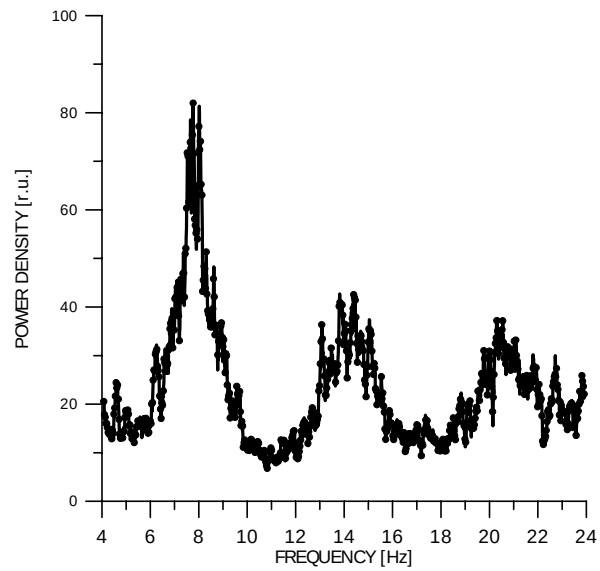


Fig. 4. Natural spectrum of Schumann resonance

8. Cyclotron resonance on the Earth

Variations of the Earth DC magnetic field and a frequency of AC Schumann resonance estimates a range of ions existing in cyclotron resonance conditions. They varied in m/e from 22 to 110. However, not all places on the Earth are under a cyclotron resonance (Fig 5).

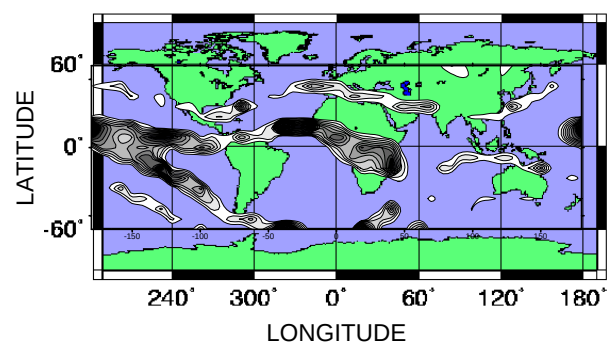


Fig. 5. Spatial distribution pattern of Cl^- cyclotron resonance on the Earth (m/q=35, 70 was used).

9. Comparison of ICR model with human activity

A set of data containing birth rates (number of births per 1000 people for over 210 countries was used for analysis (year 2002) [16]. Only Cl^- ions (A in Fig. 2) and a subharmonics (A'') in ICR were taken into calculations. Calculations of ICR indicate that average birth rate where cyclotron resonance is not observed is equal 21.3±1.8 whereas in resonance area it is 28.2±3.3. It means that in the countries with a cyclotron resonance the birth rate is higher by about 32% than in remaining countries. The birth rate may be sensitive to the amplitude of AC magnetic field of Schumann resonance. For 119 countries it is shown in Fig 6. The correlation coefficient between birth rate and AC magnetic field 0.386±0.12 is relatively high. For Africa the correlation coefficient is higher and equals 0.446±0.14.

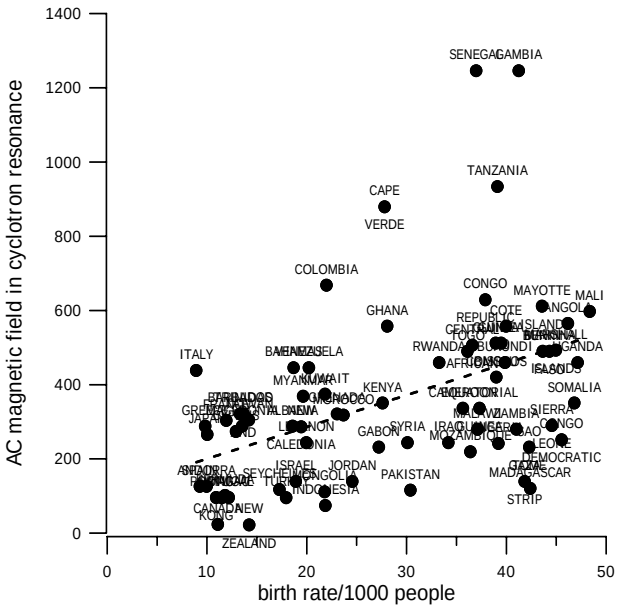


Fig. 6. Birth rate in different countries and amplitude of AC magnetic field from Schumann IRC on the Earth may indicate some relations.

Tab. 1. Average value of birth rate, death rate, population structure and radius of conflict in cyclotron resonance countries for Cl^- ions and out of ICR.

Averaged value data	{1} Countries under cyclotron resonance	{2} Countries free of ICR	{3} Ratio [1]/[2]	{4} Correlation coefficient in [1]
Birth rate	28.2 ±3.3	21.3 ±1.8	1.32	0.386 ±0.12
Death rate	11.5 ±1.5	9.12 ±0.8	1.26	0.227 ±0.12
Population Age >65 in %	6.55 ±0.82	7.75 ±0.68	0.85	-0.347 * ±0.11
Population Age <15 in %	34.71 ±4.37	29.8 ±2.63	1.16	0.371 ±0.11
HIV/AIDS Age 15-24	5.67 ±1.05	2.75 ±0.32	2.05	-0.131 ±0.18
Armed Conflict INT*R	732 ±100	410 ±150	1.78	0.365 ±0.1

* Correlation coefficient of population over 65 years of age has opposite sign than correlation for under 15. This is due to normalisation of each whole population to 100%.

Data set describing Armed Conflicts is taken from Dataset Book version "March 2004" [17] published by International Peace Research Institute, Oslo (PRIO) and Department of Peace and Conflict Research, Uppsala University. It defines intensity parameter INT of conflict and radius R of conflict. Intensity INT of conflict is defined as a two+one level assessment of the number of battle-related casualties per year in the conflict period covered by the observation, plus a special level indicating conflict history in low-intensity conflicts. The radius of conflict R assumes that all conflict areas are circles. Similarly Cl^- ion cyclotron resonance was assumed. The average value of INT*R for countries being in the cyclotron resonance conditions equals 732±100 whereas outside of resonance is 410±150.

Interesting results of HIV/AIDS events in a population in ages 15-24 [18] are seen in Table 1. The average value of HIV/AIDS in countries where we have ICR is higher (5.67) than in other places (2.75). However, no correlation between HIV/AIDS and the magnetic AC field is observed.

10. Conclusion

It is clear that biological effects of the Earth's magnetic field discussed above do not provide a clear answer to the question of whether ELF magnetic fields in cyclotron resonance conditions influence human activity. This is partly due to two problems: 1) Mechanism of ICR model for very low intensity magnetic field induced by Schumann resonance in a dense biological media is unknown, 2) Thermal fluctuations should destroy the weak biological effect of magnetic field conversion. Naturally occurring magnetic fields, which populations are exposed to, are 1-100 pT. The problem of low intensity demands non-linear and very sensitive model of ICR interaction.

Data of human activity are significant, because they refer to a whole population. However, exposure in ICR magnetic field may also be correlated with other factors such as social status. It is possible that we observe an indirect ion cyclotron resonance effect on humans by ELF electromagnetic field exposure of biological environment.

The Earth's DC magnetic field varies within a time and depends on the Sun's activity. The Sun's activity changes ELF wave propagation modulating the Schumann frequency spectrum. All these can affect human activity on the globe. This indicates that observed effect of ICR may depend also on the Sun's activity. Responsible for AC electromagnetic field Schumann resonance excitations called Q-bursts have approximately 1 second duration and typically occur every few minutes. The temporal characteristic of Q-bursts is carrying a specific code in the pulse shape according to the position of a source. However, the principal excitation mechanism of the Earth-ionosphere cavity is lighting occurring at a mean rate of approximately 100 strokes per second. They have a similar frequency spectrum as a Q-burst. An influence of electromagnetic pulses generated in the laboratory, similar to Q-bursts on heart rate has been observed [18]. The natural ELF electromagnetic field is a mixture of Q-burst and lighting. It varies according to the condition of the Earth-ionosphere cavity.

A nonlinear mechanism based on the quantum interference model developed by Binhi [19] can explain some nonlinearity of the transduction mechanisms involved in

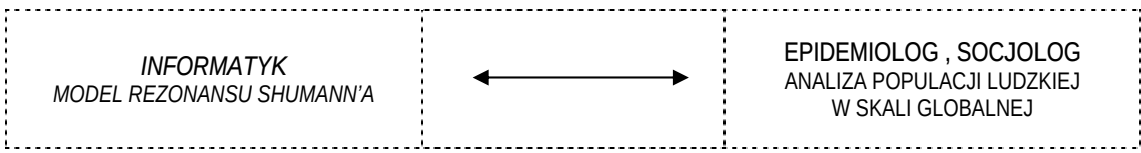
observed magnetobiological effect. However, the model works for 10-100 μT magnetic field and 10-100 Hz frequency band.

The cyclotron resonance model predicts the frequency of peaks in D2 HRV spectrum for Na^+, K^+, Ca^{2+}, Cl^- ions together with their harmonics (Fig. 2). Membrane transporters such as the Na^+-Ca^{2+} exchanger cause net ion movement across the cell membrane and may contribute to cellular activity automatically [20]. The major difference between cyclotron resonance condition and non-resonance condition lies in the charge distribution of irradiated media. The charge distribution generates the local potential. The potential perturbs ion flux and may cause profound changes in the metabolism of affected cells.

The present study carries considerable weight not only because of the large number of subjects in the whole population, but also because of relatively high correlation coefficient between AC magnetic field and birth rate reaching, for Africa, the value of 0.446±0.14.

References

1. Micek S L., Surdacki A., Dubiel J. S.: Heart Rate Variability in Ultra Low Frequency Magnetic Field, 8th National Symposium of Radio Sciences, 80 and 241, Wrocław 1996.
2. Micek S.: Wpływ pola magnetycznego na transport jonów w organizmie, V Krajowa Konferencja Modelowanie Cybernetyczne Systemów Biologicznych, 179, Kraków 2000.
3. Balcavage W. X., Alvager T., Svez J., Goff C. W., Fox M. T., Abdullyava S., King M.W.: A Mechanism for Action of Extremely Low Frequency Electromagnetic Fields on Biological Systems, Biochemical and Biophysical Research Communication, 222: 374,1996.
4. Lednev V. V.: Possible mechanism for the influence of weak magnetic fields on biological systems, Bioelectromagnetics, 12:71-75, 1991.
5. Stemp F. D. A., Arkin M. R., Barton J. K.: Electron transfer between metallointercalators bound to DNA: spectral identification of the transient intermediate, Journal of the American Chemical Society, 117: 2375-2376,1995.
6. Zamponi G. W., French R. J.: Diethylamine and phenol mimic separate modes of lidocaine block of sodium channels from heart and skeletal muscle, Biophys. J., 65: 2335-2347, 1993.
7. Consuelo Morgado-Valle, Verdugo-Diaz L., Gracia D.E., Morales-Orozco C., Drucker-Colin R.: The role of voltage-gated Ca^{2+} channels in neurite growth of cultured chromaffin cell induced by extremely low frequency (ELF) magnetic field stimulation, Cell Tissue Res, 291: 217-230, 1998.
8. Hirsch J. Y., Bishop R.: Respiratory sinus arrhythmia in humans: how breathing pattern modulates heart rate, Am. J. Physiol, Heart Circ. Physiol ,10: 241, H620, 1981.
9. Preiss G., Polosa C.: Patterns of sympathetic neurone activity with Mayer waves, Am. J. Physiol., 226: 724-730, 1994.
10. Voss A., Kuthis J., Kleiner H.J., Witt A., Wesel N., Saparin P., Osterziel K.J., Schurath R., Dietz A.: The application of methods of non-linear dynamics for the improved and predictive recognition of patients threatened by sudden cardiac death, Cardiovascular Research, 31: 419-433,1995.
11. Skinner J. E., Pratt C. M., Vybiral T.: A reduction in correlation dimension of heartbeat intervals precedes imminent ventricular fibrillation in human subjects, American Heart Journal, 125: 731-743, 1992.
12. Elbert T., .Ray W. J, Kowalik Z. J., Skinner J. E., Graf N., Birbaumer K. E.: Chaos and Physiology: Deterministic Chaos in Excitable Cell Assemblies, Physiological Reviews, 74: 1-47, 1984.
13. Micek S., Ilnicka-Micek A.: The Cyclotron Resonance for Action of Extremely Low Frequency Magnetic Field on Heart Rate Variability, Structures – Waves – Biomedical Engineering, Vol. X, 88-96, Kraków 2001.
14. Tsyganenko N. A.: Quantitative models of the magnetospheric magnetic field: Methods and results. Space Sci. Rev., 54,: 75-186, 1990.
15. Kułak A., Zięba S., Micek S., Nieckarz Z.: Solar variations in extremely low frequency propagation parameters: 1. A two-dimensional telegraph equation (TDTE) model of ELF propagation and fundamental parameters of Schumann resonances, J. Geophys. Res., 108 (A7): 1-10, 2003.
16. http://www.phrasebase.com/countries
17. http://www.geoforum.no/kartdag/kartdag.htm Wallensteen, Peter & Margareta Sollenberg: “Armed Conflict 1989–2000”, Journal of Peace Research 38(5): 629–644, 2001.
18. Micek S., Nieckarz Z., Kułak A., Michalec A., Zięba S.: Czy aktywność naturalnych pól elektromagnetycznych niskiej częstotliwości wpływa na rytm serca? Structures – Waves – Biomedical Engineering, vol. V: 191-196, Kraków 2000.
19. Binhi V. N., Savin A. V.: Molecular gyroscopes and biological effects of weak extremely low-frequency magnetic fields, Phys. Rev E, 65:051912, 2002.
20. Owsiak D., Burzynski P., Micek S.: Modelling of Linked Sinoatrial Node Cells System, Biosignals and Biomechanics, vol. XI: 197-204, 2002.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 307-310.

LOKALIZACJA KLUCZOWYCH CECH ROZPOZNAWANYCH OBIEKTÓW NA PODSTAWIE ANALIZY ŚCIEŻKI WZROKOWEJ OBSERWATORA

LOCALIZATION OF THE OBJECT'S SALIENT FEATURES BASED ON VISUAL PATH ANALYSIS

ZBIGNIEW MIKRUT, PIOTR AUGUSTYNIAK

Katedra Automatyki Akademii Górniczo-Hutniczej, Kraków, zibi@agh.edu.p,l august@agh.edu.pl

Streszczenie. W artykule przedstawiono metodykę badań, których celem jest wyznaczenie istotnych (z punktu widzenia rozpoznawania) cech obiektów. Odpowiednie eksperymenty przeprowadzono przy pomocy aparatury OBER2, służącej do rejestracji ruchu oczu człowieka. Zaplanowano i zrealizowano eksperymenty rozpoznawania ręcznie pisanych cyfr oraz opisano metodę lokalizacji cech na podstawie dwóch wskaźników.

Słowa kluczowe: cechy obiektów, lokalizacja cech, ścieżka wzrokowa, sygnał okoruchowy

Abstract. In the present paper the methodology of the salient features detection is presented. The experiments were executed using the OBER2 eye-tracker. During the particular experiment students to be examined had to recognize the more or less blurred images of handwritten digits. The salient features localization method is based on two factors derived from the visual path analysis.

Key words: salient features, feature localization, eye tracking, visual path analysis

1. Wstęp

Modelowanie fragmentów systemu wzrokowego człowieka i zwierząt wyższych może być realizowane na różnych poziomach szczegółowości, począwszy od modelu rozmieszczenia receptorów i pól receptorowych (transformacja log-polar [5]) poprzez analizę przepływu informacji traktem wzrokowym [7] na ogólnych aspektach rozumienia obserwowanej sceny skończywszy [6]. Celem niniejszej pracy jest zbadanie, czy jest możliwa lokalizacja cech obiektów, na jakie zwraca uwagę człowiek podczas procesu rozpoznawania, na podstawie analizy ścieżek wzrokowych jego oczu.

Do badań wykorzystano urządzenie OBER2, skonstruowane na Politechnice Śląskiej w Gliwicach [4]. Badanemu zakłada się rodzaj okularów bez szkieł („gogle"), na których są umieszczone dwie pary oświetlaczy podczerwieni oraz tyle samo odbiorników odbitego promieniowania. Diody oświetlają gałkę oczną, a odbiorniki rejestrują odbite promieniowanie. Odbywa się to niezależnie w dwóch wzajemnie prostopadłych osiach. Ruch gałki ocznej powoduje zmiany natężenia odbitego promieniowania.

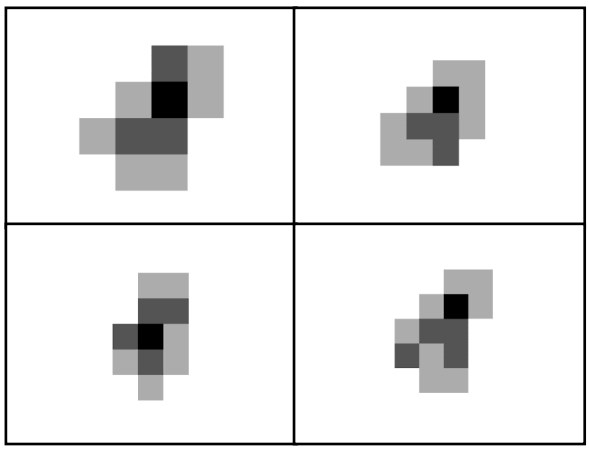
Sygnały z „gogli" dochodzą do przetwornika analogowo-cyfrowego, następnie są zapisywane w buforze urządzenia i transmitowane łączem szeregowym do komputera klasy PC. Aparatura OBER2 dokonuje akwizycji z maksymalną częstotliwością próbkowania 2 kHz przy 12 bitowej rozdzielczości kwantowania.

Konstruktorzy aparatury OBER2 dostarczają wraz z urządzeniem podstawowe oprogramowanie pracujące pod systemem DOS i umożliwiające dynamiczną kalibrację oraz akwizycję ścieżki wzrokowej podczas prezentacji statycznych obrazów. W Laboratorium Biocybernetyki AGH w Krakowie powstała rozszerzona wersja oprogramowania, działająca w systemach Windows. Umożliwia ona kalibrację statyczną (badany ma za zadanie obwieść wzrokiem wyświetlony na monitorze prostokąt) oraz wielokrotną akwizycję ścieżki wzrokowej w trakcie ściśle zaplanowanego eksperymentu. Pojedynczy eksperyment może składać się z pokazu filmu video lub prezentacji sekwencji różnych obrazów, przy czym zarówno kolejność, jak i czasy ekspozycji są z góry zadane w odpowiednim pliku konfiguracyjnym. Tę ostatnią możliwość wykorzystano w opisywanych badaniach.

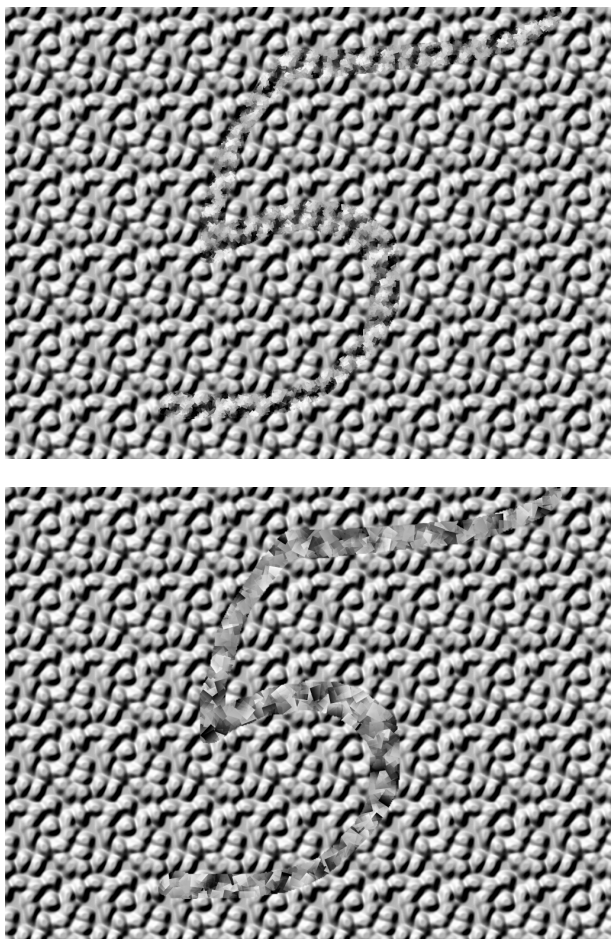
2. Eksperymenty

Cel pracy, czyli badanie możliwości lokalizacji istotnych cech obiektów, był realizowany na przykładzie rozpoznawania ręcznie pisanych cyfr. Obydwa te problemy (lokalizacji cech i rozpoznawania) są intensywnie badane, lecz nie do końca rozwiązane [1].

W celu uniknięcia sytuacji rozpoznawania znaku „jednym rzutem oka" zaplanowano dwie serie eksperymentów, różniące się rodzajem prezentowanych sekwencji obrazów. Prezentacje polegały na kolejnych pokazach obrazów o coraz większej wierności odwzorowania. W serii pierwszej zastosowano tzw. pikselizację, czyli uśrednianie całego obrazu znaku (czarny znak na białym tle) w oknach o zmiennej wielkości. Na rys. 1 przedstawiono kilka kolejnych obrazów cyfry „1". W serii drugiej znaki „wyłaniały się" z obrazu tekstury, czyli były rysowane przy pomocy tekstur coraz bardziej różniących się od tła (por. rys. 2).



Rys. 1. Przykład serii obrazów (cyfra „1") o malejącym stopniu trudności rozpoznawania otrzymanych metodą „pikselizacji"



Rys. 2. Dwa obrazy (cyfra „5") o malejącym stopniu trudności rozpoznawania otrzymane metodą „teksturową".

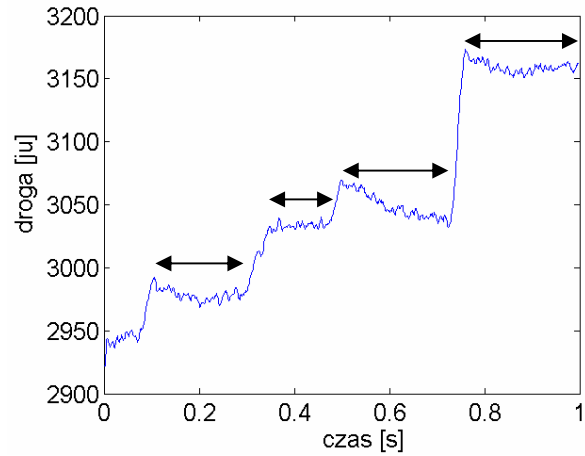
Eksperymenty polegały na przeprowadzeniu kalibracji statycznej (por. pkt 1), sprawdzeniu jej wyników, a następnie na rejestracji ruchów oczu podczas pokazu kolejnych obrazów danego znaku. Czasy ekspozycji kolejnych obrazów zmieniały się od około 5 s dla projekcji „najtrudniejszych" do ok. 2 s dla obrazów mało zakłóconych. Urządzenie OBER2 dokonywało akwizycji poszczególnych składowych ścieżki wzrokowej z częstotliwością 750 Hz.

3. Analiza wyników

W poprzednich pracach autorów głównym kryterium zainteresowania badanego był czas koncentracji uwagi na danym fragmencie obrazu [2], [3]. Podczas analizy sygnałów ruchu oczu, przeprowadzonej w ramach niniejszej pracy, zdecydowano się (głównie z przyczyn wizualizacyjnych) na zastosowanie tego kryterium w formie wydzielenia „obszaru zainteresowania", czyli wyodrębnienia fragmentów sceny, na których oko obserwatora (przez pewien czas i w sposób ciągły) *najdłużej* wykonywało tylko ruchy z małą prędkością.

Postać drugiego kryterium wyniknęła z metody prowadzenia eksperymentów. Podczas przeglądania wyników okazało się, że badani – nie będąc pewni swojego rozpoznania – wielokrotnie powracali wzrokiem do „niepewnych" fragmentów obrazu cyfry. Można postawić hipotezę, że cechy te stanowią elementy różnicujące poszczególne cyfry. Tak więc jako drugie kryterium przyjęto „liczbę powrotów" uwagi badanego w dany rejon sceny. W ten sposób wyznaczono drugi obszar sceny o szczególnej zawartości informacyjnej – obszar, dla którego „liczba powrotów" była największa.

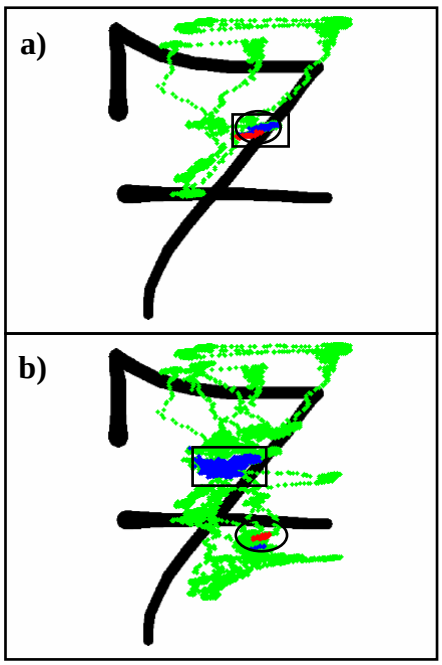
Algorytm obliczania obu wskaźników opiera się na analizie chwilowej prędkości ruchu oka, wspomaganej informacją o jego aktualnym położeniu. Przykładowy przebieg drogi chwilowego punktu koncentracji uwagi jednego oka pokazano na rys. 3. Jak widać, można względnie łatwo wyodrębnić odcinki ruchu wolnego, odpowiadającego obserwacji danego fragmentu sceny (na rys. 3 fragmenty te zaznaczono strzałkami) oraz szybkiego, oznaczającego przenoszenie punktu koncentracji uwagi w inny obszar.



Rys. 3. Fragment rejestracji drogi punktu koncentracji uwagi (oko prawe). [ju] oznacza „jednostki umowne" współrzędnych obserwowanej sceny

Punkty odpowiadające kolejnym odcinkom ruchu wolnego tworzą kolejne obszary skupienia punktu koncentracji uwagi. Po zakończeniu wyszukiwania, określane są granice każdego

z obszarów, a następnie – dla każdego znalezionego obszaru – sprawdzane jest, ile razy dany obszar nałożył się na inny. W końcu wybierany jest obszar największy oraz obszar o największej liczbie nałożeń, czyli powrotów punktu koncentracji uwagi. Obszary te to wspomniane wyżej wskaźniki lokalizacji cech rozpoznawanego obiektu. Zostają one naniesione na źródłowy obraz cyfry. Przykład lokalizacji cech pokazano na rys. 4.



Rys. 4. Wizualizacja dwóch wskaźników lokalizacji cech: obszaru najdłuższego skupienia uwagi (oznaczony elipsą) oraz obszaru o największej liczbie powrotów „uwagi" obserwatora (obszar oznaczony prostokątem): a) do momentu rozpoznania (1 obraz), b) dla całego eksperymentu. Obraz b) został uzyskany w wyniku analizy rejestracji osoby 5 (por. tab. 1).

Na podstawie wizualizacji możliwe jest określenie kluczowych cech, na jakie zwracał uwagę badany. W tabeli 1 zestawiono przykładowe wyniki uzyskane od 10 badanych osób dla rozpoznawania cyfr „1" i „7". Wyniki te otrzymano, analizując cały eksperyment, włącznie z trajektoriami zarejestrowanymi po rozpoznaniu obiektu (jak na rys. 4b). Wybrano akurat tę parę cyfr, ponieważ są one dość podobne – szczególnie na początkowych etapach prezentacji. Obszar obrazu cyfry podzielono na 4 podobszary oznaczone kolejnymi literami alfabetu (por. rys. 5):

a – środek obiektu, w przypadku cyfry „7" odpowiada poziomej kresce,
b – górna część cyfry,
c – część lewa (okolice daszka),
d – podobszar poza obiektem.

Pomimo niewielkiej (statystycznie) liczby badanych i eksperymentów można pokusić się o wstępną interpretację zestawionych wyników.

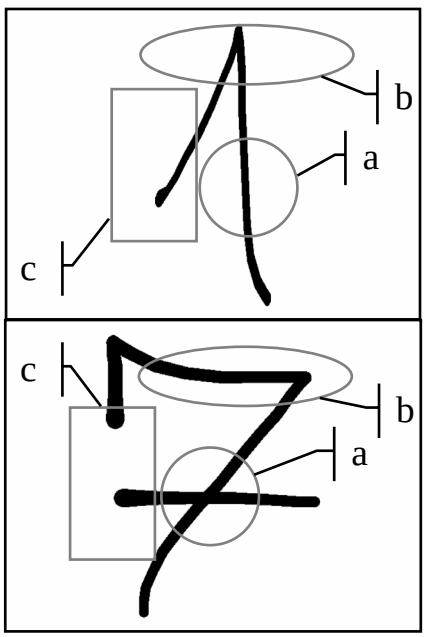
W ponad połowie eksperymentów badani wybierali takie same cechy dla obu wskaźników, czyli obszar skupienia wypadał w pobliżu obszaru częstych powrotów punktu koncentracji uwagi. W tej grupie dwie osoby (nr 9 i 10) szukały cech z użyciem obu metod w tym samym obszarze w przypadku obu cyfr. Był to obszar „b", czyli górna część znaku. W przypadku cyfr „1" i „7" odpowiada to nachyleniu „daszka", co stanowi bardzo ważną cechę różnicującą także w przypadku innych obserwatorów.

Na obszarze „c", czyli na lewej części obrazu znaku, koncentrowano uwagę podczas 9 eksperymentów. W 4 przypadkach obszary najdłuższej koncentracji uwagi i największej liczby powrotów pokrywały się.

Kolejną cechą jest pozioma kreska na obrazie cyfry „7" (obszar „a"). Rejon ten był przeszukiwany podczas 5 eksperymentów (na 20 przeprowadzonych), przy czym w 3 przypadkach zachodziła sytuacja pokrywania się obszarów określonych obydwoma wskaźnikami.

Tab. 1. Zestawienie położeń kluczowych cech cyfr „1" i „7" dla 10 badanych osób (objaśnienia w tekście i na rys. 5)

lp.	cyfra „1"		cyfra „7"	
	skupienia	„powroty"	skupienia	„powroty"
1.	b	c	c	c
2.	a	a	b	b
3.	c	c	d	c
4.	c	c	b	b
5.	a	b	a	a
6.	a	c	b	c
7.	d	d	c	c
8.	a	a	c	b
9.	b	b	b	b
10.	b	b	b	b



Rys. 5. Położenie podobszarów, dla których zlokalizowano kluczowe cechy (por. tab. 1): okrąg oznacza podobszar „a", elipsa – „b", prostokąt – „c". Podobszar „d" znajduje się poza obrysem znaku

4. Podsumowanie

Na podstawie przeprowadzonych badań składających się z serii eksperymentów z użyciem urządzenia OBER2 oraz z algorytmów interpretacji otrzymanych rejestracji można stwierdzić, że postawiony na wstępie cel został osiągnięty: okazało się, że jest możliwa lokalizacja cech obiektów, na jakie zwraca uwagę człowiek podczas procesu rozpoznawania

na podstawie analizy trajektorii oczu. Zaproponowane wskaźniki: obszar najdłuższego skupienia wzroku oraz obszar o największej liczbie powrotów uwagi dobrze lokalizują istotne cechy rozpoznawanych cyfr.

Odrębnym zagadnieniem jest interpretacja uzyskanych wyników, których przykład zawiera tabela 1. Na ich podstawie nie można wskazać jednej metody rozpoznawania. Obserwuje się zindywidualizowane strategie rozpoznawania, choć w części przypadków obydwa wskaźniki (obszary) pokrywają się lub znajdują się blisko siebie. Analiza położenia cech dla pojedynczych cyfr lub dla innych par (jak pokazano to w tabeli 1) będzie tematem dalszych badań. Kolejnym etapem interpretacji wyników może być także określenie kolejności przenoszenia uwagi, uwzględnienie większej liczby obszarów (nie tylko największych) oraz badanie otrzymanych rejestracji przedziałami odpowiadającymi np. okresom przed i po rozpoznaniu obiektu.

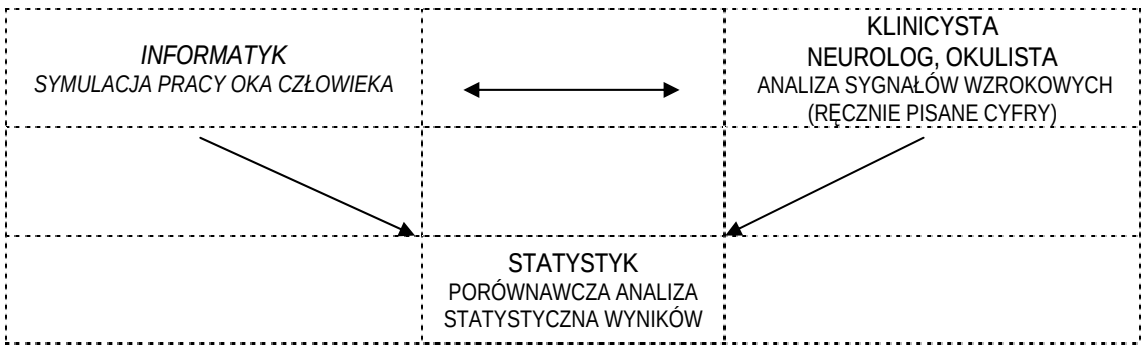
Podziękowania

Autorzy dziękują studentom III roku kierunku Automatyka i Robotyka Bartłomiejowi Kluzie i Adrianowi Korczyńskiemu za pomoc w realizacji eksperymentów oraz w interpretacji wyników.

Opisane prace zostały zrealizowane w ramach badań statutowych (umowa AGH nr 11.11.120.249).

Literatura cytowana

1. Alpaydın E., Salah A. A., Akarun L.: A Selective Attention Based Method for Visual Pattern Recognition with Application to Handwritten Digit Recognition and Face Recognition, IEEE Trans. Pattern Analysis and Machine Intelligence, 24(3): 420-425, 2002.
2. Augustyniak P., Mikrut Z.: Correlating the degree of observer's preoccupation and the observation time: visual tasks with OBER2 eyetracker, J. of Medical Informatics & Technologies, 4: MT3-MT10, 2002.
3. Mikrut Z., Augustyniak P.: Estimation of execution times for tasks of objects counting and localization using the OBER2 device, Proc. 12th Nordic Baltic Conf. on Biomedical Engineerig and Medical Physics, 144-145, Reykjavik 2002.
4. Ober J., Hajda J., Loska J., Jamicki M.: Application of eye movement measuring system OBER2 to medicine and technology, Proc. of SPIE, Infrared technology and applications, , vol. 3061, part 1, pp. 327-33, Orlando, USA 1997.
5. Schwartz E. L.: Spatial mapping in the primate sensory projection: analytic structure and relevance to perception, Biological Cybernetics, 25:181-194, 1977.
6. Tadeusiewicz R., Ogiela M. R.: Medical Image Understanding Technology, Springer, Berlin 2004.
7. Zeki S.: Obrazy wzrokowe w mózgu i umyśle, Świat Nauki, 11:43-51, Warszawa 1992.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 311-316.

WYKORZYSTANIE GRAMATYK GRAFOWYCH DO SYNTAKTYCZNEGO OPISU PRZESTRZENNYCH STRUKTUR NACZYŃ WIEŃCOWYCH

APPLICATION OF GRAPH GRAMMARS FOR SYNTACTIC DESCRIPTION OF CORONARY ARTERIES SPATIAL STRUCTURES

MAREK R. OGIELA, MIROSŁAW TRZUPEK

Katedra Automatyki Akademii Górniczo-Hutniczej, Kraków, mogiela@agh.edu.pl

Streszczenie. W pracy zostanie zaprezentowane nowe podejście do tworzenia semantycznego opisu struktur naczyń wieńcowych serca. Opis taki ma na celu modelowanie przestrzennej topologii naczyń i jest ukierunkowany na możliwości dokonywania znaczeniowej interpretacji morfologii poszczególnych odcinków i odgałęzień tętnic wieńcowych, w celu ich szybkiej identyfikacji oraz automatyzacji procesów detekcji miejsc istotnych przewężeń ich światła. Modelowanie takie zostanie przeprowadzone dzięki użyciu grafowych języków obrazowych opartych na gramatykach ekspansywnych typu edNLC, które dzięki dużej mocy generacyjnej pozwolą na opracowanie sposobów tworzenia uniwersalnego opisu znaczeniowego tych struktur. Takie opisy semantyczne mogą następnie tworzyć integralne części inteligentnych, medycznych systemów informacyjnych lub systemów wspomagania wczesnej diagnostyki stanów niedokrwienia mięśnia sercowego. Zastosowanie grafowych języków obrazowych do analizy przestrzennych struktur unaczynienia serca jest przykładem ukazującym rozwój percepcyjnych metod komputerowego rozumienia zobrazowań diagnostycznych serca. Metody te w ostatnim czasie są niezwykle dynamicznie rozwijane, co ma doniosłe znaczenie dla rozwoju dziedziny biocybernetyki.

Słowa kluczowe: rozpoznawanie obrazów medycznych, sztuczna inteligencja, diagnostyka chorób serca, inżynieria biomedyczna, rozumienie obrazów

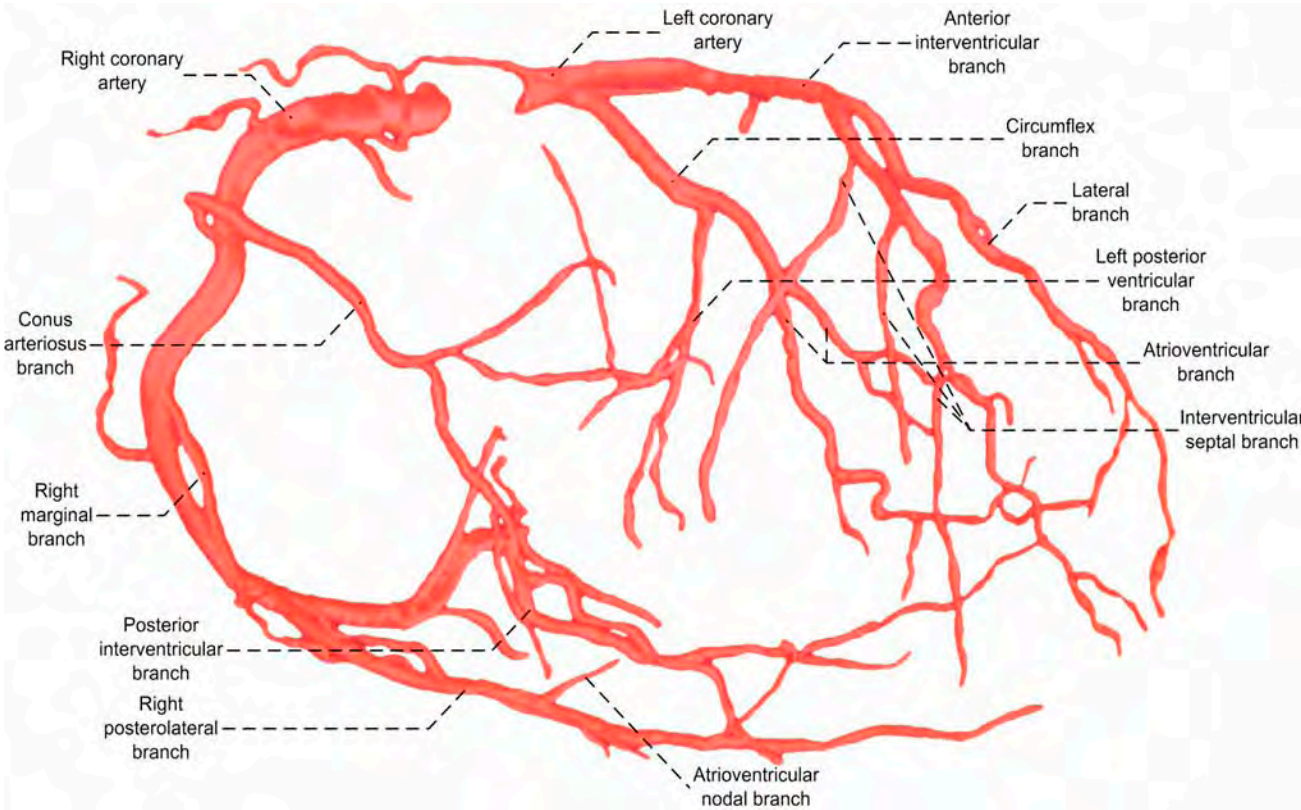
Abstract. The paper presents the new approach to the creation of the semantic description of coronary arteries structure. Such description aims at modeling the spatial topology of vessels, and is directed at possibilities of making semantic interpretation of coronary arteries morphology. It is leading towards their fast identification and automatisation of processes of lumen stricture detection. Such modeling will be possible owing to the ication of graph image languages based on the expansive graph grammars of edNLC type. These types of grammars, thanks to their great generation power, allow to define the ways of creating an universal description of such structures. Such semantic descriptions may be later used as an integral part of intelligent medical information systems or Computer-Aided Diagnosis Systems supporting the early detection of cardiac ischemic states. Application of the graph image languages for the analysis of spatial structures of coronary vessels is an example showing the development of perceptual methods of computer understanding of cardiac medical visualization. Now these methods are more extensivety elaborated, which is most important for evolution of bio-cybernetics.

Key words: medical pattern recognition, artificial intelligence, heart disease diagnosis, biomedical engineering, image understanding

1. Wstęp

Ogromny postęp technologiczny w dziedzinie metod medycznej diagnostyki obrazowej oraz wizualizacji przestrzennej różnorodnych organów wpłynął na zwiększenie skuteczności zadań diagnostycznych praktycznie wszystkich części ludzkiego ciała. Szerokie możliwości przestrzennego modelowania badanych struktur nie tylko umożliwiły wierne odwzorowanie przestrzenne wybranego narządu, ale również dokładną obserwację jego zewnętrznej lub wewnętrznej morfologii, a także zmian jego tekstury. Takie zaawansowane techniki przetwarzania obrazów są obecnie stosowane praktycznie we wszystkich rodzajach badań diagnostycznych opartych na technice cyfrowej oraz zadań operacyjnych z wykorzystaniem robotów medycznych [1], [5], [12]. Nietrudno jednak zauważyć, że wszystkie te osiągnięcia na płaszczyźnie technik wizualizacji mają ograniczone możliwości w zakresie automatyzacji zadań interpretacyjnych prowadzonych dla analizowanych obrazów. Wynika to głównie z trudności, na jakie napotyka biocybernetyka i inżynieria biomedyczna w dziedzinie modelowania skomplikowanych procesów myślowych przebiegających w umyśle człowieka, a prowadzących do znaczeniowej interpretacji uzyskiwanych zobrazowań medycznych. Jak pokazują badania, znaczna część takich zobrazowań jest wciąż trudna do automatycznej analizy prowadzonej w oparciu o komputerową interpretację informacji o znaczeniu semantycznym.

Okazuje się jednak, że ta istotna luka badawcza ma szansę zostać zapełniona poprzez rozwinięcie przedstawionego w niniejszej pracy lingwistycznego podejścia do modelowania struktur przestrzennych z wykorzystaniem grafowych języków opisu cech kształtów [5], [6, [7], [12]. Języki takie pozwalają na dokonanie automatycznej identyfikacji istotnych punktów informatywnych w obrazie danego narządu (np. rekonstrukcji naczyń wieńcowych), a następnie poprzez odpowiednio wprowadzone relacje przestrzenne na lokalizację oraz określenie stopnia zaawansowania zmian o charakterze patologicznym. Warto w tym momencie zwrócić uwagę na fakt, iż metody lingwistyczne sztucznej inteligencji znacząco wpływają na podniesienie możliwości cybernetycznego modelowania procesów analizy znaczeniowej i rozumienia obrazów poprzez wykorzystanie informacji o charakterze semantycznym. Języki obrazowe charakteryzują się nie tylko możliwościami klasyfikacyjnymi, ale również możliwościami gromadzenia i syntezy informacji o charakterze semantycznym i to zarówno niskiego, jak i wysokiego poziomu. Właśnie na takie zadania będą ukierunkowane metody prezentowane w dalszej części pracy i zaproponowane w celu semantycznego modelowania naczyń wieńcowych serca (rys. 1). Jest to istotny postęp naukowy, gdyż jak do tej pory najbardziej rozwiniętym i nadal rozwijanym obszarem zastosowań przetwarzania i analizy obrazów w kardiologii jest identyfikowanie i lokalizowanie (często manualnie) przewężeń oraz zatorów w tętnicach wieńcowych [3], [4], [6]. Wiele rozbudowanych algorytmów wstępnego przetwarzania i analizy obrazów stanowi kluczowy składnik badania angiograficznego serca, ale także jego diagnostyki za pomocą rentgenowskiej tomografii komputerowej, badań rezonansu magnetycznego, aż po diagnostykę USG struktury wewnątrznaczyniowej (IVUS) [4], [10].



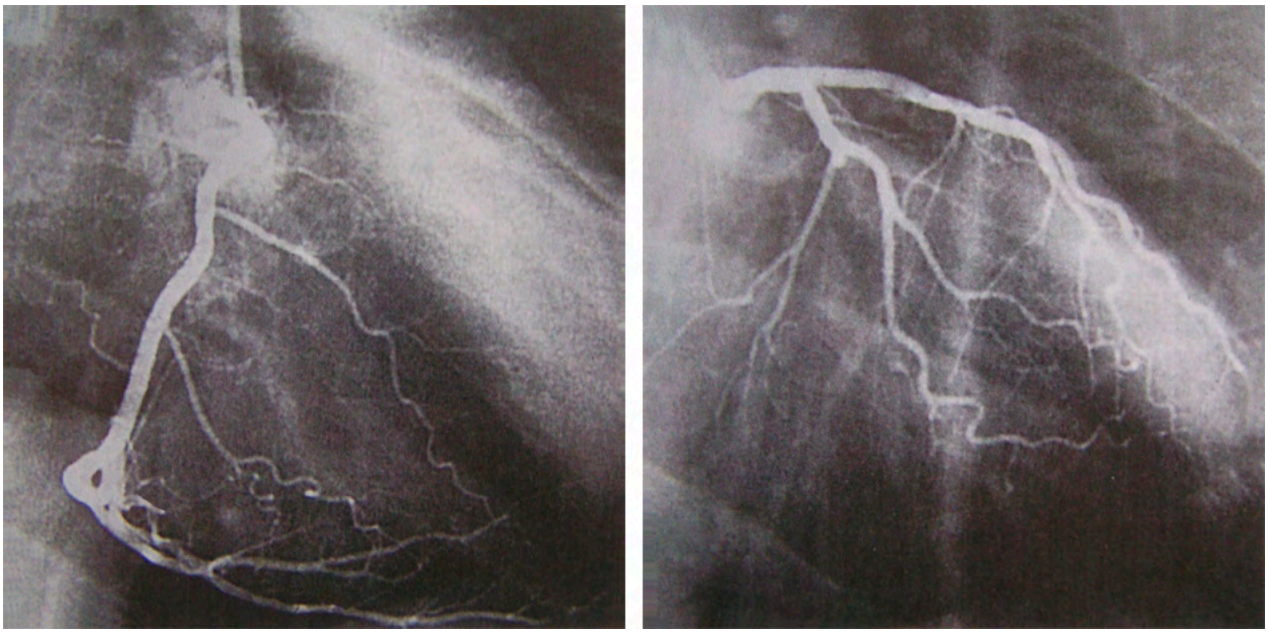
Rys. 1. Przestrzenna rekonstrukcja naczyń wieńcowych

Wszystkie wymienione badania wzajemnie się uzupełniają, a każde z nich umożliwia na różnym poziomie szczegółowości odtworzenie informacji na temat trójwymiarowej struktury naczyń wieńcowych. Ich projekcja 2D została przedstawiona na rys. 1. Natomiast w zakresie komputerowego odtwarzania struktury trójwymiarowej znajdują zastosowanie metody geometryczne i algorytmy służące do wyznaczania wzajemnych odpowiedników w wykorzystywanej grupie „płaskich" obrazów (tzw. matching) oraz techniki tzw. renderingu, czyli wizualizacji struktury trójwymiarowej w postaci płaskiego obrazu wyświetlanego na ekranie komputera [1], [3], [9]. Wszystkie te techniki są wykorzystywane do otrzymywania materiału diagnostycznego mogącego zawierać istotne informacje dotyczące występowania lub stanu zaawansowania pewnych schorzeń. Nie dają one jednak możliwości ich automatycznej interpretacji, tak jak czyni to specjalista. Dlatego też w dalszej części opisany zostanie sposób tworzenia lingwistycznej reprezentacji znaczeniowej dla struktur naczyń wieńcowych wykazujących pewne zmiany chorobowe. Opis ten będzie umożliwiał również określenie semantyki wykrytych zmian oraz wspomaganie zadań interpretacji znaczeniowej występującej patologii.

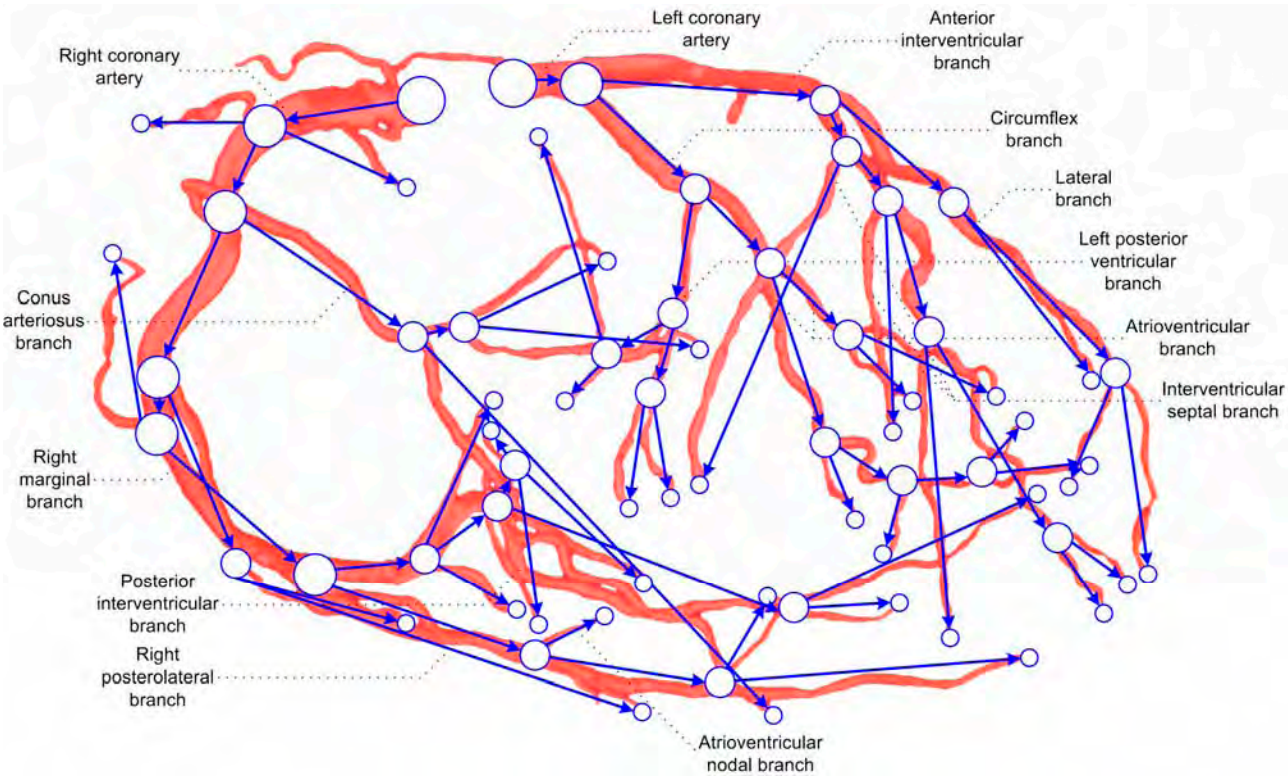
2. Materiał i metody

Aby umożliwić prawidłowe zdefiniowanie grafowej gramatyki obrazowej, która umożliwi określenie lokalizacji oraz oszacowanie wielkości zmiany chorobowej (np. w postaci przewężenia światła jednej z tętnic), zostały wykorzystane sekwencje DICOM zobrazowań tętnic wieńcowych otrzymywanych w trakcie wykonywania badań koronarograficznych. Przykłady takie zostały zilustrowane na rys.2.

Obrazy takie dokładnie ukazują zmiany morfologii tętnic oraz umożliwiają w sposób optymalny przestrzenny rendering naczyń wieńcowych.



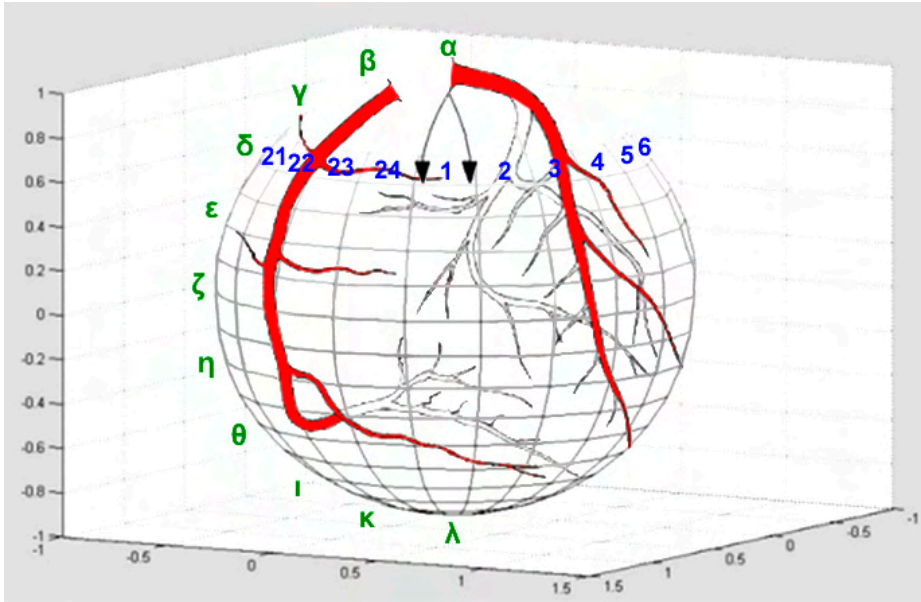
Rys. 2. Obrazy koronarograficzne tętnic wieńcowych



Rys. 3. Schemat przestrzennego etykietowania tętnic oraz relacji pomiędzy nimi

Rys. 3 ukazuje trójwymiarową wizualizację, wykorzystującą koronarogramy po wstępnym przetworzeniu i eksponującą naczynia na obrazach binarnych. Naczynia tam uwidocznione zostały opisane za pomocą struktur grafowych

w ten sposób, że poszczególne węzły grafu identyfikują wszystkie punkty początkowe oraz końcowe naczyń, a także wszystkie punkty odgałęzień lub przejścia naczyń głównych w naczynia niższego stopnia. Struktura grafowa utworzona w ten sposób będzie tworzyła elementy języka definiującego prawidłowe naczynia i ich topologię oraz unaczynienie serca wraz ze zmianami morfologicznymi.



Rys. 4. Zbiór etykiet krawędziowych grafów opisujących wzajemne relacje przestrzenne poszczególnych tętnic wieńcowych.

Na rys. 4 przedstawiono wzajemne relacje, jakie mogą zachodzić pomiędzy elementami struktury naczyniowej reprezentowanej w postaci grafu. Uwzględnienie tych relacji oraz etykietowanie wierzchołków grafu rozpinającego arterie pozwala na utworzenie lingwistycznego opisu semantyki poszukiwanych zmian i rozpoznawania (lokalizacji) patologicznych przewężeń (rys. 3). Dzięki temu możliwe jest np. przeprowadzanie komputerowej analizy otrzymanej struktury ukierunkowanej na automatyczną detekcję miejsca przewężenia, jego rodzaju (koncentryczne lub ekscentryczne), a także jego stopnia.

Przy takiej reprezentacji otrzymuje się zwięzły oraz jednoznaczny opis wszystkich elementów struktury naczyniowej, co pozwala na ścisłe określenie kierunku analizy w poszukiwaniu miejsc patologii np. w postaci istotnych przewężeń naczyń wieńcowych. Przeszukiwanie jest wykonywane w trakcie analizy składniowej badanych reprezentacji grafowych (otrzymywanych dla kolejnych przypadków diagnostycznych) utworzonych z wykorzystaniem gramatyk typu edNLC.

Poniżej przytoczono formalną definicję tego typu gramatyki. Warto jednak zaznaczyć, że klasa tych gramatyk posiada bardzo efektywne analizatory składniowe działające w czasie wielomianowym.

Def. Krawędziowo-etykietowaną, skierowaną, sterowaną etykietą wierzchołka gramatyką grafową (gramatyką G_{edNLC} – ang. *edge-labelled directed Node-Label Controlled graph grammar*) nazywamy zbiór pięciu elementów definiowanych w następujący sposób [12], [14]:

$$G_{edNLC} = (\Sigma, \Delta, \Gamma, P, Z),$$

gdzie:

Σ – jest skończonym, niepustym zbiorem etykiet wierzchołkowych,

Δ (podzbiór Σ) – jest zbiorem terminalnych (końcowych) etykiet wierzchołkowych,

Γ – jest skończonym, niepustym zbiorem etykiet krawędziowych,

P – jest skończonym zbiorem produkcji o postaci (l, D, C), gdzie $l \in \Sigma$, D jest grafem EDG,
$C : \Gamma \times \{in, out\} \rightarrow 2^{\Sigma \times \Sigma \times \Gamma \times \{in, out\}}$ jest transformacją osadzenia,

Z – jest startowym grafem, od którego rozpoczyna się proces generacji wszystkich postaci grafowych dla tętnic wieńcowych.

W celu określenia miejsca występowania patologii naczyń, w grafie tętnic wieńcowych (z rys. 3) zaproponowano gramatykę następującej postaci:

$$G_{edNLC} = (\Sigma, \Delta, \Gamma, P, Z),$$

gdzie $\Sigma = \Sigma_N \cup \Sigma_T$ jest zbiorem zarówno terminalnych, jak i nieterminalnych etykiet wierzchołkowych opisujących badane grafy i zdefiniowanych w następujący sposób:

Zbiór terminalnych etykiet wierzchołkowych (rys. 3).

Σ_T = {coronary artery, right coronary artery,
conus arteriosus branch, right marginal branch,
posterior interventricular branch,
right posterolateral branch,
atrioventricular nodal branch,
left coronary artery,
anterior interventricular branch,

circumflex branch, lateral branch,
left posterior ventricular branch,
atrioventricular branch,
interventricular septal branch}

Zbiór nieterminalnych etykiet wierzchołkowych.

Σ_N = {Coronary_Artery,
C_Right, conus_arteriosus_br,
C_Right_marg_br, C_Right_post_intervent_br,
C_Right_posterolat_br,
C_Right_atriovent_nod_br, C_Left,
C_Left_ant_intervent_br, C_Left_circumflex,
C_Left_later_br, C_Left_post_vent_br,
C_Left_atriovent_br,
C_Left_intervent_sept_br}

$\Delta = \Sigma_T$ zbiór terminalnych etykiet wierzchołkowych,
Γ = {α, β, γ, δ} jest zbiorem etykiet opisujących krawędzie grafu,
Z = {Coronary_Artery} jest pierwotnym grafem startowym.

3. Wyniki

Dzięki zastosowaniu przedstawionej w pracy gramatyki grafowej możliwe jest przestrzenne modelowanie morfologii oraz dokładne wykrywanie miejsc tzw. istotnych przewężeń tętnic wieńcowych. W przypadku analizy strukturalnej takich reprezentacji, składniowy analizator rozpoznający (*parser*) w sposób automatyczny dostarcza pełnej informacji określającej przestrzenną topologię badanego grafu opisującego naczynia wieńcowe oraz jego poszczególnych składników. Informacja taka jest następnie wykorzystywana do analizy strukturalnej i semantycznej poszczególnych odcinków tętnic (wyznaczanych przez terminalne etykiety wierzchołkowe), co pozwala na dokładne określenie morfologicznych nieprawidłowości badanych naczyń [8], [13].

Metody syntaktyczne wykazują tutaj ogromną przydatność do określania semantycznych informacji dotyczących zmian tętnic serca oraz mogą stanowić dodatkowe narzędzie służące do wspomagania wczesnej diagnostyki choroby niedokrwiennej.

Możliwości semantycznej interpretacji przestrzennych rekonstrukcji naczyń wieńcowych z użyciem opisanych gramatyk poszerzają spektrum zastosowań metod sztucznej inteligencji w kierunku określania medycznego znaczenia badanych patologii, a także w kierunku metod komputerowego rozumienia obrazów medycznych.

4. Podsumowanie

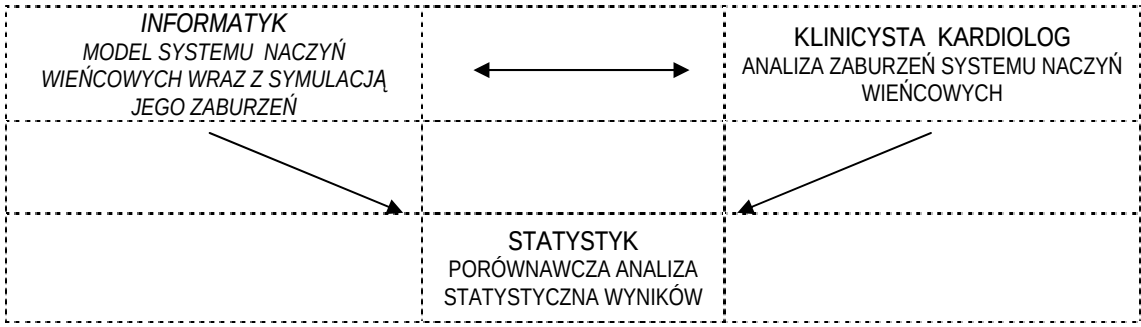
Przedstawione w pracy podejście wykorzystania grafowych języków obrazowych w przestrzennym modelowaniu struktur wieńcowych oraz określania semantyki występujących tam zmian stanowi nową jakość w zadaniach komputerowej diagnostyki medycznej, stwarzając ogromne możliwości jej ukierunkowania na dokonanie automatycznej analizy i detekcji zmian w sposób percepcyjny. Możliwości takie istnieją dzięki opracowaniu na potrzeby analiz medycznych nowego typu kognitywnych systemów informacyjnych, bazujących na zaawansowanych formalizmach sztucznej inteligencji, soft-computingu oraz grafowych językach obrazowych.

Techniki takie mają doniosłe znaczenie, gdyż pozwalają nie tylko na wykrywanie zmian, ale w pewnym zakresie pozwalają również na określenie ich semantyki, co w przypadku zobrazowań medycznych (zwłaszcza trójwymiarowych) może prowadzić do komputerowego „rozumienia" ich medycznego znaczenia, a także precyzowania diagnozy i określania optymalnych możliwości terapeutycznych. Bardzo istotna z punktu widzenia mechanizmów biocybernetyki jest głęboka analogia funkcjonowania modelu analizy strukturalnej do interpretacyjnych mechanizmów kognitywnych zachodzących w umyśle człowieka. Analogia ta polega na wykorzystaniu interferencji pomiędzy oczekiwaniami (wiedzą gromadzoną w postaci grafowych reguł gramatycznych), a strumieniem danych płynącym z systemu analizującego badane zobrazowanie. Interferencja taka jest charakterystyczna dla modelu percepcji wzrokowej człowieka.

Literatura cytowana

1. Bankman I. (ed.): Handbook of Medical Imaging: Processing and Analysis, Academic Press, 2002.
2. Davis L. S. (eds.): Foundations of image understanding, Kluwer Academic Publishers, Norwell 2001.
3. Higgins W. E., Karwoski R. A., Ritman E. L., Spyra W. J. T.: System for analyzing true three-dimensional angiograms, IEEE Trans. Med. Imag., 15:377-385, 1996.
4. Khan M. G.: Heart Disease Diagnosis and Therapy, Williams & Wilkins, Baltimore 1996.
5. Meyer-Baese A.: Pattern Recognition in Medical Imaging, Elsevier- Academic Press, 2003.
6. Ogiela M. R., Tadeusiewicz R.: Syntactic reasoning and pattern recognition for analysis of coronary artery images, Artificial Intelligence in Medicine, 26: 145-159.
7. Ogiela M. R., Telejko K.: Grafowe języki obrazowe w percepcyjnej analizie zjawisk chorobowych i wspomaganiu diagnostyki medycznej, Przegląd Lekarski, 61: 121-123.
8. Ogiela M. R., Tadeusiewicz R.: Picture Languages in Intelligent Retrieval of Visual Data Semantic Information, LNAI 3336: 389-396.
9. Ogiela M. R.: Strukturalne metody rozpoznawania obrazów w kognitywnej analizie zobrazowań medycznych, Wydawnictwa AGH, Kraków 2004.
10. Higgins W. E., Reinhardt J. M.: Cardiac image processing, 789-804, in Handbook of Video and Image Processing. ed. A. Bovik, Academic Press, 2000.
11. Robb R. A.: Biomedical Imaging Visualization, and Analysis, Wiley-Liss, 2000.
12. Tadeusiewicz R., Ogiela M. R.: Medical Image Understanding Technology, Springer Verlag, Berlin-Heidelberg 2004.
13. Tadeusiewicz R., Ogiela M. R.: Ogiela, Structural Approach to Medical Image Understanding, Bulletin of the Polish Academy of Sciences – Technical Sciences, 52(2): 131-139.
14. Tadeusiewicz R., Flasiński M.: Rozpoznawanie obrazów, Wydawnictwo Naukowe PWN, Warszawa 1991.

Praca realizowana w ramach Badań Własnych AGH Nr 10.10.120.39.


INFORMATYK
MODEL SYSTEMU NACZYŃ WIEŃCOWYCH WRAZ Z SYMULACJĄ JEGO ZABURZEŃ
KLINICYSTA KARDIOLOG
ANALIZA ZABURZEŃ SYSTEMU NACZYŃ WIEŃCOWYCH
STATYSTYK
PORÓWNAWCZA ANALIZA STATYSTYCZNA WYNIKÓW

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 317-320.

KOMPUTEROWA ANALIZA OBRAZU W OCENIE ZMIAN ZACHODZĄCYCH W STREFACH GRANICZNYCH: KOŚĆ-CEMENT-IMPLANT

ASSESSMENT OF THE CHANGES OCCURRING IN THE BONE-CEMENT-IMPLANT BORDER ZONES USING THE COMPUTER IMAGE ANALYSIS

JERZY WŁODARSKI

Zakład Inżynierii Biomedycznej, Politechnika Częstochowska, Częstochowa, wlodarski@iop.pcz.czest.pl

Streszczenie. W pracy analizowano zmiany w obrazie radiologicznym zachodzące w obszarze styku: kość-cement-implant pod wpływem obciążeń zmiennych realizowanych w badaniach preparatów na symulatorze stawu biodrowego. Analiza widma amplitudowego posłużyła do obiektywnego wnioskowania o charakterze zmian zachodzących w strefach granicznych: kość-cement-implant pod wpływem obciążeń i rokowania co do trwałości połączenia.

Słowa kluczowe: radiogram, widmo amplitudowe, endoproteza, implant

Abstract. In the paper, changes in the radiological image at the bone-cement-implant contact region under changing loads have been analyzed. The tests of specimens have been performed on a hip-joint simulator. The analysis of the amplitude spectrum has served to come to the objective conclusions of the nature of changes occurring in border zones: bone-cement-implant under the influence of loads. Furthermore, the results obtained from the analysis allow to predict the durability of the joint.

Key words: radiogram, amplitude spectrum, endoprosthesis, graft

1. Materiał i metoda

Metoda komputerowej analizy obrazu jest coraz chętniej używana w medycynie do obrazów powstających podczas badań diagnostycznych. Przykładami mogą być preparaty histopatologiczne, obrazy USG czy też – jak w przypadku tej pracy – obrazy rentgenowskie [5].

Przekształcenia komputerowej analizy obrazu można wykorzystać do jakościowej analizy zdjęć rentgenowskich. Odejmując od siebie obrazy kości udowej z endoprotezą w chwili jej osadzania i po określonej liczbie cykli obciążenia, znacznie łatwiej zauważyć ewentualne zmiany w obrazach styku „kość-cement" i „cement-implant". Obrazy składowe do tej analizy muszą być otrzymane w jednakowych warunkach (parametry naświetlania) i jednakowym położeniu.

A zatem można w przyszłości badać zmiany w czasie, zachodzące po wszczepieniu endoprotezy jednego i tego samego pacjenta.

Badaniom poddano preparaty anatomiczne z osadzoną cementowo endoprotezą, które podlegały różnym sposobom obciążenia. Badania modelowe prowadzono na symulatorze stawu biodrowego umożliwiającego odwzorowanie dynamiki obciążeń w stawie biodrowym człowieka.

Obciążenie na głowę endoprotezy w postaci siły skupionej było przenoszone poprzez panewkę polietylenową i odpowiadało normalnej pracy stawu biodrowego (tzn. podczas chodu). W ustawieniu kości na stanowisku badawczym brano pod uwagę jej fizjologiczne położenie z uwzględnieniem wypadkowej siły obciążającej staw biodrowy. Wielkość maksymalnej wypadkowej siły obciążającej odpowiadała 3 lub 4-krotności masy ciała człowieka [1], [4].

Po każdorazowym obciążeniu połączenia (1-1,5) x 10^6 razy (odpowiada to liczbie cykli obciążenia jednego stawu w ciągu roku) wykonywano radiogram kości z osadzoną endoprotezą. Radiogramy wykonywano także bezpośrednio po osadzeniu endoprotezy.

Preparaty kości umieszczano w specjalnym przyrządzie umożliwiającym jednakowe dokładne zachowanie pozycji względem lampy rentgenowskiej i kasety.

Wykonane radiogramy mogły być wówczas porównywalne, gdyż dotyczyły tego samego ustawienia preparatu.

Zdjęcia preparatów uzyskiwano, naświetlając je przy jednakowych parametrach, wykonując dwa radiogramy: bez powiększenia i przy x2 powiększeniu. Powiększony radiogram służył do komputerowej analizy uzyskanego obrazu rtg.

Radiogramy kości udowych służyły do rejestracji zmian zachodzących na powierzchniach granicznych „kość-cement"

i „cement-implant" badanych preparatów pod wpływem obciążeń.

Otrzymane negatywy radiologiczne przetwarzano w postać cyfrową w pliki w formacie TIFF z użyciem skanera VitaScan firmy UMAX z rozdzielczością 300 dpi. Pliki te następnie wprowadzano do cyfrowego analizatora obrazu IBAS firmy KONTRON.

Analizę cyfrową tych plików przeprowadzono w dwóch etapach, z wykorzystaniem dwóch sposobów. Opracowano w tym celu dwa specjalne programy komputerowe umożliwiające realizację tych celów.

Narzędziami programistycznymi były procedury oferowane przez oprogramowanie IBAS wersja 2.5 (dla celów analizy obrazu) oraz język C++ (dla celów analizy matematycznej danych cyfrowych uzyskanych w pierwszym etapie).

I sposób:

1. Segmentacja progowa profilu implantu.
2. Utworzenie maski poprzez dilatację profilu implantu o 16 pixeli (oktagonalny operator strukturalny).
3. Wydzielenie obszaru implantu i jego otoczenia metodą maskowania.
4. Odjęcie profilu implantu (pkt. 1) od wydzielonego obszaru w pkt. 3.
5. Zapis stopni szarości wszystkich pikseli obszaru wynikowego obejmującego szesnastopikselowy pas otaczający implant w postaci tabeli zawierającej następujące dane: współrzędne x-y pixela, stopień szarości w skali 0-256.
6. Przetworzenie cyfrowych danych densytometrycznych (pkt. 5) w 16 wektorów opisujących ciągi szarości w kolejnych 16 „warstwach" otaczające implant.
7. Analiza Fourierowska wektorów uzyskanych w pkt. 6.
8. Porównanie wyników (widm) z obrazów dotyczących tych samych implantów otrzymanych po poszczególnych cyklach badanego preparatu anatomicznego.

II sposób:

1. Segmentacja progowa profilu implantu.
2. Utworzenie maski poprzez dilatację profilu implantu o 16 pixeli (oktagonalny operator strukturalny).
3. Wydzielenie obszaru implantu i jego otoczenia metodą maskowania.
4. Odjęcie profilu implantu (pkt.1) od wydzielonego obszaru w pkt. 3.
5. Porównanie obrazów wynikowych obejmujących szesnastopikselowe pasy otaczających implanty w obrazach dotyczących tych samych implantów otrzymanych po poszczególnych cyklach obciążenia danego preparatu. Porównania tego dokonywano przez nałożenie i odjęcie ich od siebie.

Celem obu tych sposobów było wykrycie ewentualnych zmian (mikropęknięć) w obszarach styku „kość-cement" i „cement-implant" [2], [5].

Tezą, jaka przyświecała hipotezie o skuteczności pierwszego z podanych sposobów, było założenie, że ewentualne mikropęknięcia pojawiające się po obciążeniu modelu mają jakąś charakterystyczną i powtarzalną lokalizację i przebieg. Ich wykrywanie, jeśli okazałoby się skuteczne, byłoby mało wrażliwe na zmienność pozycji przestrzennej preparatu względem aparatu rentgenowskiego.

Drugi ze sposobów wymaga dokładnej fiksacji przestrzennej preparatów naświetlanych w poszczególnych etapach eksperymentu i jedynie identyfikuje fakt ewentualnych mikropęknięć substancji cementującej wszczep.

2. Transformata Fouriera dla obrazu powierzchni przylegania endoprotezy

Podstawę do obliczeń stanowiły pliki tekstowe zawierające dane o wartościach jasności obrazu w okolicy brzegu protezy stawu biodrowego. Dla każdego zdjęcia wyznaczono kontur obrazu protezy o grubości od 16 do 64 pikseli w zależności od zdjęcia.

Z otrzymanych danych dla każdego zdjęcia wygenerowano dwie prostokątne macierze o wymiarach M = 64 kolumny na tyle wierszy, ile zajmowała długość protezy na zdjęciu (przeważnie N = 480 – 500 wierszy). Macierze takie utworzono dla lewej i prawej strony protezy, po jednej dla każdej ze stron. W przypadku, gdy liczba pikseli konturu dla danej linii obrazu była mniejsza niż 64, wiersz macierzy był uzupełniany zerami do tej wartości. Wygenerowane macierze zostały zapisane w plikach tekstowych. Do wykonania obliczeń i ich zapisu do plików wykorzystano program Turbo Pascal firmy Borland wersja 7.0.

Dla otrzymanych macierzy wykonano dwuwymiarową dyskretną transformatę Fouriera opisaną wzorem:

$$F(p,q)=\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} f(m,n)\cdot e^{-j(2\pi/M)pm}\cdot e^{-j(2\pi/N)qn} \qquad (1)$$

$$p=0,1,.....,M-1; \quad q=0,1,...,N-1$$

gdzie:

f(m,n) – element macierzy o współrzędnych *(m,n)*,
F(p,q) – współczynnik Fouriera dla elementu *f(m,n)*,
p,q – częstotliwości przestrzenne,
M,N – wymiary przetwarzanej macierzy.

Nastepnie wyznaczono moduł zespolonej funkcji *F(p,q)* (w celu określenia widma amplitudowego) oraz logarytm tego modułu, aby uwypuklić składowe widma o niewielkiej amplitudzie i wytłumić składowe o dużych wartościach amplitudy. Kolejną operacją było takie przesunięcie uzyskanego widma, aby jego składowa stała (element *F(0,0)*) znajdowała się w środku obrazu. Powyższe operacje (logarytmowanie i przesunięcie widma) ułatwiają analizę uzyskanego widma amplitudowego.

Ciąg powyższych operacji opisany jest wzorem:

$$F(p,q)=shift\left(\log\left|F(p,q)\right|\right) \qquad (2)$$

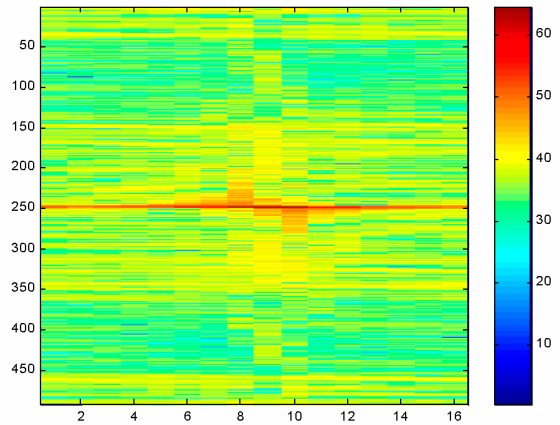
Obraz zawiera cały zakres częstotliwości (w zakresie od π do π dla osi X i Y, element *F(0,0)* reprezentujący zerową częstotliwość znajduje się w środku obrazu), przy czym skala nie jest jednakowa: w kierunku X obraz zawiera 64 składowe widma, zaś w kierunku Y od 480 do 500. Wszystkich obliczeń i wykresów obrazów widma dokonano za pomocą programu Matlab firmy MathWorks wersja 5.0 z wykorzystaniem Image Processing Toolbox (zbiór procedur do przetwarzania obrazów). Analiza widma amplitudowego.

Zapis stopni szarości obrazu wynikowego zawierającego powierzchnię przylegania obejmującą implant i przetworzenie cyfrowych danych densytometrycznych w wektory opisujące ciągi stopni szarości w kolejnych „warstwach" przylegających do implantu umożliwiły przeprowadzenie analizy Fouriera uzyskanych w ten sposób wektorów.

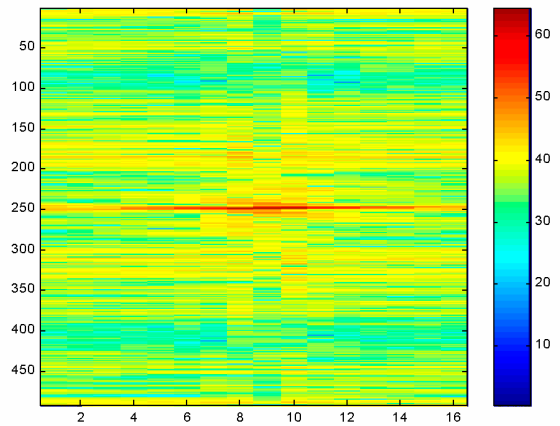
3. Wyniki badań

Na rysunkach 1-4 zamieszczono przykładowe rozkłady widma amplitudowego dla obrazów brzegów endoprotezy stawu biodrowego po obciążeniu 1 x 10^6 cykli i 8 x 10^6 cykli obciążenia.

Rysunki 1 i 3 dotyczą lewej strony endoprotezy, natomiast rysunki 2 i 4 prawej strony endoprotezy.

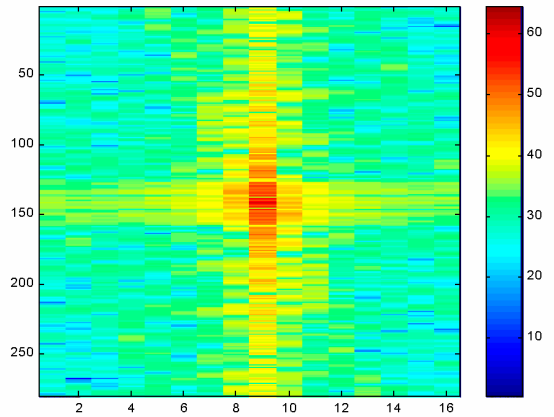


Rys. 1. Rozkład widma amplitudowego dla obrazu brzegu endoprotezy (cement-implant) po stronie lewej. Liczba cykli obciążenia 1 x 10^6

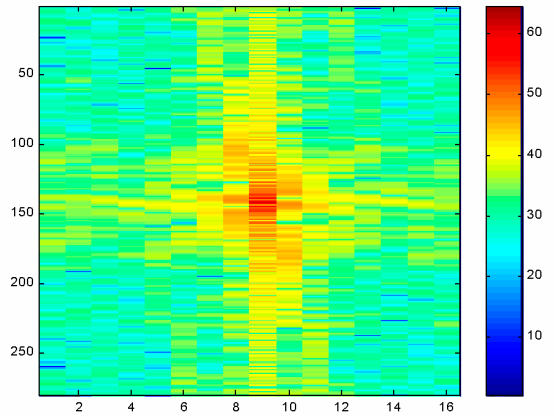


Rys. 2. Rozkład widma amplitudowego dla obrazu brzegu endoprotezy (cement-implant) po stronie prawej. Liczba cykli obciążenia 1 x 10^6

Wydruki modułów transformat Fouriera dla obrazów brzegów endoprotezy stawu biodrowego umożliwiają ocenę zawartości częstotliwości w tych obrazach.



Rys. 3. Rozkład widma amplitudowego dla obrazu brzegu endoprotezy (cement-implant) po stronie lewej. Liczba cykli obciążenia 8 x 10^6



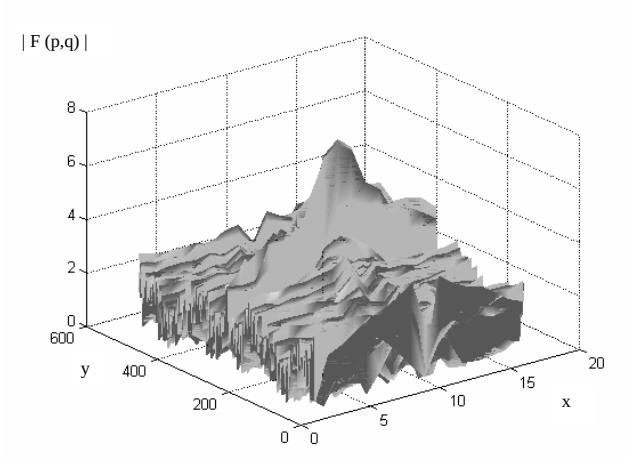
Rys. 4. Rozkład widma amplitudowego dla obrazu brzegu endoprotezy (cement-implant) po stronie prawej. Liczba cykli obciążenia 8 x 10^6

Z porównania rysunków wynika, że występują zmiany widma częstotliwości w czasie dotyczące tej samej strony endoprotezy. Charakter tych zmian jest podobny na obu stronach endoprotezy.

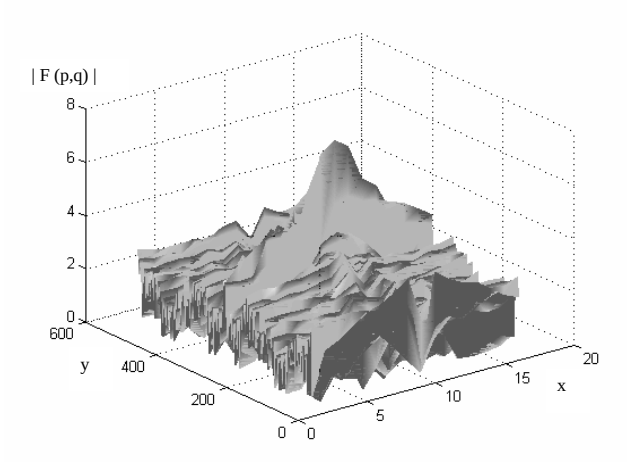
Wysoka częstotliwość w widmie amplitudowym odpowiada gwałtownym zmianom przejścia w poziomach szarości na zdjęciu rentgenowskim. Ta gwałtowna zmiana w poziomach szarości dotyczy strefy granicznej „implant-cement".

Widmo o szybko zanikających wysokich składowych częstotliwościach świadczy o obrazie rentgenowskim zawierającym łagodne przejścia, bez gwałtownych zmian poziomów jasności. W tym przypadku występuje obraz „rozmyty", sugerujący występowanie mikropęknięć w cemencie kostnym, które prowadzą do obluzowania endoprotezy. Wysoki „pik" na rysunku 5 odpowiada składowej stałej (wartości średniej) i odpowiada środkowi widma na rysunku 1.

Na rysunkach 5 i 6 zamieszczono wykresy rozkładu wartości widma amplitudowego (opracowane dla rozkładu widma amplitudowego z rysunku 1 i 2). Wykresy te pozwalają na określenie wartości widma częstotliwości w układzie przestrzennym odniesione do współrzędnych x-y pikseli.



Rys. 5. Wykres wartości rozkładu widma amplitudowego dla obrazu brzegu endoprotezy (cement-implant) po stronie lewej. Liczba cykli obciążenia 1 x 10^6. x, y – współrzędne pikseli



Rys. 6. Wykres wartości rozkładu widma amplitudowego dla obrazu brzegu endoprotezy (cement-implant) po stronie prawej. Liczba cykli obciążenia 1 x 10^6. x, y – współrzędne pikseli

Analiza wydruków modułów transformat Fouriera dla obrazów brzegów endoprotezy pozwoliła ustalić, że mikropęknięcia płaszcza cementowego przylegającego do endoprotezy miały charakterystyczny przebieg oraz powtarzalną i charakterystyczną lokalizację na rozkładach widma amplitudowego. Ich występowanie potwierdzały radiogramy, używane następnie do przetworzenia na postać cyfrową wykorzystywaną następnie do opracowania tych modułów.

Analiza widma amplitudowego może więc służyć do wnioskowania o charakterze zmian zachodzących w strefach granicznych „kość-cement-implant" pod wpływem obciążeń i rokowania co do trwałości połączenia.

Dokonując przekształceń (II sposób) i odejmując od siebie obrazy składowe otrzymane w tych samych warunkach i mających identyczne rozmiary, otrzymano obraz, który, zamieniony na postać binarną, umożliwia analizę zmian zachodzących w strefie przylegania pod wpływem czasu tego samego preparatu.

4. Wnioski

1. Analiza widma amplitudowego pozwala wnioskować o strukturze obrazu rentgenowskiego. Widmo o szybko zanikających wysokich składowych częstotliwości świadczy o obrazie radiologicznym zawierającym łagodne przejścia bez gwałtownych zmian poziomów jasności.
2. Korzystając z komputerowej analizy obrazu, można badać zmiany zachodzące w czasie po wszczepieniu endoprotezy. Metoda ta pozwala na dokonanie jakościowej analizy, a binarna postać obrazu może być przydatna do pomiaru strefy przejaśnień z radiogramu.

Literatura cytowana

1. Będziński R., Bernakiewicz M., Rybaczuk M.: Analysis of material properties of bones. Biology of Sport, s. 192-197, Spała 1997.
2. Callaghan I. I., Fulghum C. S., Glisson R. R., Stranne S. K.: The effect of femoral stem geometry on interface motion in uncemented porous-coated hip prstheses, Comparison of straight-stem and curved-stem designs. I. Bone Joint Surg., 74-A, 1992.
3. Gruen T. A.: Radiographic criteria for the clinical performance of uncemented total hip replacement, Quantatitative characterization and performance of porous implants for hard tissue aplications, (red.): Lemous, I. E.: 1985.
4. Gruen T. A., McNeice G. M., Amstutz H. C.: Modes of failture of cemented stem-type femoral components: a radiographic analysis of loosening, Clin. Orthop., 141, 1979.
5. Tadeusiewicz R., Korohoda P.: Komputerowa analiza i przetwarzanie obrazów. Wyd. Fundacji Postępu Technicznego, Kraków 1997.

INFORMATYK
FIZYK
ANALIZA OBRAZU RADIOLOGICZNEGO NA STYKU KOŚĆ-CEMENT-IMPLANT
KLINICYSTA ORTOPEDA
DANE O FUNKCJONOWANIU NARZĄDU RUCHU

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 321-324.

NEURONAL MOVEMENT DETECTOR IN THE MODEL OF SIMULATED MAMMALIAN VISUAL SYSTEM

NEURONALNY DETEKTOR RUCHU W MODELOWANYM UKŁADZIE WZROKOWYM SSAKA

GRZEGORZ M. WÓJCIK, WIESŁAW A. KAMIŃSKI

Zakład Fizyki Systemów Złożonych, Uniwersytet Marii-Curie Skłodowskiej,
Plac Marii Curie-Skłodowskiej 5, 20-031 Lublin, gmwojcik@hektor.umcs.lublin.pl

Abstract. A simple model of mammalian visual system was created. This structure was simulated by means of several hundred Hodgkin-Huxley neurons. We investigated signal processing properties of the model. Energetic function – a concept taken from information theory was applied to the analysis of Primary Visual Cortex dynamics. Discussion of such method's efficiency in movement detection is presented in some detail.

Key words: Liquid State Machine, Hodgkin-Huxley neuron, movement detection

Streszczenie. Stworzono prosty model układu wzrokowego ssaka. Struktura taka była symulowana przy pomocy kilkuset neuronów Hodgkina-Huxleya. Badano zdolności przewodzenia sygnałów przez taką strukturę. Funkcja energetyczna – metoda zaczerpnięta z teorii informacji – została wykorzystana do analizy dynamiki pierwotnej kory wzrokowej (PVC) mózgu. Zostanie przeprowadzona dyskusja na temat wykorzystania funkcji energetycznej do detekcji ruchu.

Słowa kluczowe: maszyna LSM, neuron Hodgkina-Huxleya, detekcja ruchu

1. Introduction

In Maass's Liquid State Machine (LSM) theory [7] the fragment of brain neocortex is treated as a liquid. It was proved that such structures may have great computational abilities [4]. In this paper we simulate simple mammalian visual system using LSM model [7]. However, we do not analyse its computational power here. Instead of this, we concentrate on investigating the dynamics of *Primary Visual Cortex* (PVC) with energetic function known from informational theory [2].

Mammalian brains are built of neurons organised in microcircuits [5]. Microcircuits form columns, and the function of each column depends on its position in the brain.

In our model we use 4-neuron microcircuits. An input device (ID) representing the retina is built of 25 microcircuits. As a part of primary visual cortex and *Lateral Geniculate Nucleus* (LGN) – the liquid – we build a set of 25 Hodgkin-Huxley[1] Liquid State Machines (HHLSMs). There are 24 neurons in each HHLSM. Randomly chosen microcircuits of ID are connected with randomly chosen HHLSMs. The same architecture of connections is arranged between the cortex and the "readout" device (for detail see. Fig. 1) also consisting of 100 neurons.

[1] Neurons are built according to Hodgkin-Huxley model [5].

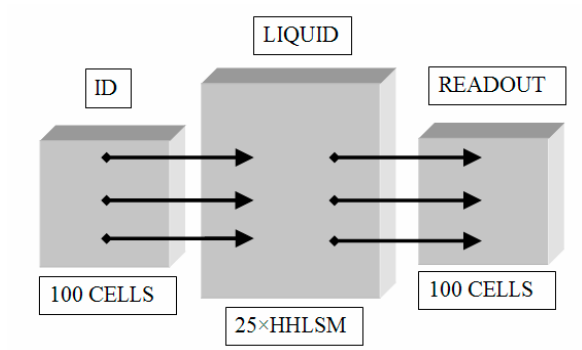


Fig. 1. Scheme of simulated biological-like visual system. For detail see text.

The readout plays role of PVC part and in this case it is specialised to movement detection. To achieve this, we arrange additionally special architecture of connections inside the readout itself. Weight values of these connections increase linearly in horizontal direction from left to right. The value of each weight is chosen using the following rule:

$$w = \alpha \cdot i \pm 5\% , \qquad (1)$$

where normalising constant α=10, and i is the number of readout's vertical row of cells which varies from 1 to 10. Readout neurons in each vertical row are full-connected

Pattern recognition

without auto-connections. Note that values of weights are randomly diffused according to (1) with the accuracy of ±5%. Fig. 2 represents typical maps of readout weights. One can choose some other architectures of readout connections, i.e. with weight values increasing exponentially.

There are also some random connections arranged among HHLSMs, however, their number is not very large (as in reality). All of the structures described above were simulated using the General Neural Simulation System (GENESIS) [1].

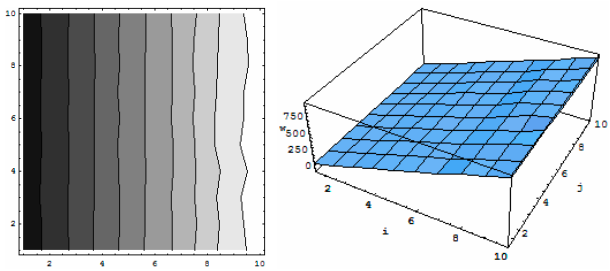


Fig. 2. Map of readout weights 2D in the left panel and 3D in the right panel.

The neurons of ID are then stimulated by series of four different changing in time – moving patterns (see Fig. 3). Nevertheless, our readout device is specialised only in the recognition of horizontal movements.

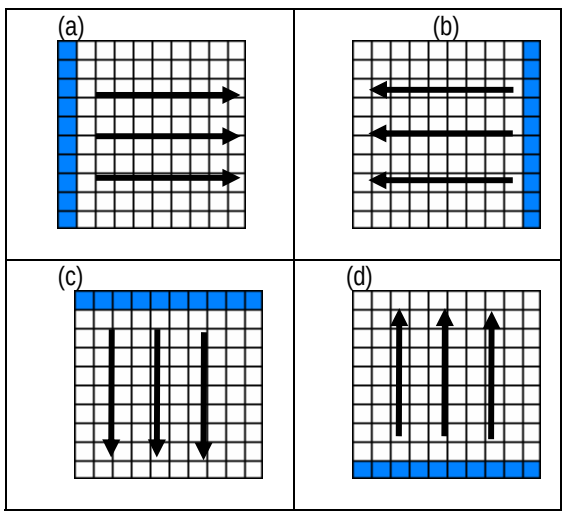


Fig. 3 Four different stimulation patterns: bar moving from left to right (a), from right to left (b), from top to bottom (c) and from bottom to the top of ID (d). Arrows present the direction of the movement.

2. Energetic Function and Results

Applications of informational theory and energetic function to analysis of signal processing were discussed in [6], [8]. For this contribution we will show that the energetic method can also be used for movement detection.

Energetic function is also a useful method in informational analysis for the discussed systems. For the i-th interval of time one can define its individual

$$T_0^i = \sum_{j,k}\sum_{l,m} w_{jklm} \cdot y_{jk} \cdot y_{lm} \qquad (2)$$

and global formula

$$T = \sum_t T_0^i = \sum_i \sum_{j,k}\sum_{l,m} w_{jklm} \cdot y_{jk} \cdot y_{lm}, \qquad (3)$$

where w_{jklm} is the weight of connection among the readout neurons with j,k and l,m coordinates. Activity y_{jk} (y_{lm}) corresponds to the mean value of the neuron j, k (l, m) potential (taken in time interval t_i=0.1).

Fig. 4 presents changes of readout's individual energetic function for the first pattern (Fig. 3a) during 100 ms of time. Results confirm our expectations that in such case the energetic function should be growing in time, together with weights' growth. Fig. 5 shows general tendency of energetic function growth calculated for time intervals of 5 ms.

Readout device specialised in this way can then be used as effective detector of every kind of movement in horizontal direction (compare Fig. 6 and Fig. 7). In contrast, readout with weights arranged in abovementioned way, cannot be applied to detection of vertical movements (Fig. 8 and Fig. 9).

We will not discuss the global energetic function here. However, the analysis of (3) proves that we would come to very similar conclusions.

3. Summary

In conclusion, we proved that in networks with especially arranged architecture the energetic function can be used as effective movement classifier. In near future, we will apply two readouts to our model and this will let us detect both horizontal and vertical movements. First applications of informational theory methods will probably appear soon in visual systems of intelligent robots, much later in medicine.

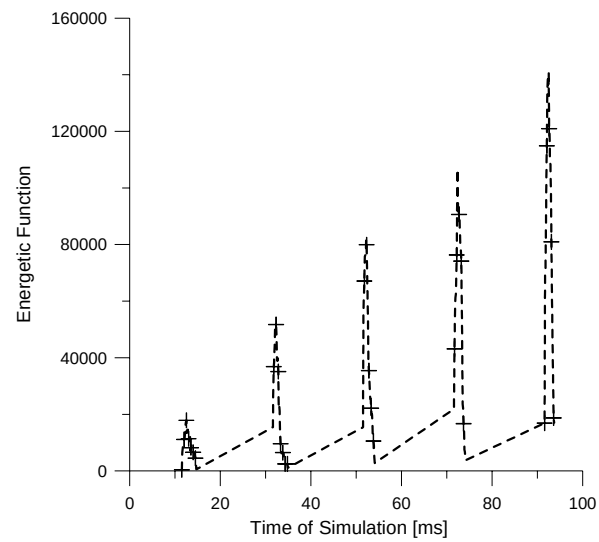


Fig. 4. Changes of readout's individual energetic function for the Fig. 3a pattern.

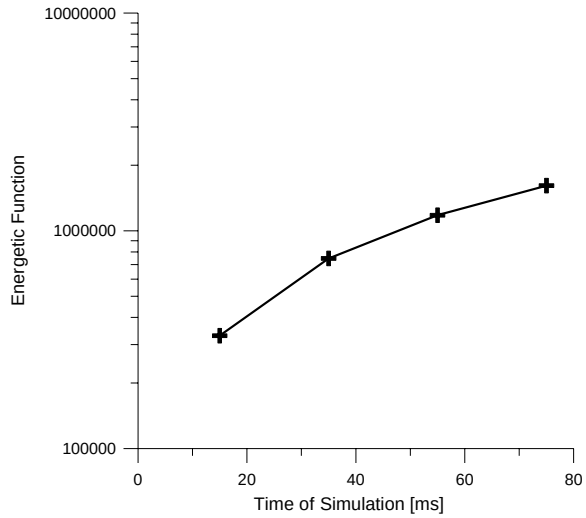


Fig. 5. Tendency of readout's energetic function changes for time interval of 5 ms (see: Fig. 4).

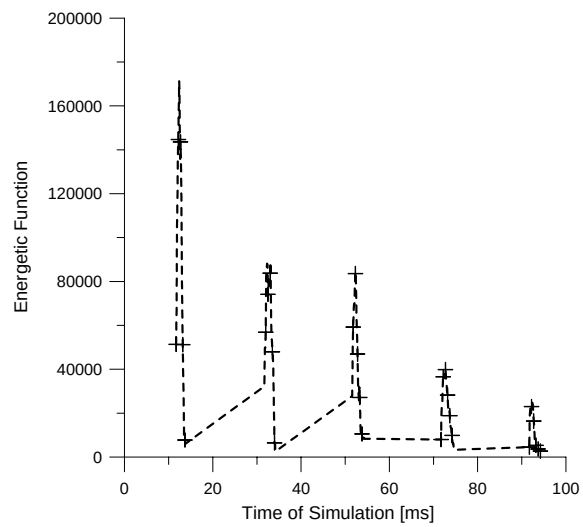


Fig. 6. Changes of readout's individual energetic function for the Fig. 3b pattern.

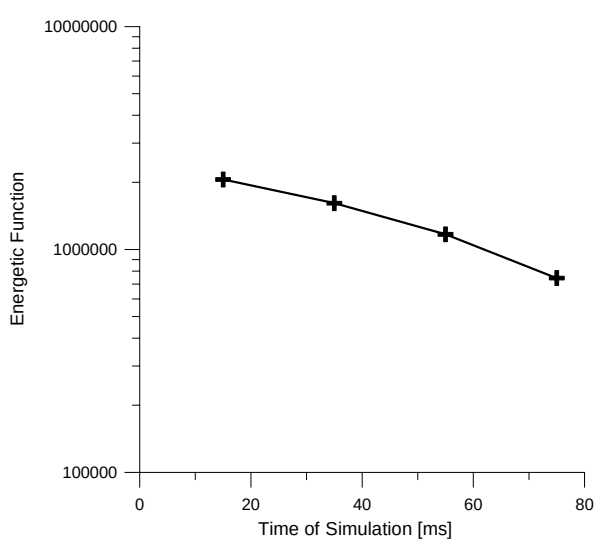


Fig. 7. Tendency of readout's energetic function changes for time interval of 5 ms (see: Fig. 6).

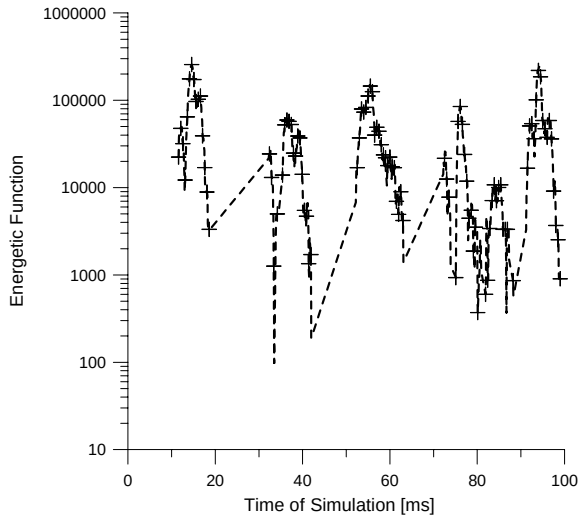


Fig. 8. Changes of readout's individual energetic function for the Fig. 3c pattern.

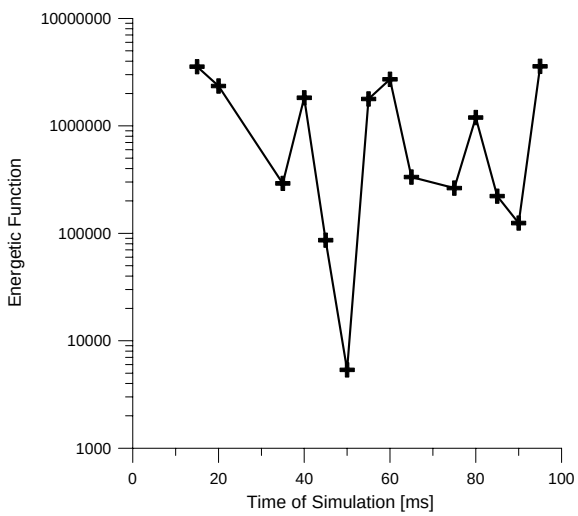


Fig. 9. Tendency of readout's energetic function changes for time interval of 5 ms (comp. Fig. 8).

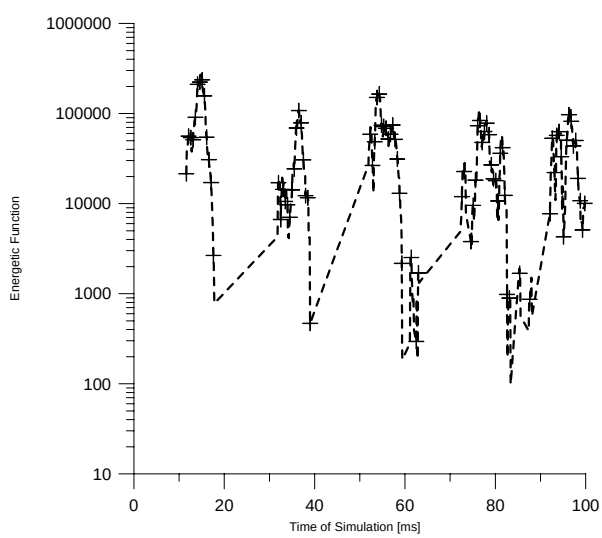


Fig. 10. Changes of readout's individual energetic function for the Fig. 3d pattern.

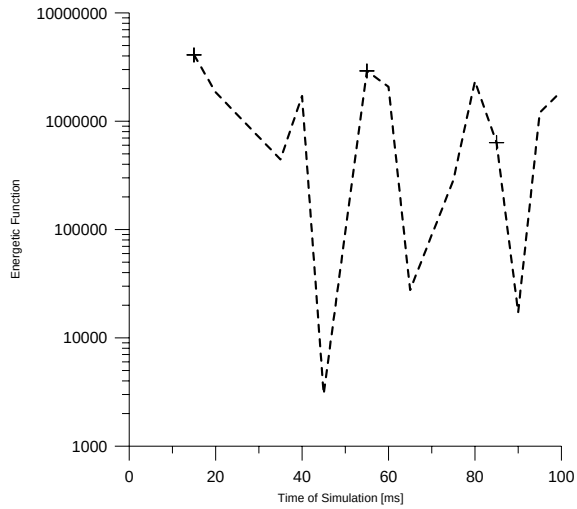
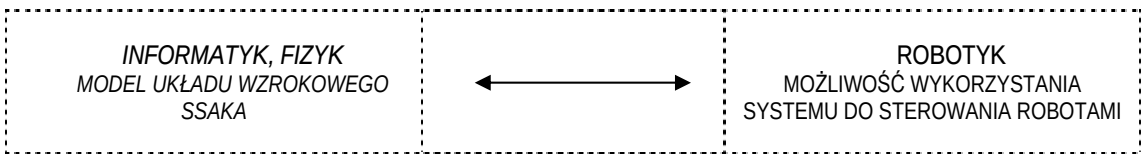


Fig. 11. Tendency of readout's energetic function changes for time interval of 5 ms (see: Fig. 10).

References

1. Bower J. M, Beeman D.: The Book of GENESIS – Exploring Realistic Neural Models with the GEneral NEural SImulation System, Telos, New York 1995.
2. Cover T., Thomas J.: Elements of Information Theory. Wiley, New York.
3. Hodgkin A., Huxley A.: Currents carried by sodium and potassium ions through the membrane of the giant axon of Loligio, J. Physiol., London 1952.
4. Holden A. V., Tucker J. V., Thompson B. C.: Can excitable media be considered as computational systems?, Physica D, 49, 240-246, 1991.
5. Gupta A., Wang Y., Markram H.: Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex, Science 287, 273-278, 2000.
6. Kamiński W. A., Wójcik G. M.: Liquid State Machines Built of Hodgkin-Huxley Neurons – Pattern Recognition and Informational Entropy, Annales UMCS Informatica, Vol. I, Lublin 2003.
7. Maass W., Natschlaeger T., Markram H.: Real-time Computing Without Stable States: A New Framework for Neural Computation Based on perturbation, Neural Computations, 14(11):2531-2560, 2002.
8. Wójcik G. M., Kamiński W. A.: Badania maszyn neuronalnych HHLSM metodami fizyki statystycznej, Obliczenia naukowe, (Polskie Towarzystwo Informatyczne), 9-16, Lublin 2003.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 325-332.

PROBLEMY ZWIĄZANE Z TOPOGRAFIĄ ELEKTROD DLA BADAŃ EEG - SYSTEM ICH DIGITALIZACJI

SOME PROBLEMS WITH TOPOGRAPHY OF ELECTRODES FOR EEG EXAMINATION – A SYSTEM OF THEIR DIGITALISATION

TOMASZ ZYSS

Klinika Psychiatrii Dorosłych Collegium Medicum Uniwersytetu Jagiellońskiego, Kraków
Pracownia EEG i PW, Klinika Neurologii, Szpital Uniwersytecki Kraków
mzzyss@cyf-kr.edu.pl

Streszczenie. Praca opisuje zagadnienia związane z lokalizacją elektrod do rejestracji czynności bioelektrycznej EEG w układzie standardowym i rozszerzonym. Zaprezentowano procedurę dokonywania pomiarów na głowie. Wskazano na konieczność stosowania dokładnych przestrzennych koordynat elektrod w badaniach naukowych. W rozwiązaniu problemu dokładnego rozlokowywania elektrod na powierzchni głowy może być pomocny system do digitalizacji.

Słowa kluczowe: elektrody do EEG, lokalizacja, digitalizacja topografii

Abstract. The work deals with problems connected with location of electrodes for recording EEG signal in standard and enhanced system (arrangement). A procedure of taking measurements on the head has been presented. The authors indicate the necessity of applying precise spatial coordinates of electrodes in scientific research. This can be done with the help of a digitizing system.

Key words: EEG electrodes, localization, topography digitization

1. Badanie EEG

Badanie EEG polega na rejestrowaniu czynności bioelektrycznej mózgu poprzez pomiar różnicy potencjałów między określonymi miejscami – punktami na powierzchni głowy badanej osoby. Rutynowe badanie EEG wykorzystuje przy tym 19-21 standardowych miejsc, w których umieszczane są elektrody rejestrujące. Wprowadzenie technik komputerowych umożliwiło dalszy rozwój techniki EEG. Możliwym stało się rejestrowanie sygnału EEG z większej ilości kanałów wraz z ich wizualizacją na ekranie komputera. Dla większej ilość rejestrowanych kanałów EEG – zniknęła również konieczność dokonywania w trakcie samego procesu rejestrowania sygnału EEG,tzw. remontaży. Sygnał biologiczny rejestrowany jest jednobiegunowo (aktywność spod czynnej elektrody *versus* „ziemia"). Dopiero przy przeglądaniu zapisu możliwe jest dokonywanie cyfrowych remontaży, polegających na przeliczeniu sygnału jednobiegunowego spod dwu wybranych elektrod na sygnał dwubiegunowy – dokładnie z tych samych dwóch elektrod [1].

2. Standardowy układ 10-20

Teoretycznie liczba miejsc na powierzchni głowy, z których możliwe jest rejestrowanie sygnału EEG, jest nieograniczona. W początkowym okresie stosowania EEG rozmaici neurofizjolodzy wykorzystywali dowolnie przez siebie wybrane miejsca na głowie, tworząc i propagując własne schematy topografii elektrod [4]. Tego rodzaju sytuacja wymagała pewnej unifikacji. Już na pierwszym Międzynarodowym Kongresie EEG, jaki miał miejsce w Londynie w 1947 r., stwierdzono konieczność opracowania pewnego standardu lokowania elektrod EEG na głowie.

Jedenaście lat później (1958 r.) Jasper zaproponował lokalizowanie elektrod według układ (schematu) 10-20, który stał się obowiązującym we wszystkich pracowniach EEG [2].

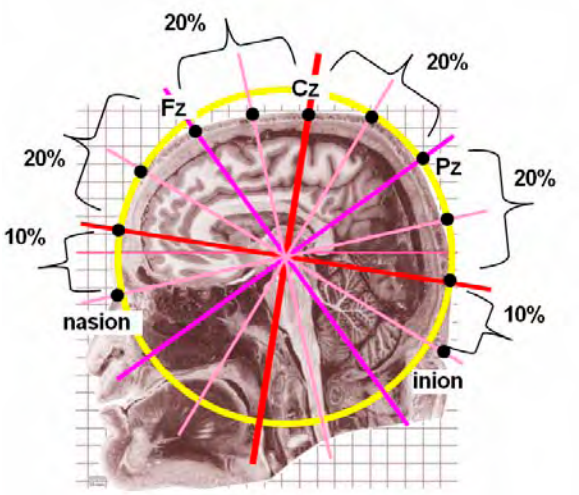
Zasadniczy problem w stworzeniu tego rodzaju zunifikowanego systemu związany był z wielkością głowy człowieka. U człowieka dorosłego obwód głowy wahać się może od 46 do 62 cm. Należy również pamiętać, iż głowa dziecka jest mniejsza niż dorosłego człowieka i zwiększa się z wiekiem – od 22 do 46 cm. Ponadto określone okolice na powierzchni głowy człowieka odpowiadają zasadniczo określonym okolicom mózgu ludzkiego. Tym samym odległości między elektrodami nie mogły być pewnymi sztywnymi wartościami bezwzględnymi, lecz wartościami względnymi wyrażanymi albo w odległościach kątowych, albo w procentach. I właśnie system 10-20 polega na oznaczaniu odległości miedzy elektrodami w odległościach procentowych.

3. Zasady lokalizacji elektrod

Wyznaczanie lokalizacji elektrod według układu 10-20 polega na znalezieniu kilku stałych (niezmiennych) punktów

orientacyjnych na powierzchni głowy, dokonywaniu pomiarów odległości między tymi punktami i dzieleniu tych odległości na 10- lub 20-procentowe odcinki. Zasadnicze pomiary dokonuje się w trzech płaszczyznach: strzałkowej, wieńcowej i poziomej [3].

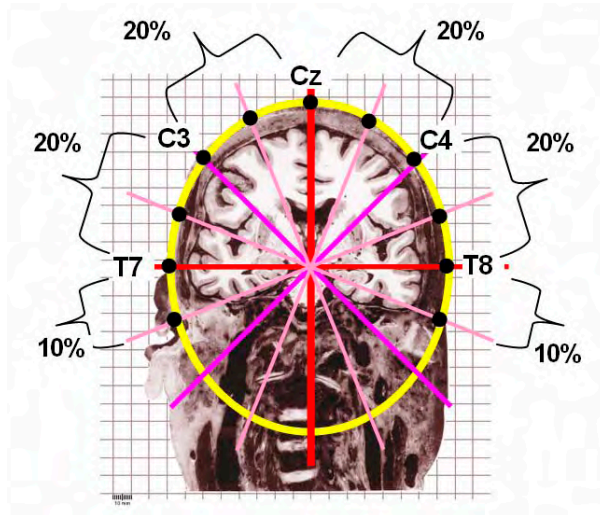
1) Wyznaczenie elektrod w płaszczyźnie strzałkowej:
Na powierzchni głowy wyszukuje się dwa punkty: *nasion* (Nz – zagłębienie na szczycie nosa) i *inion* (Iz – wypukłość na postawie potylicy). Oba punkty łączy się linią przechodzącą przez szczyt głowy. W punktach Nz i Iz nie są lokowane żadne elektrody – są to jedynie punkty orientacyjne (przynajmniej w układzie 10-20). W połowie odległości (50%) między punktami Nz i Iz znajduje się punkt szczytowy głowy (vertex), gdzie lokowana jest elektroda centralna Cz.
Odległość Nz-Iz dzieli się na 10 równych odcinków (np. dla odległości Nz-Iz równej 42 cm długość 10% odcinka będzie wynosić 4,2 cm). W punktach odległych o 10% od Nz i Iz podstawowy układ 10-20 nie przewiduje żadnej elektrody; w układzie rozszerzonym (dla większej ilości elektrod) lokowane są tam odpowiednio następujące elektrody: Fpz (frontopolar – czołowa biegunowa lub przedczołowa) i Oz (occipital – potyliczna). Przez punkty te przechodzi płaszczyzna pozioma (równik głowy). W odległości dalszych 20% (stąd układ 10-20) lokowane są kolejne dwie elektrody: Fz (frontal – czołowa) i Pz (parietal – ciemieniowa).



Rys. 1. Płaszczyzna strzałkowa

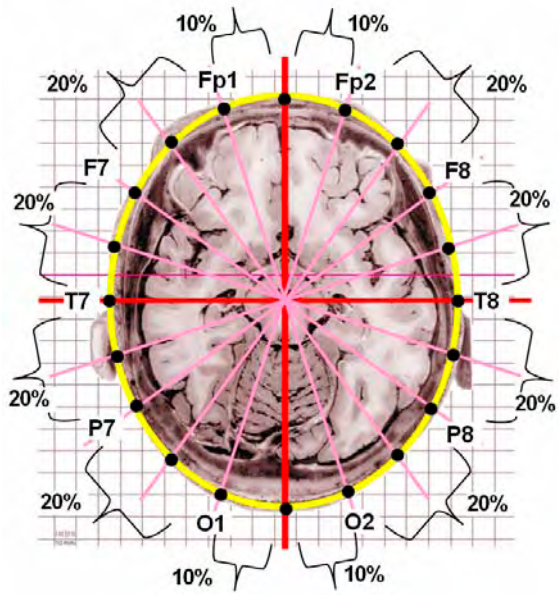
Litera z (zero) określa elektrody znajdujące się na linii środkowej (płaszczyźnie strzałkowej). Została ona wybrana zamiast cyfry zero (0), aby nie dochodziło do pomyłek z literą O, oznaczającą okolicę potyliczną (O – occipital).

2) Wyznaczenie elektrod w płaszczyźnie wieńcowej:
Punktami orientacyjnymi w przypadku płaszczyzny wieńcowej są punkty leżące tuż przed skrawkiem ucha (chrząstka wystająca z przodu ucha zewnętrznego – tragus). U układzie rozszerzonym elektrod w punktach tych lokują się elektrody skroniowe (temporal) T9 i T10. Płaszczyznę wieńcową tworzy linia łącząca punkty T9 i T10 przez elektrodę Cz.
Odległość T9-T10 dzieli się na 10 równych odcinków. W punktach odległych o 10% od T9 i T10 lokowane są odpowiednio następujące elektrody środkowo-skroniowe: T7 (po stronie lewej; dawniej T3) i T8 (po stronie prawej; dawniej T4). W odległości dalszych 20% odległości lokowane są kolejne dwie elektrody centralne: C3 i C4 (central).
Odległość T7-C3 jest równa C3-Cz (i analogicznie T8-C4 = C6-Cz). Z faktu tego wynika, iż elektrody C3 i C4 lokowane są w połowie (50%) odcinka odpowiednio T7-Cz i T8-Cz.
Elektrody C3 i C4 zaliczane są do elektrod leżących w oddaleniu o 20% od linii przyśrodkowej (*20% parasagittal electrodes*); zgodnie z tą konwencją elektrody T7 i T8 to elektrody leżące w oddaleniu o 50% od linii przyśrodkowej (*50% parasagittal electrodes*).
Wszystkie elektrody położone po stronie lewej określane są liczbami nieparzystymi (1, 3, 5, 7, 9), a po stronie prawej – parzystymi (2, 4, 6, 8, 10).



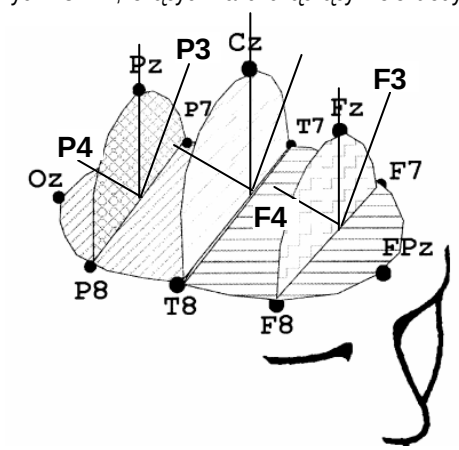
Rys. 2. Płaszczyzna wieńcowa

3) Wyznaczenie elektrod w płaszczyźnie poziomej:
Przy wyznaczaniu płaszczyzny poziomej wykorzystuje się już wcześniej poczynione pomiary. Płaszczyzna ta przechodzi przez punkty oddalone o 10% od punktów Nz i Iz (odpowiednio Fpz i Oz w układzie rozszerzonym) oraz przez punkty oddalone o 10% od okolicy przed skrawkami usznymi (odpowiednio T7 i T8). Płaszczyzna ta odpowiada południkowi zerowemu, tj. równikowi mózgu.
Odległość T3-T4 do przodu i do tyłu dzielona jest na 10 równych odcinków. Punkty leżące w odległości 10% od punktu powyżej *nasion* (Fpz) określają położenie elektrod czołowych biegunowych Fp1 i Fp2 (frontopolar). W odległości 10% od punktu powyżej *inion* (Oz) znajdują się elektrody potyliczne O1 i O2.
Punkty zlokalizowane o kolejne 20% długości ku tyłowi od Fp1 i Fp2 wyznaczają położenie kolejnych elektrod F7 i F8 (czołowe dolne). Analogicznie - elektrody położone o dalsze 20% długości ku przodowi od elektrod O1 i O2 to elektrody P7 i P8 (dawniej określane jako tylnio-skroniowe T5 i T6).



Rys. 3. Płaszczyzna pozioma

4) Wyznaczenie ostatnich elektrod:
Standardowy układ elektrod 10-20 składa się z 19 (lub 21, jeśli uwzględnić elektrody Fpz i Oz) elektrod. Dzięki powyżej zaprezentowanym zasadom wyznaczona została lokalizacja 17 elektrod. Dwie kolejne – tzw. uszne (A1 i A2) umieszczane są na płatkach uszu. W standardowym układzie 10-20 służą one jako elektrody referencyjne, względem których rejestrowany jest sygnał w przypadku odprowadzeń tzw. jednobiegunowych. W rozszerzonym układzie, gdzie wykorzystuje się większą ilość elektrod, elektrody A1 i A2 mają swoją dokładną topograficzną lokalizację i służą do odbioru sygnału bioelektrycznego z okolic środkowo-skroniowych.
Pozostałe cztery elektrody zaliczane są do elektrod leżących w oddaleniu o 20% od linii przyśrodkowej (*20% parasagittal electrodes*). I tak elektroda czołowa lewa F3 leży w odległości 20% od elektrody Fz w kierunku do elektrody F7. Ponieważ również odległość F3-F7 jest równa 20% długości, to dla celów praktycznych uznaje się, iż elektroda F5 leży w połowie odcinka Fz-F7 (przez podobieństwo z lokalizacją elektrod C3 i C4). Analogicznie – prawa elektroda czołowa F4 znajduje się w połowie odległości odcinka Fz-F8. Na takiej samej zasadzie wyznacza się położenie elektrod ciemieniowych P3 i P4, leżących na łuku łączącym elektrody P7-Pz-P8.



Rys. 4. Lokalizacja elektrod F3, F4, P3 i P4

4. Metody przestrzennej lokalizacji elektrod

Opisane powyżej zasady pozwalają na średnio precyzyjne rozmieszczanie podstawowych 19 elektrod w standardowym układzie 10-20 do zwykłego rutynowego badania EEG. Współczesne badania naukowe polegają jednak na rejestrowaniu sygnału EEG lub tzw. potencjałów wywołanych zwłaszcza typu poznawczego (ERP – *event related potentials*) z określonych generatorów mózgowych (niekiedy niewielkich okolic-centrów kory mózgowej). Badania takie wymagają bardzo precyzyjnego rozlokowywania elektrod, jak również dokładnego definiowania ich położenia.

Wykorzystuje się przy tym definiowane położenia elektrod zgodnie z układem kartezjańskim (osie XYZ) oraz układem biegunowym/sferycznym (kąt theta – z osią Z, oraz kąt phi – z osią X). Oba układy są wzajemnie przechodnie zgodnie z równaniami:

$$r = \sqrt{x^2 + y^2 + z^2},$$

$$x = r \sin\theta \cos\varphi,$$

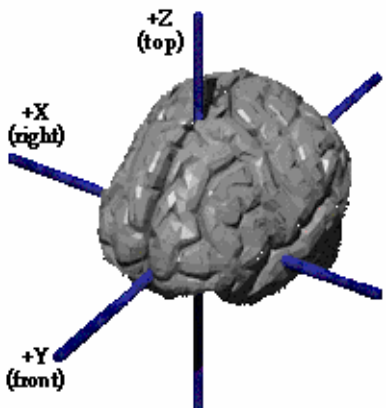
$$y = r \sin\theta \sin\varphi,$$

$$z = r \cos\theta,$$

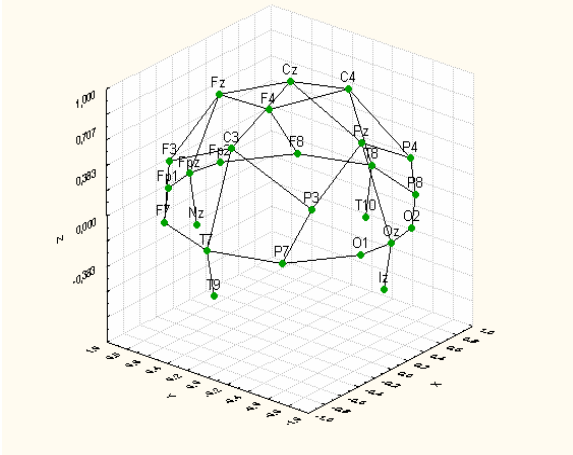
$$\varphi = \arctan\frac{y}{x},$$

$$\theta = \arctan\frac{\sqrt{x^2 + y^2}}{z} = \arcsin\frac{z}{\sqrt{x^2 + y^2 + z^2}} = \arccos z$$

Brak jest jednak standardów w zakresie definiowana osi systemu kartezjańskiego. W badaniach własnych wykorzystano konwencję, według której oś X wskazuje w kierunku lewego ucha, oś Y – w kierunku nosa, a oś Z – do szczytu głowy. Istnieją jednak systemy, w których osie XYZ są przesunięte o 90-180^0.



Rys. 5. Układ kartezjański do lokalizacji elektrod EEG



Rys. 6. Przestrzenna lokalizacja elektrod EEG

5. Technika hrEEG

W standardowym badaniu EEG sygnał bioelektryczny rejestruje się z 19(21) elektrod rozlokowanych na powierzchni skóry głowy. Dla celów kliniczno-diagnostycznych jest to liczba w pewnym zakresie wystarczająca.

Wymieniona liczba elektrod pozwala na umieszczenie od 1 do 3 elektrod nad każdym z głównych płatów mózgowych (3 pary elektrod nad płatami czołowymi, 2-3 pary elektrod nad okolicami skroniowymi, po jednej parze nad płatami potylicznymi i ciemieniowymi oraz okolicą centralną oraz 3-5 elektrod w linii pośrodkowej mózgu). Odległość między poszczególnymi elektrodami może wahać się w przypadku głowy dorosłego człowieka między 3 a 8 cm. Tym samym możliwym jest rejestrowanie aktywności znad odległych od siebie generatorów mózgowych.

W przypadku oceny tzw. zapisu analogowego (krzywych EEG), opisana „dokładność" metody jest wystarczająca. Dodatkowe informacje o czynności EEG można pozyskać, stosując określone montaże, tj. połączenia między elektrodami (tzw. odprowadzenia dwubiegunowe, w przeciwieństwie do odprowadzeń jednobiegunowych, gdzie sygnał z jednej czynnej elektrody porównywany jest z elektrodą uziemioną czy uśrednioną).

Aktualne badania naukowe mózgu próbują jednak korelować czynność EEG określonych – niekiedy punktowych – okolic mózgu z rozmaitymi zjawiskami neuropsychologicznymi. W takim przypadku liczba 19-21 elektrod jawi się jako dalece niewystarczająca.

W przypadku tzw. mapowania (napięciowego lub częstotliwościowego), czynności EEG aktywność bioelektryczna okolic między elektrodami wyliczana jest na zasadzie estymowania odpowiednich wartości spod sąsiadujących elektrod. Tym samym dla dużych odległości między elektrodami wartości estymowane mogą dalece odbiegać od wartości rzeczywistych.

Jeszcze do niedawna sądzono, że wystarczającą ilością odprowadzeń może być liczba ok. 100-128 elektrod. Przy tej ilości elektrod (znajdują się one w odległości około 2,5 cm jedna od drugiej) – rejestrując pola elektryczne pochodzące od różnych generatorów korowych. Wydawało się wówczas, iż dalsze zwiększanie liczby elektrod, a więc zmniejszenie odległości pomiędzy nimi, prowadzi już nie do zwiększenia rozdzielczości przestrzennej, lecz do rejestracji tych samych potencjałów przez różne (sąsiednie) elektrody. Aktualnie jednak przyjmuje się tę zasadę jako nieobowiązującą i konstruuje się zestawy EEG z możliwością rejestracji sygnału EEG z coraz większej ilości elektrod.

Wprowadzenie technik komputerowych umożliwiło „rozwój techniki EEG wysokich rozdzielczości (hr-EEG = *high resolution EEG*), polegającej na równoczesnej rejestracji czynności bioelektrycznej mózgu z kilkudziesięciu, a nawet kilkuset elektrod zlokalizowanych na powierzchni głowy osoby badanej. Aktualnie najbardziej zaawansowane systemy cyfrowego EEG umożliwiają rejestrację sygnału z 512 elektrod.

Podstawą stosowania techniki hr-EEG było stworzenie szybkich wzmacniaczy elektro-fizjologicznych, umożliwiających próbkowanie sygnału EEG z częstotliwością przynajmniej 200 Hz. W przypadku badań polegających na analizie tzw. poznawczych potencjałów wywołanych (ERP – *event related potentials*) wymagana jest znacznie wyższa częstotliwość próbkowania sygnału – około 1-2 kHz.

Formalnie nie konstruuje się wzmacniaczy z bardzo dużą ilością kanałów. Zestawy do hrEEG konstruowane są poprzez równoległe łączenie 2-4 głowic wzmacniaczy – zwykle dzięki zastosowaniu interfejsów z portami typu USB.

W rozwój techniki hr-EEG duży wkład miał również sam postęp technik komputerowych. Duże pamięci masowe są w stanie magazynować olbrzymie pliki wynikowe (dla wyobrażenia: zapis sygnału EEG z 23 elektrod z częstotliwością próbkowania 600 Hz i o czasie trwania 10 minut osiąga wielkość około 20 Mb pamięci). Dopiero technologia Windows (szczególnie XP) radzi sobie z analizą *on-line* sygnału EEG i jego wizualizacją na ekranie komputera.

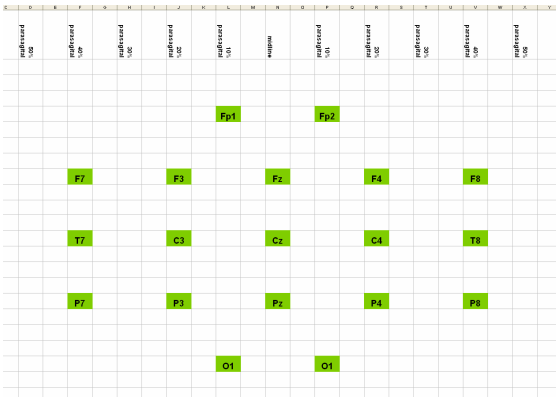
6. Rozszerzony układ 10-10 (10%)

Wykorzystanie dużej ilości elektrod pociąga za sobą konieczność dokładnego definiowana ich położenia, jak również precyzyjnego rozmieszczania na powierzchni głowy. Zaproponowany przez Jaspera w 1958 r. układ 10-20 okazał się być układem otwartym. Pierwszym, który zaproponował rozszerzenie ilości elektrod z 22 do 74 był Chatrian i wsp. (1985 r.). Opracowany przez niego system lokalizacji elektrod został nazwany rozszerzonym układem 10-20 lub też układem 10-10 (10%) [5]. Propozycja Chatriana i wsp. została zaakceptowana i jest uznawanym standardem Amerykańskiego Towarzystwa EEG (American Electroencephalographic Society), jak również Międzynarodowej Federacji Towarzystw EEG i Klinicznej Neurofizjologii (International Federation of Societies for Electroencephalography and Clinical Neurophysiology).

W przypadku standardowego układu 10-20, poszczególne elektrody lokalizowane są w odległościach 10 lub 20% odcinków między stałymi punktami głowy. W układzie rozszerzonym zwiększona została „gęstość" elektrod poprzez wykorzystanie już tylko 10% odległości.

Z historycznego punktu widzenia należy stwierdzić, iż już układ Jaspera przewidywał możliwość wprowadzenia dodatkowych elektrod poprzez niewykorzystanie liczb 1 i 5 dla elektrod lewostronnych (lub 2 i 6 po stronie prawej); i tak elektroda C3 miała być elektrodą leżącą pomiędzy elektrodami C1 i C5.

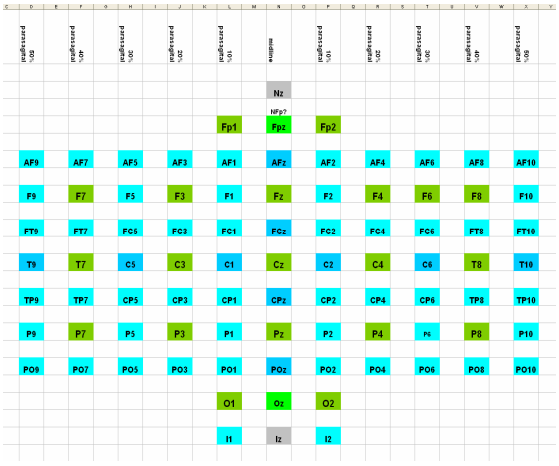
Nazwy dodatkowych elektrod lokowanych w odstępach 10% zostały poddane adaptacji i modyfikacji komitetu ds. standaryzacji. Trzeba zaznaczyć, iż procedury unifikacyjne doprowadziły do tego, że niektóre starsze nazwy elektrod zostały zastąpione nowymi. I tak elektrody T3 i T4 – leżące w oddaleniu o 40% od linii przyśrodkowej (*40% parasagittal electrodes*) zostały zastąpione nazwami - odpowiednio: T7 i T8; bowiem wszystkie elektrody leżące w tej odległości procentowej od linii przyśrodkowej są określane liczbami 7 i 8. Podobna zmiana objęła również elektrody T5 i T6, które otrzymały nowe nazwy – P7 i P8. Wymienione elektrody są elektrodami leżącymi na równiku głowy.



Rys. 7. 19 elektrod w standardowym układzie 10-20

Do rozszerzonego układu 10-10 trafiły elektrody, których lokalizacja musiała być wyznaczona już w układzie standardowym 10-20, jak: Fpz (biegunowo-czołowa) i Oz (potyliczna) – położone na linii środkowej z (zero) oraz elektrody skroniowe T9 i T10 (okolica przed skrawkami usznymi) stanowiące punkty konieczne do wykreślenia płaszczyzny wieńcowej.

Na linii środkowej wyznaczone zostały następujące elektrody: AFz, FCz (czołowo-centralna), CPz (centralno-ciemieniowa), POz (ciemieniowo-potyliczna). Kombinacja pierwszych dwóch liter mówi o położeniu danej elektrody: elektroda FCz umieszczana jest w połowie odległości między elektrodami Fz i Cz, elektroda CPz – w połowie odległość Cz i Pz itp. Zgodnie z tą konwencją elektroda położona w połowie odległości między elektrodami Fp i F winna mieć określenie FpFz. Określenie to zostało jednak zastąpione nazwą AFpz (*anterio-frontal* – przednio-czołowa).



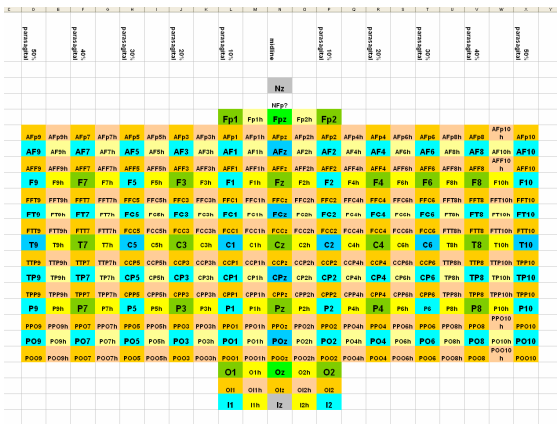
Rys. 8. 87 elektrod w rozszerzonym układzie 10-10

Wszystkie elektrody oznaczone liczbami 7 i 8 znajdują się na płaszczyźnie poziomej (równiku głowy). Elektrody o numerach 3 i 4 umiejscawiane są w połowie odległości między odpowiednią elektrodą z linii środkowej (z), a elektrodą 7 lub 8. Z kolei elektrody znaczone cyframi 1 znajdują się w odległości 50% od elektrod z linii środkowej-strzałkowej (z) do elektrody 3. W połowie odcinka między elektrodą oznaczoną cyfrą 3 a elektrodą 7 znajduje się miejsce dla szeregu elektrod o numerze 5. Odpowiednie zasady obowiązują dla elektrod po stronie prawej, oznaczonych cyfrą 6.

Rząd elektrod biegunowo-czołowych Fp kończy swoją numerację na cyfrze 1 i 2; podobnie sytuacja ma się w przypadku elektrod potylicznych O i elektrod I – leżących na bok od punktu-elektrody inion.

Elektrody z rzędu elektrod AF, F, P i PO – między numerami 1 i 7 po stronie lewej (oraz numerami 2 i 7 po stronie prawej) nie zmieniają swoich literowych określeń. Elektrody z rzędu elektrod FC (czołowo-centralne) kończą swoje literowe określenie na elektrodach o numerze 5 i 6. Elektrody 7-9 i 8-10 noszą w tym rzędzie litery FT (czołowo-skroniowe). W przypadku rzędu centralnego C ostatnimi elektrodami są elektrody skroniowe T, a w przypadku rzędu CP (centralno-ciemieniowego) – elektrody skroniowo-ciemieniowe TP.

Elektrody o liczbach 9 i 10 leżą już o 10% dalej - poniżej równika głowy. Teoretycznie możliwe jest jeszcze niższe rozmieszanie elektrod. W literaturze można znaleźć takie nazwy elektrod, jak: TP11 i TP12 oraz PO11 i PO12.



Rys. 9. 357 elektrod w rozszerzonym układzie 10-5

Badania literaturowe pozwalają znaleźć inne próby definiowania elektrod w układzie rozszerzonym 10-10. I tak np. oryginalna propozycja Chatriana opierała się na wykorzystaniu znaczników zwanych „prim" (‘). Pojedynczy prim otrzymywała elektroda leżąca za elektrodą standardową: np. elektroda leżąca między elektrodami C4 i P4 otrzymywała nazwę C4’. Podwójnym primem znaczona była elektroda do przodu od elektrody standardowej: elektroda C4’’ to elektroda między C4 i F4.

W innych propozycjach schematów proponowano wykorzystanie dodatkowych liter B, D, E i H do oznakowywania dodatkowych płaszczyzn wieńcowych. Sugerowano dodawanie liter a *vel* A (*anterior*) lub p *vel* P (*posterior*) do oznakowywania dodatkowych elektrod znajdujących się do przodu lub do tyłu od elektrod z układu standardowego, np. FA zamiast AFF, FP zamiast FC, PA zamiast CP czy PP w miejsce PO. Wymienione propozycje nie zostały jednak powszechnie zaakceptowane.

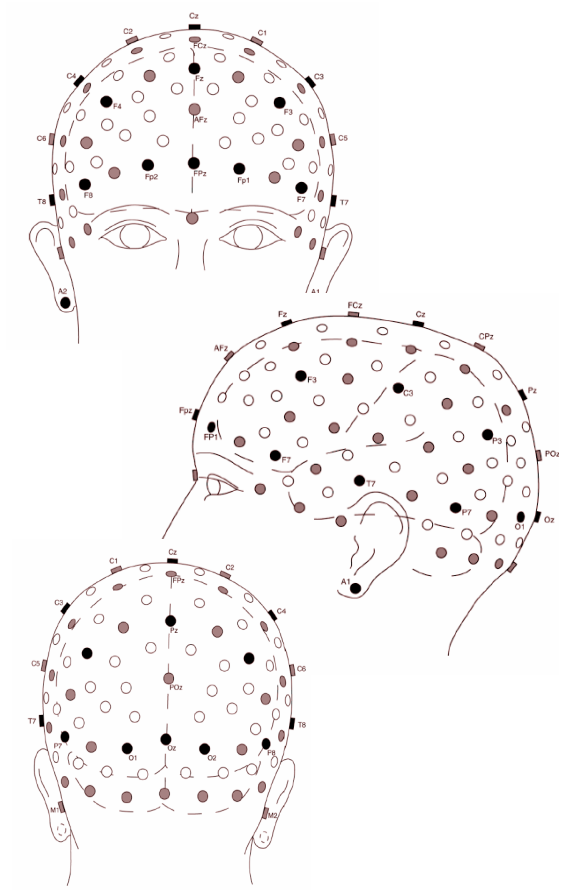
7. Układ 10-5 lub 5-5 (5%)

Naprzeciw zapotrzebowaniu na zdefiniowanie jeszcze większej ilości elektrod wyszli w 2001 r. Oostenveld i Praamstra [7]. Wykorzystując dalsze zagęszczanie elektrod, zaproponowali oni układ 10-5 polegający na wykorzystaniu 5% odległości między kolejnymi elektrodami.

Oparto się na wyżej opisanej konwencji tworzenia kolejnych rzędów (płaszczyzn wieńcowych) elektrod, polegającej na łączeniu liter nazw elektrod zewnętrznych do tworzenia nazwy elektrody leżącej pośrodku.

I tak elektroda leżąca w połowie odległości między elektrodami AF i F to elektroda AFF; elektroda PPO – to elektroda położona w ½ odcinka między elektrodami P i PO.

W standardowym układzie 10-20 na linii środkowej znajdują się jedynie trzy elektrody (wszystkie ze znacznikiem z – zero): F, C i P. Rozszerzony układ 10-10 pozwala na wyznaczenie następujących elektrod: N, Fp, AF (zamiast FpF), F, FC, C, CP, P, PO, O, I. Dalsze zwiększenie ilości elektrod, proponowane przez Oostenvelda i Praamstrę, pozwoliło na utworzenie następujących rzędów elektrod: N, Fp, AFp (zamiast FpAF), AF, AFF, F, FFC, FC, FCC, C, CCP, CP, CPP, P, PPO, PO, POO, O, OI, I. Nie określono nazwy elektrody leżącej między elektrodą Nz i Fpz. Nadawanie numerów poszczególnych elektrod podlega takim samym zasadom, jak w przypadku układu 10-10. Wprowadzenie dodatkowych płaszczyzn wieńcowych z nowymi rzędami elektrod umożliwiło zdefiniowanie łącznie 178 elektrod.



Rys. 10. Topografia elektrod w układzie 5-5

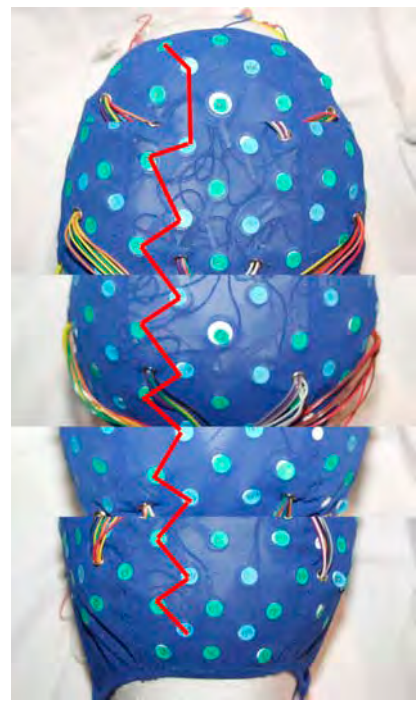
Dla stworzenia szeregów dodatkowych elektrod leżących o 5% odległości od istniejących już płaszczyzn strzałkowych (*5, 15, 25, 35, 45% parasagittal electrodes*) wykorzystano znacznik literowy „h" (od *half* – połowa). I tak elektroda leżąca w połowie odległości między elektrodami Cz i C1 to elektroda C1h. Elektroda o nazwie F6h to elektroda znajdująca się w połowie drogi między elektrodami C4 i C6.

Łącznie układ 10-5 umożliwia zdefiniowanie 357 elektrod.

Jak się wydaje, Oostenveld i Praamstra skorygowali błąd Nuwera i wsp., którzy zaliczyli elektrody Fp1 i Fp2 do elektrod położonych na płaszczyźnie oddalonej o 30% (zamiast o 10%) od płaszczyzny strzałkowej. Również błędnie zostały określone przez Nuwera i wsp. elektrody O1 i O2 jako oddalone o 20 zamiast 10% od płaszczyzny strzałkowej [8].

8. Problemy praktyczne z techniką hr-EEG

Samo określenie nazwy elektrody to jeden problem, innym jest natomiast zdefiniowanie jej dokładnego położenia w koordynatach kartezjańskich XYZ czy biegunowych. W przypadku badań rutynowych, wykorzystujących standardowe 19-21 elektrody (specjalne „grzybki" elektrodowe zakładane ręcznie jeden za drugim na głowę osoby badanej) ogólne reguły układu 10-20 są wystarczające, aby z dość dobrą precyzją rozmieścić elektrody. Manualne zakładanie większej ilości elektrod – ze względu na długi czas ich zakładania, jak i dokładność ich rozmieszczania – jawi się jako trudne do zaakceptowania. Dość dobrym wyjściem rozwiązującym ten problem jest zastosowanie specjalnych komercyjnych czepków, gdzie elektrody są na sztywno zintegrowane z elastycznym materiałem czepka. W przypadku zakładania czepka na większą lub mniejszą głowę rozciągnięciu ulega materiał elastyczny czepka gwarantując zachowanie stałych odległości (wyrażanych w procentach odległości) między poszczególnymi elektrodami.



Rys. 11. 128-elektrodowy czepek Elektro-Cap

Wykorzystanie do konstrukcji czepka określonych materiałów (elektrody ze spiekanego srebra, kabelki z włókien węglowych) umożliwia wykonywanie badań EEG równocześnie z badaniami neuroobrazowania w skanerze MRI.

Badania własne czepka Elektro-Cap wykazały jednak, iż jakość czy dokładność wykonania czepka pozostawia wiele do życzenia. I tak linia łamania na rysunku 18 łączy elektrody oznaczone cyfrą 1, które powinny znajdować się na jednej linii.

Na poziomie dalszego przetwarzania cyfrowego zarejestrowanego sygnału EEG istnieje jeszcze jeden problem techniczny. Standardowy układ 10-20 został skonstruowany w ten sposób, że głowa ludzka została uznana za idealną sferę (kulę) o tym samym promieniu r na wszystkich osiach. Tymczasem głowa człowieka jest pewnego rodzaju elipsoidą z najdłuższym wymiarem w osi strzałkowej. Stąd też matematycznie wyznaczone koordynaty elektrod niezbyt nadają się do zastosowania w zaawansowanych badaniach naukowych nad czynnością naturalnych generatorów mózgowych czynności EEG rzutowanych na realistyczny model głowy.

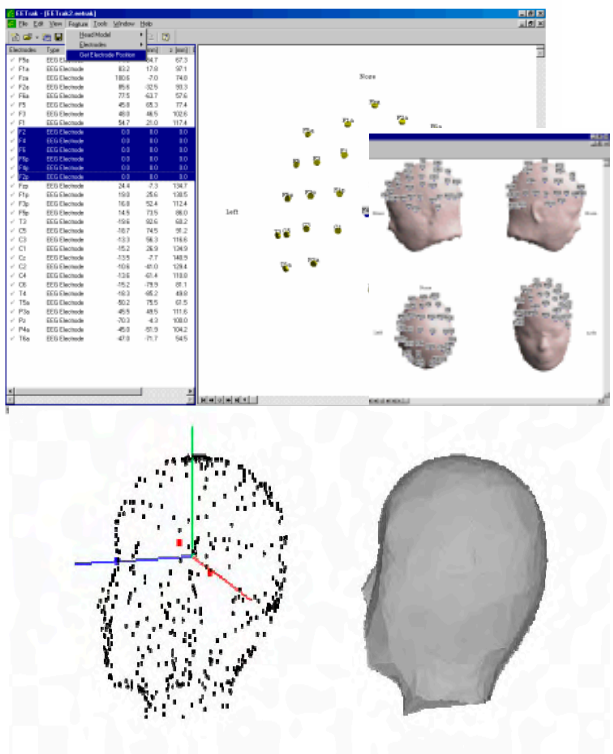
W tabeli obok zostały zaprezentowane wartości koordynat kartezjańskich dla standardowych 19 elektrod, wyliczonych według zasad układu 10-20 (idealnych) oraz estymowanych z realistycznego modelu powierzchni głowy (realnych). Podane wartości są wyliczone dla promienia głowy r=1. W przypadku dokonywania obliczeń dla rzeczywistych wymiarów głowy dokonuje się procedury przeliczenia (*fitting*). O ile jeszcze wartości dla elektrod położonych na linii strzałkowej oraz linii wieńcowej są w dalekim zakresie równe, to dla takich elektrod, jak F3, F4, P3 i P3 stwierdza się znaczne różnice w wartościach. Pełna lista zawierająca koordynaty dla wszystkich elektrod układu 10-5 znajduje się na stronie internetowej Oostenwelda.

1x3	1 name	2 X	3 Y	4 Z	5	6 X	7 Y	8 Z
1								
2		KOORDYNATY IDEALNE				KOORDYNATY REALNE		
3								
4	Fp1	-0,309	0,951	0		-0,309	0,9511	0,0001
5	Fpz	0	1	0		0	1	0,0001
6	Fp2	0,309	0,951	0		0,3091	0,9511	0
7	F7	-0,809	0,588	0		-0,809	0,5878	0
8	F3	-0,6	0,703	0,383		-0,5399	0,6726	0,5043
9	Fz	0	0,707	0,707		0	0,7067	0,7067
10	F4	0,6	0,703	0,383		0,5399	0,6726	0,5043
11	F8	0,809	0,588	0		0,809	0,5878	0
12	T7	-1	0	0		-1	0	0
13	C3	-0,707	0	0,707		-0,7066	0,0001	0,7066
14	Cz	0	0	1		0	0,0002	1
15	C4	0,707	0	0,707		0,7066	0,0001	0,7066
16	T8	1	0	0		1	0	0
17	P7	-0,809	-0,588	0		0,809	-0,5878	-0,0001
18	P3	-0,6	-0,703	0,383		-0,5401	-0,6724	0,5045
19	Pz	0	-0,707	0,707		0	-0,7063	0,7065
20	P4	0,6	-0,703	0,383		0,5401	-0,6724	0,5045
21	P8	0,809	-0,588	0		0,809	-0,5878	-0,0001
22	O1	-0,309	-0,951	0		-0,309	-0,9511	0
23	Oz	0	-1	0		0	-1	0
24	O2	0,309	-0,951	0		0,309	-0,9511	0

Rys. 12. Koordynaty idealne i realne 19 elektrod w układzie 10-20

Dla przezwyciężenia problemów związanych z możliwie jak najbardziej dokładnym położeniem elektrod na głowie osoby badanej od kilku ostatnich lat do dyspozycji stoją specjalistyczne digitizery (np. 3Space Fastrack firmy Polhemus), które czynią możliwym oznaczanie punktów w trójwymiarowej przestrzeni kartezjańskiej z dokładnością do 0,025 stopnia kątowego. Pomiary koordynat elektrod poczynione na głowie osoby badanej mogą być zaimportowane do oprogramowania analizującego sygnał EEG [9].

Rys. 13. Digitizer firmy Polhemus umożliwiający pomiar rzeczywistego położenia przestrzennego elektrod do EEG



Rys. 14. Oprogramowanie do digitalizacji elektrod EEG

9. Wizualizacja wyników

W przypadku techniki EEG dużych rozdzielczości zarejestrowanych kilkadziesiąt-kilkaset kanałów sygnału EEG nie jest poddawanych analogowej ocenie wzrokowej, lecz jest poddawanych dalszej analizie cyfrowej i przetworzeniu pod

kątem stworzenia map napięciowych lub częstotliwościowych (*mapping*).

W starszych systemach czynność EEG była prezentowana w płaskich dwuwymiarowych mapach (projekcja głowy z góry lub – rzadziej – z boku).

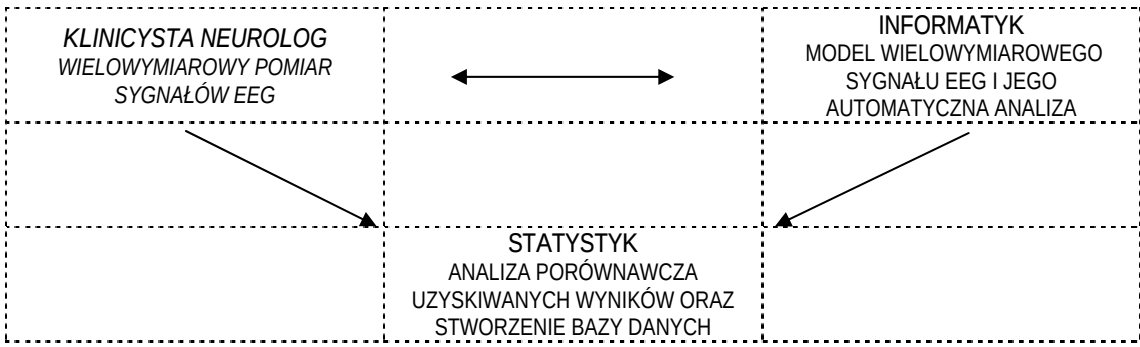
Nowsze oprogramowania pozwalają już na trójwymiarową, w pełni realistyczną prezentację sygnału EEG.

Trójwymiarowy model głowy może być obracany we wszystkich stopniach swobody, pozwalając na obserwację zmian czynności EEG poszczególnych obszarów głowy.

Kolejnymi krokami analizy może być wyszukiwanie korelacji między sygnałem EEG a różnymi wydarzeniami (*events*) mogącymi wpływać na czynność EEG), analiza falkowa, itp. [6].

Literatura cytowana

1. American Electroencephalographic Society: Guidelines for standard electrode position nomenclature. J. Clin. Neurophysiol., 8: 200-202, 1991.
2. Jasper H. H.: The ten-twenty electrode system of the International Federation, EEG. Clin. Neurophysiol., 10: 371-375, 1958.
3. Klem G. H., Lüders H. O., Jasper H. H., Elger C.: The ten-twenty electrode system of the International Federation, EEG. Clin. Neurophysiol., Supp. 52: 3-6, 1999.
4. Niebeling H. G.: Einführung in die Elektroencephalograpie, Johann Ambrosius Barth, Leipzig 1980.
5. Chatrian G. E ., Lettich E., Nelson P. L.: Ten percent electrode system for topographic studies of spontaneous and evoked EEG activity, Am J. EEG Technol., 25: 83-92, 1985.
6. Nuwer M. R., Lehmann D., Lopes de Silva F.: IFCN guidelines for topographic and frequency analysis of EEGs and EPs, EEG. Clin. Neurophysiol., Suppl. 52: 15-20, 1999.
7. Oostenveld R.: www.smi.auc.dk/~roberto/electrode.htm
8. Oostenveld R., Praamstra P.: The five percent electrode system for high-resolution EEG and ERP measurements, Clin. Neurophysiol. 112: 713-719, 2001.
9. 3Space Fastrak, Polhemus: www.antsoftware.nl/eetrak/fastrak.htm



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 333-338.

ZASTOSOWANIE VIDEODERMATOSKOPII W DIAGNOSTYCE ZMIAN BARWNIKOWYCH SKÓRY

VIDEODERMATOSCOPY IN THE DIAGNOSIS OF MELANOCYTIC SKIN LESIONS

EWA ŻABIŃSKA-PŁAZAK[*], ANNA WOJAS-PELC[*], GRZEGORZ DYDUCH[**]

[*]*Katedra Dermatologii Collegium Medicum Uniwersytetu Jagiellońskiego, Kraków, zabinskae@wp.pl*
[**]*Katedra Patomorfologii Collegium Medicum Uniwersytetu Jagiellońskiego, Kraków*

Streszczenie. Czerniak skóry należy do najbardziej złośliwych nowotworów człowieka. Obserwowany jest również na całym świecie bardzo dynamicznie narastający wzrost zachorowalności na czerniaka. Wczesna diagnostyka zmian jest warunkiem zastosowania skutecznego leczenia. Videodermatoskopia (mikroskopia epiluminescencyjna) jest metodą bardzo pomocną we wczesnej diagnostyce czerniaka skóry oraz monitorowaniu znamion atypowych. Jej dalszy rozwój jest ściśle uzależniony od rozwoju systemów komputerowych umożliwiających precyzyjną ocenę obrazów uzyskanych w badaniu dermatoskopowym.

Słowa kluczowe: videodermatoskopia, epiluminescencja mikroskopowa, czerniak skóry, znamiona atypowe

Abstract. Melanoma is one of the most malignant human neoplasmas. The dynamic increase of incidence of melanoma is observed all over the world. The early diagnose of melanoma is the condition of succesful treatment. Videodermatoscopy (epiluminescence microscopy) is a very useful method in early diagnose of melanoma and monitoring of atypical nevi. The development of videodermatoscopy is strictly conected with improving computer systems used in evaluation of skin lesions.

Key words: videodermatoscopy, epiluminescence microscopy, melanoma, atypical nevi

Czerniak skóry należy do nowotworów wykazujących najbardziej dynamiczny wzrost zachorowalności. Zjawisko to obserwowane jest we wszystkich rejonach świata, niezależnie od znacznych różnic w zapadalności na czerniaka, sięgającej od 0.8/100 000 ludności w Hong-Kongu do 30/100 000 w stanie Queensland w Australii [14].

Czerniak jest jednym z najbardziej złośliwych nowotworów, dającym wcześnie przerzuty drogą naczyń krwionośnych i chłonnych. Ponadto metody uzupełniające leczenie chirurgiczne czerniaka, takie jak chemioterapia, radioterapia, immunoterapia wykazują bardzo niską skuteczność. Na obecnym etapie wiedzy jedyną skuteczną metodą leczenia czerniaka pozostaje usunięcie chirurgiczne zmiany we wczesnym etapie zaawansowania nowotworu. Usunięcie zmiany o grubości poniżej 0.75 mm w skali Breslowa daje 10-letnie przeżycie w 97% przypadków, podczas gdy przy grubości usuniętej zmiany powyżej 4.0 mm spada ono do 32% [23]. W świetle powyższych danych niezwykle istotnym zagadnieniem staje się profilaktyka oraz wczesna diagnostyka czerniaka skóry.

Metodą pomocną w różnicowaniu zmian barwnikowych oraz diagnozowaniu czerniaka skóry we wczesnych etapach jego rozwoju jest dermatoskopia [3], [4], [24]. Jest ona popularna w krajach Europy i w Australii, podczas gdy w Stanach Zjednoczonych jest niewystarczająco rozpowszechniona [9]. Polega ona na oglądaniu powierzchni zmiany skórnej w powiększeniu, z wykorzystaniem światła oraz immersji. Początki dermatoskopii sięgają XVIII wieku, kiedy to w 1663 roku Johan Christophorus Kolhaus zastosował po raz pierwszy mikroskop do oceny naczyń powierzchownych wału paznokciowego. W 1878 roku Unna uzyskał znaczną poprawę jakości obrazu dzięki wykorzystaniu olejku immersyjnego. Początki współczesnej dermatoskopii sięgają lat dwudziestych XX wieku, kiedy to Saphier zastosował tę metodę diagnostyczną do oceny szeregu zmian skórnych, w tym barwnikowych [22].

Znaczący rozwój dermatoskopii nastąpił jednak dopiero w momencie wykorzystania do oceny obrazu powierzchni badanych zmian programów komputerowych. Umożliwiło to uzyskiwanie większych powiększeń, dodatkowej obróbki obrazu oraz – co jest szczególnie istotne – porównywanie

obrazu danej zmiany w odstępach czasu. Odzwierciedleniem popularyzacji videodermatoskopii jest wzrost liczby publikacji na jej temat, które pojawiły się w indeksowanych czasopismach w 16 w latach 1987-1993, do 76 w latach 1998-2000 [7].

Uważa się, że dermatoskopia wykazuje 90-95% czułość wykrywania czerniaka skóry, podczas gdy w badaniu klinicznym wynosi ona 70-80% [7].

Obecnie wykorzystywanych jest na świecie kilkanaście systemów videodermatoskopii. Do najszerzej stosowanych należą Dermogenius oraz Molmax II [22].

Istnieje również kilka algorytmów oceny obrazu uzyskanego w badaniu dermatoskopowym. Należą do nich: zasada ABCD, skala Menzies'a, skala 7-punktowa, zasada ABCDE, ABC [1], [5], [8], [11]. Do najczęściej cytowanych w literaturze należą skala Menzies'a (tab. 1), skala 7-punktowa (tab. 2) oraz zasada ABCD (tab. 3).

Tab. 1. Skala Menzies'a

zmiana łagodna	symetria układu zmiany i koloru
czerniak	asymetria układu zmiany więcej niż 1 kolor minimum 1 z 9 cech: niebiesko-biaława „woalka" liczne brązowe kropki pseudowypustki smugi promieniste odbarwienia bliznopodobne obwodowe czarne kropki/globule liczne kolory liczne niebieskie lub szare plamy szeroka siatka barwnikowa

Tab. 2. Skala 7-punktowa

kryteria	punktacja
kryteria duże	
atypowa siatka barwnikowa	2
atypowy wzór naczyniowy	2
niebiesko-biaława „woalka"	2
kryteria małe	
nieregularne smugi (pseudowypustki, smugi promieniste)	1
nieregularne zabarwienie	1
nieregularne plamy/globule	1
obszary regresji	1
wynik	< 3 = zmiana nie jest czerniakiem ≥ 3 = czerniak

Tab. 3. Zasada ABCD

zróżnicowanie zmian melanocytowych i niemelanocytowych	
regularna siatka barwnikowa smugi gałązkowate ciałka skupione	zmiana melanocytowa
obszary stalowo-niebieskie	znamię błękitne
pseudocysty rogowe otwory pseudomieszkowe	brodawka łojotokowa
zatoki naczyniowe	naczyniak
struktury podobne do liścia klonowego	rak
zmiany nie wykazujące powyższych cech	zmiana melanocytowa

ocena ABCD	
oceniana cecha	TDS (total dermatoscopy score)
Asymmetry	x 1.3
Border (granica)	x 0.1
Color (kolor)	x 0.5
Different structural components (zróżnicowane składniki strukturalne)	x 0.5
wynik – rodzaj zmiany	< 4.75 = łagodna 4.8-5.45 = budząca podejrzenie czerniaka > 5.45 prawdopodobnie czerniak

Badania porównujące 3 powyższe schematy oceny zmian barwnikowych wykazały największą czułość skali 7-punktowej i największą specyficzność zasady ABCD w stosunku do pozostałych [2]. Podkreślany jest również fakt, że skala 7-punktowa i skala Menzies'a wydają się być prostsze w zastosowaniu dla mniej doświadczonych badaczy [3].

Rewolucją w dziedzinie diagnostyki zmian skórnych stała się możliwość przesyłania cyfrowych obrazów badania dermatoskopowego poprzez sieci telekomunikacyjne [18].

Telemedycyna stanowi obecnie istotny element codziennej praktyki w wielu dziedzinach, szczególnie radiologii i patologii. W wielu krajach również dermatologia stała się specjalnością wykorzystującą Internet. Dermatoskopia jest badaniem szczególnie „telemedycznym". Wpływają na to następujące fakty: obraz zapisywany jest techniką cyfrową, a więc jest łatwy do przesłania siecią, jego analiza zależy w głównym stopniu od doświadczenia badacza, dermatolog w wielu regionach jest specjalistą trudno dostępnym, jednym z najistotniejszych elementów leczenia czerniaka jest szybka diagnoza zmian podejrzanych.

Analiza diagnozy stawianej na podstawie badania bezpośredniego zmiany oraz poprzez ocenę obrazu uzyskanego przez Internet wykazała porównywalną trafność obu metod [19], [20]. Stwierdzono, że doświadczenie osoby oceniającej obraz oraz stopień trudności diagnostycznej danej zmiany wpływały na ostateczny wynik badania bardziej znacząco niż jakość przesłanego obrazu [20].

W niektórych krajach Europy od kilku lat tworzone są centra diagnostyki zmian barwnikowych ze względu na szczególnie istotny czynnik – konieczne jest posiadanie dużego doświadczenia klinicznego oceniającego zmiany. Szersze zastosowanie teledermatoskopii pozwoli w przyszłości zwiększyć dostępność diagnozy w takich wysoce wyspecjalizowanych ośrodkach.

Odrębnym zagadnieniem są próby zastosowania tzw. automatycznej dermatoskopii, wykorzystującej programy komputerowe do oceny zmian barwnikowych. Zainteresowanie możliwością zastosowania automatycznej komputerowej analizy znacznie wzrosło w ostatnich latach ze względu na narastającą zachorowalność na czerniaka skóry oraz niską trafność diagnozy stawianej przez lekarzy nie będących specjalistami. Bardzo istotny jest również fakt rozwoju systemów komputerowych mogących sprostać skomplikowanym algorytmom diagnozowania zmian skórnych [17]. Konsorcjum DANAOS (Diagnostic and Neuronal Analysis of Skin Cancer), mieszczące się w Zentrum fur Neuroiformatik w Bochum w Niemczech, grupuje obecnie 13 europejskich ośrodków dysponujących systemem kalibrowanej 3-CCD (*charge-coupled device*) videodermatoskopii [16]. Prowa-

dzone są tam badania mające na celu udoskonalenie systemów autodermatoskopii tak, aby mogły one znaleźć zastosowanie w codziennej praktyce lekarzy podstawowej opieki. Jednym ze stosowanych modeli komputerowej analizy zmian barwnikowych jest system stworzony w przez Burroni-Dell'Eva Biomedical Engineering Group (Siena, Włochy), wykorzystujący kalibrowaną 3 CCD videokamerę. Badanie dotyczące analizy 57 przypadków czerniaka i 90 przypadków znamion barwnikowych wykazało 80.7% czułość i 87.8% swoistość tej metody [1]. Najistotniejszym ograniczeniem autodermatoskopii jest niedoskonałość programów w różnicowaniu zmian melanocytowych i niemelanocytowych, co powoduje, że łagodne zmiany barwnikowe, takie jak dermatofibroma czy brodawki łojotokowe uzyskują wysoki wynik TDS i kwalifikowanie są do zabiegu chirurgicznego [16]. W przypadku wykorzystania dermatoskopii przez doświadczoną osobę ta niedogodność nie występuje, a wręcz w niektórych szczególnie trudnych przypadkach dermatoskopia umożliwia różnicowanie wyżej wymienionych zmian. Do takich przypadków należą także zmiany zlokalizowane pod paznokciem, podrażnione, mocno pigmentowane brodawki łojotokowe, niektóre przypadki raka podstawnokomórkowego barwnikowego [6].

Jednym z najistotniejszych elementów badania dermatoskopowego jest możliwość obserwacji zmian o typie znamienia atypowego, zwłaszcza u chorych z grupy ryzyka rozwoju czerniaka, tj. z dodatnim osobniczym lub rodzinnym wywiadem w kierunku czerniaka, osób z zespołem znamion atypowych [1], [10]. Ze względu na fakt, że jedynie 30-40% przypadków czerniaka skóry rozwija się na podłożu wcześniej istniejącego znamienia barwnikowego, istnieje obecnie tendencja do ograniczania ilości usuwanych profilaktycznie znamion atypowych. Nie jest możliwe usunięcie wszystkich znamion u danej osoby, o wiele istotniejsze wydaje się dokładne monitorowanie zmian i usuwanie tych, które ulegają jakiejkolwiek zmianie, co może świadczyć o ich transformacji nowotworowej. W badaniu dotyczącym 245 chorych, u których oceniono 318 znamion atypowych w odstępie od 2.5 do 4.5 miesięcy, pośród znamion, w obrębie których stwierdzono jakiekolwiek zmiany w badaniu dermatoskopowym (w porównaniu z badaniem wcześniejszym), 11% okazało się być czerniakiem [15]. Szczególnie istotne jest, że żadna ze zmian zdiagnozowanych w badaniu histopatologicznym jako czerniak nie wykazywała cech typowych w badaniu dermatoskopowym. Zmiany nadal spełniały kryterium znamienia atypowego, a jedynie obserwowane w ich obrębie zmiany (koloru, struktur, rozmiaru) decydowały o zakwalifikowaniu do usunięcia chirurgicznego. Ponadto wszystkie zmiany rozpoznane jako czerniak wykazywały niski stopień zaawansowania w skali Breslowa (maksymalnie 0.28 mm). Wydaje się, że możliwość monitorowania zmian stanowi jeden z istotniejszych kierunków rozwoju dermatoskopii [11], [13], [21], a szczególnie duże nadzieje pokładane są w autodermatoskopii, która mogłaby zwiększyć precyzję tego typu procedury.

Poniżej przedstawiamy przypadki kliniczne mogące sprawiać trudności w rutynowym badaniu, w diagnostyce których wykorzystano videodermatoskopię.

Badania dermatoskopowe wykonane zostały przy użyciu standaryzowanego systemu komputerowego MoleMax II (Derma Instruments, Wiedeń, Austria) w Katedrze Dermatologii CMUJ. Zmiany oceniane były w 30-krotnym powiększeniu.

Badania histopatologiczne wykonano w Katedrze Patomorfologii CMUJ.

Przypadek nr 1

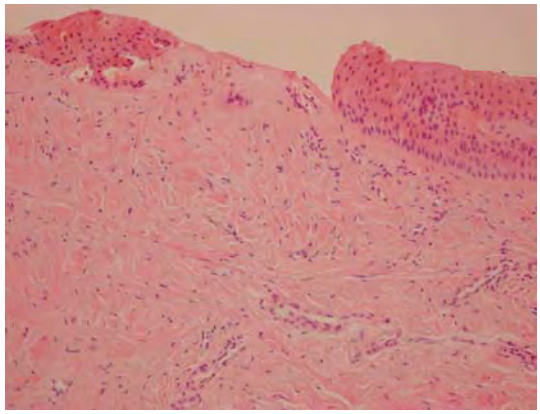
33-letnia chora zgłosiła się z powodu zmiany barwnikowej, która pojawiła się pod paznokciem trzeciego palca prawej dłoni. Zmiana powiększała się w ciągu kilku miesięcy. Starsza siostra chorej operowana była z powodu czerniaka skóry 3 lata wcześniej, obecnie bez nawrotu choroby. W badaniu przedmiotowym zwraca uwagę występowanie przebarwienia również w obrębie wału paznokciowego (fot. 1). Zjawisko to, znane jako objaw Hutchinsona, jest często obserwowane w przypadkach czerniaka podpaznokciowego.

Fot.1.

Natomiast w wykonanym badaniu dermatoskopowym stwierdzono cechy typowe dla łagodnego znamienia barwnikowego: regularne, równoległe linie, wyraźne boczne granice (fot. 2).

Fot 2.

Ze względu na dodatni wywiad rodzinny zmianę usunięto w całości. Badanie histopatologiczne potwierdziło łagodny charakter zmiany (fot. 3).



Fot. 3.

Komentarz. Zmiany zlokalizowane pod paznokciem są trudno dostępne badaniu klinicznemu. Czerniak w tej lokalizacji charakteryzuje się wysoką dynamiką rozwoju. Badanie dermatoskopowe pozwala często uniknąć zabiegu zdejmowania płytki paznokciowej w przypadku tej niepokojącej dla dermatologa lokalizacji zmiany barwnikowej.

Dla porównania – przypadek krwiaka podpaznokciowego u 20-letniej chorej (fot. 4).

Fot. 4.

Przypadek nr 2

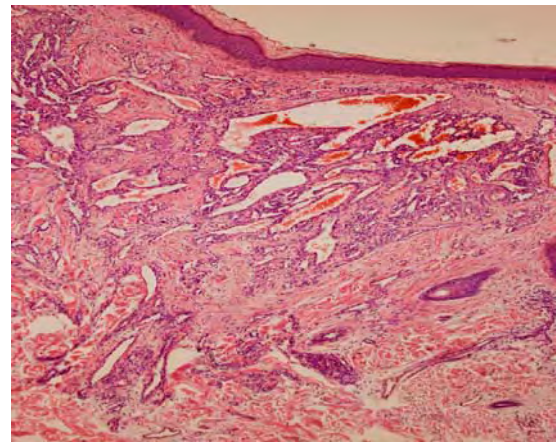
Chory lat 35, zgłosił się do Kliniki zaniepokojony szybkim wzrostem zmiany „barwnikowej" zlokalizowanej na plecach, którą obserwował od kilku lat. W wywiadzie podał intensywne nasłonecznienie 3 miesiące wcześniej. W badaniu klinicznym zmiana wykazywała cechy zmiany naczyniowej (żywoczerwony kolor, zblednięcie przy ucisku) (fot. 5).

Fot. 5.

W badaniu dermatoskopowym nie stwierdzono cech typowych dla naczyniaka (typowe zatoki), brak jednak również cech zmiany melanocytowej (fot. 6).

Fot. 6.

Ze względu na wywiad oraz znaczny niepokój chorego zdecydowano o usunięciu zmiany w całości. Badanie histopatologiczne potwierdziło rozpoznanie kliniczne naczyniaka (fot. 7).

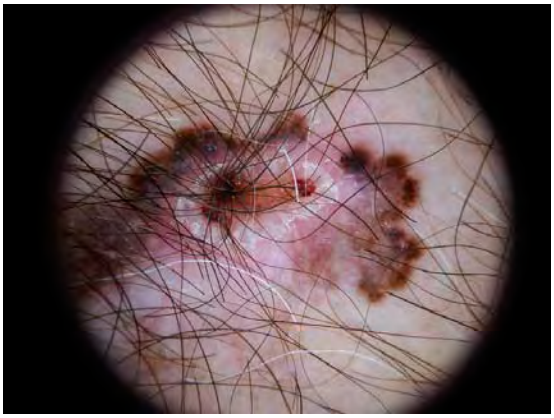


Fot. 7.

Komentarz. Należy pamiętać, że diagnoza wstępna zmiany barwnikowej zawsze powinna opierać się na wywiadzie, badaniu przedmiotowym i – jeśli to możliwe – badaniu dermatoskopowym. Rozstrzygającym pozostaje badanie histopatologiczne.

Przypadek nr 3

Do Kliniki skierowano z Przychodni Rejonowej pacjenta, u którego ze zmiany zlokalizowanej na skórze owłosionej głowy pobrano biopsję do badania histopatologicznego. W wykonanym badaniu stwierdzono czerniaka skóry. Przed usunięciem zmiany wykonaliśmy badanie dermatoskopowe, którego obraz wykazał cechy typowe dla czerniaka: ostre ograniczenie zmiany, asymetria zmiany, obszar regresji w centrum, atypowe globule (ciałka skupione) na obwodzie, TDS: 7.2 (fot. 8).

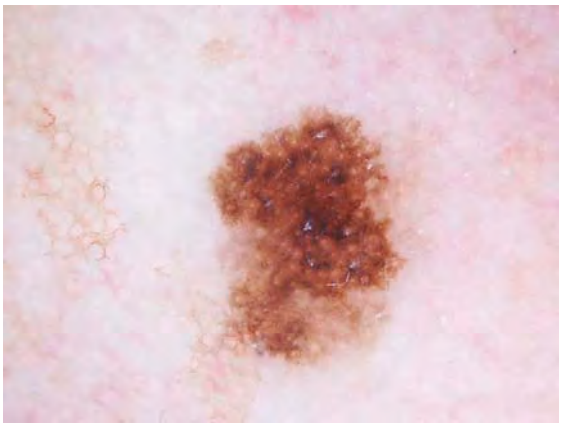


Fot. 8.

Komentarz. Należy podkreślić, że zalecanym postępowaniem w przypadku zmiany budzącej podejrzenie czerniaka jest usuwanie jej w całości z małym marginesem (tzw. *excision biopsy*) i ewentualne jego poszerzanie w zależności od stopnia zaawansowania czerniaka w skali Breslowa. Zastosowanie dermatoskopii pozwala zmniejszyć liczbę niefortunnie pobranych biopsji.

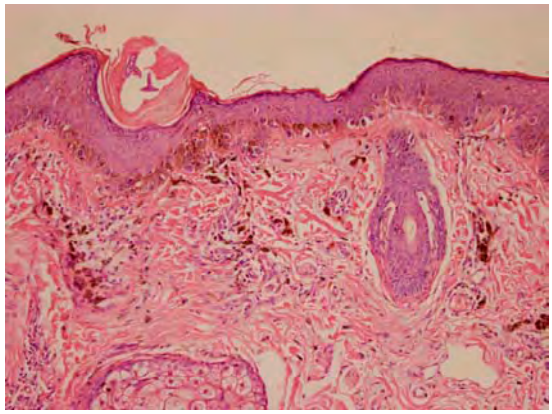
Przypadek nr 4

Chory lat 67, 2 lata wcześniej operowany w Klinice z powodu raka podstawnokomórkowego. Wszyscy chorzy operowani w Klinice z powodu nowotworów skóry zgłaszają się do okresowej kontroli. W trakcie kolejnego rutynowego badania kontrolnego lekarz zauważył niewielką, 5 mm zmianę barwnikową, silnie wysyconą, zlokalizowaną w okolicy skroniowej prawej. Chory nie pamiętał, kiedy zmiana pojawiła się. W wykonanym badaniu dermatoskopowym otrzymano wynik TDS: 6.1 (fot. 9). Zmianę usunięto w całości. W badaniu histopatologicznym stwierdzono *lentigo maligna* (fot. 10).



Fot. 9.

Komentarz. Wydaje się, że okresowe kontrole u doświadczonego lekarza mogą zwiększyć liczbę przypadków czerniaka zdiagnozowanych we wczesnym okresie rozwoju, kiedy usunięcie zmiany w całości daje szansę całkowitego wyleczenia.



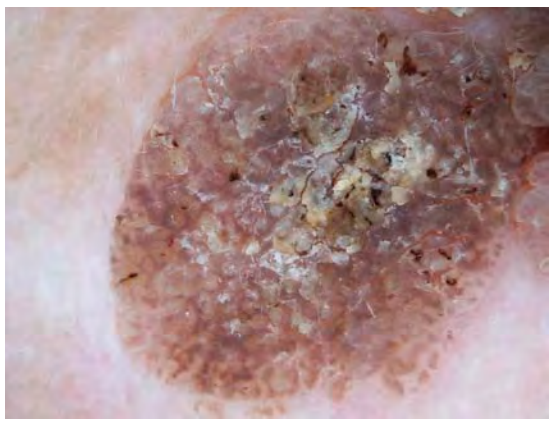
Fot. 10.

Przypadek nr 5

Chora lat 67, zgłosiła się do badania z powodu zmiany barwnikowej na czole, stopniowo powiększającej się od kilku lat (fot. 11).

W badaniu dermatoskopowym stwierdzono cechy typowe dla brodawki łojotokowej: otwory pseudomieszkowe, pseudocysty rogowe, obszary „wygryzione przez mole", ostre granice zmiany (fot. 12). Ze względu na bardzo typowy obraz dermatoskopowy zmiany nie usunięto.

Fot. 11.



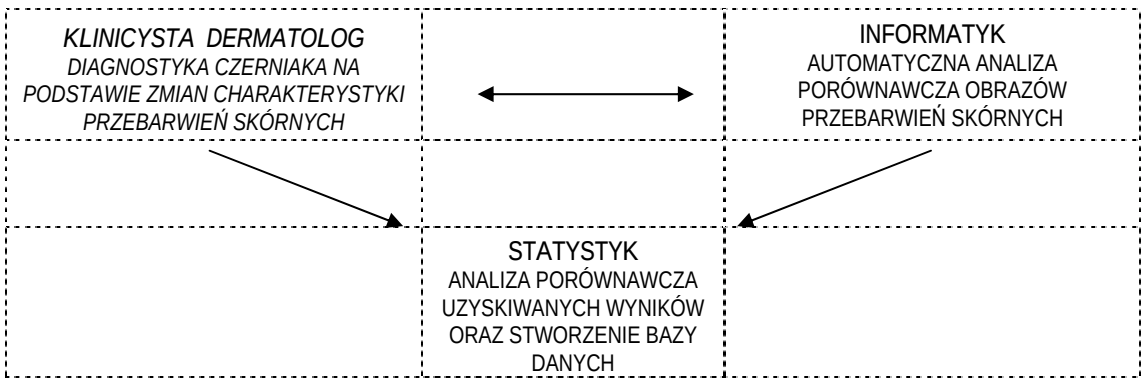
Fot. 12.

Komentarz. Nietypowe brodawki łojotokowe mogą budzić podejrzenie czerniaka, najczęściej typu czerniaka wywodzącego się z plamy soczewicowatej lub typu guzkowego.

Podsumowując, stały wzrost zachorowalności na czerniaka, jednego z bardziej złośliwych nowotworów człowieka, zmusza do udoskonalania metod jego diagnozowania. Ogromny rozwój systemów komputerowych wspomagających medycynę daje nadzieję na szersze zastosowanie biocybernetyki w diagnostyce zmian barwnikowych skóry.

Literatura cytowana

1. Andreassi L., Perotti R., Rubegni P. et al.: Digital Dermatoscopy analysis for the differentiation of atypical nevi and early melanoma: a new quantitative semiology, Arch Dermatol, 135: 1459-1465, 1999.
2. Argenziano G., Fabbrocini G., Carli P. et al.: Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: comparison of the A BCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis, Arch Dermatol, 134: 1563-70, 1998.
3. Argenziano G., Soyer H. P., Chimenti S. et al.: Impact of dermatoscopy on the clinical management of pigmented skin lesions, Clin Dermatol, 20: 200-202, 2002.
4. Argenziano G., Soyer H.P.: Dermatoscopy of pigmented lesions a valuable tool for early diagnosis of melanoma, Lancet Oncol 2: 443-449, 2001.
5. Blum A., Rassner G., Garbe C.: Modified ABC-point list of dermatoscopy: a simplified and highly accurate dermatoscopic al.gorithm for the diagnosis of cutaneous melanocytic lesions, J Am Acad Dermatol, 48: 672-678, 2003.
6. Braun R. P., Rabinovitz H., Oliviero M. et al.: Dermoscopic diagnosis of seborrheic keratosis, Clin Drmatol, 20: 270-272, 2002.
7. Carli P., De Giorgi V., Gianotti B. et al.: Dermatoscopy and early diagnosis of melanoma, Arch Dermatol, 137: 1641-1644, 2001.
8. Carli P., De Giorgi V., Palli D. et al.: Preoperative assessment of melanoma thickness by ABCD score of dermatoscopy, J Am Acad Dermatol, 43: 459-466, 2000.
9. Federman D. G., Kravetz J. D., Kirsner R. S.: Skin cancer screening by drmatologist: prevalence and barriers, J Am Acad Dermatol, 46: 710-714, 2002.
10. Hofmann-Wellenhof R., Blum A., Wolf I. H. et al.: Dermatoscopic classification of atypical melanocytic nevi (Clark nevi), Arch Dermatol, 137: 1575-1580, 2001.
11. Johr R., Menzies S. W.: Lessons on Dermatoscopy. Dermatol Surg 27: 911-912, 2001.
12. Johr R H.: Dermatoscopy: al.ternative melanocytic al.gorithms-the ABCD rule of dermatoscopy, Menzies scoring method, and 7-point checklist. Clinics in Dermatology 20: 240-247, 2002.
13. Malvehy J., Susana P.: Dermatoscopic patterns of binign volar melanocytic lesions in patients with atypical mole syndrom, Arch Dermatol, 140: 538-544, 2004.
14. Marks R.: Epidemiology of melanoma, Clin Exp Dermatol, 25: 459-463, 2000.
15. Menzies S. W., Gutenev A., Avramidis M. et al.: Short-term digital surface microscopic monitoring of atypical or changing melanocytic lesion, Arch Dermatol, 137: 1583-1589, 2001.
16. Menzies S. W.: Automated epiluminescence microscopy: human VS machine in the diagnosis of melanoma, Arch Dermatol, 135: 1538-1540, 1999.
17. Paine S., Cockburn J., Noy S. et al.: Early detection of skin cancer: knowledge, perceptions and practices of general practitioners in Victoria, Med J Aust, 161: 188-195, 1994.
18. Perednia D. A., Brown N. A.: Teledermatolody: one application of telemedicine, Bull Med Libr Assoc, 83: 42-47, 1995.
19. Piccolo D., Smolle J., Argenziano G. et al.: Teledermatoscopy-results of a multicentre study on 43 pigmented skin lesions, J Telemed Telecare, 6: 132-137, 2000.
20. Piccolo D., Smolle J., Wolf I.H. et al.: Face-to-face diagnosis vs telediagnosis of pigmented skin tumors: a teledermatoscopic study, Arch Dermatol, 135: 1467-1471, 1999.
21. Robinson J. K., Nickoloff B. J.: Digital epiluminescence microscopy monitoring of high-risk patients, Arch Dermatol, 140: 49-56, 2004.
22. Stolz W., Braun-Falco O., Bielk P. et al.: Color atlas of dermatoscopy. Blackwell Publishing Company, 2nd edition, 2002.
23. Thörn M., Pontén F., Bergström R. et al.: Clinical and histopathologic predictors of survival in patients with malignant melanoma: a population-based study in Sweden, J Natl Cancer Inst, 86: 761-769, 1994.
24. Westerhoff K., McCarthy W. H., Menzies S. W.: Increase in the sensitivity for melanoma diagnosis by primary care physicians using skin surface microscopy, Br J Dermatol, 143: 1016-1020, 2000.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 339-347.

SZACOWANIE ISTOTNOŚCI STATYSTYCZNEJ ZMIAN W MOCY SYGNAŁU EEG ZWIĄZANEGO Z BODŹCEM

ESTIMATION OF STATISTICAL SIGNIFICANCE OF EVENT RELATED EEG POWER CHANGES

JAROSŁAW ŻYGIEREWICZ*, PIOTR DURKA*, JOANNA MAZURKIEWICZ*, NATHAN CRONE**, PIOTR FRANASZCZUK**

**Zakład Fizyki Biomedycznej, Instytut Fizyki Doświadczalnej, Uniwersytet Warszawski, Warszawa*
jarekz@fuw.edu.pl, durka@fuw.edu.pl, asia@mazurkiewicz.org
***Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, USA*
necrone@jhmi.ed, pfranasz@jhmi.edu

Streszczenie. Interesujących danych na temat przetwarzania informacji przez mózg dostarczają badania zmian synchronizacji aktywności komórek piramidowych kory mózgowej związanych z reakcjami na proste zdarzenia. Zmiany synchronizacji przejawiają się jako zmiany mocy sygnału EEG lub ECoG. Szacowanie istotności statystycznej tych zmian mocy stwarza kilka praktycznych problemów. W artykule zaproponowano kompleksowe ich rozwiązanie. Składa się ono z następujących kroków: oszacowania gęstości energii sygnału EEG w dziedzinie czas-częstość, wyznaczenia średnich zmian mocy, oszacowania ich istotności statystycznej wraz z poprawkami na wykonywanie testów wielokrotnych oraz weryfikowania hipotez dotyczących współzależności zmian zachodzących w różnych pasmach częstości.

Słowa kluczowe: ERD, ERS, mapy statystyczne, czasowo-częstotliwościowe rozkłady gęstości energii sygnału, korelacje krótkoczasowe

Abstract. Event related changes in the synchronization of pyramidal neurons activity are a valuable source of information about the information processing in the brain. These changes can be measured as the changes in the power of EEG or ECoG signal. This paper presents a methodology of estimating the statistical significance of those changes. It consists of the following steps: estimation of signal energy density in the time-frequency space, computation of mean relative changes of power and assessment of their significance with the respect of multiple comparison problem, and testing the hypothesis concerning the possible covariance of different band-power changes. Usefulness of this methodology is presented on ECoG data.

Key words: ERD, ERS, statistical maps, time-frequency energy density, short-time covariance

1. Wstęp

Funkcjonalna aktywność mózgu jest związana z wieloma zmianami przejawiającymi się między innymi jako wzrosty i spadki mocy sygnału EEG lub ECoG. Zmiany te przebiegają różnie w różnych pasmach częstości. Na przykład [23] przy świadomie kontrolowanym ruchu palcem moc sygnału EEG w paśmie alfa (8-12Hz) spada na około 2 sekundy przed początkiem ruchu i powraca do wartości spoczynkowej dopiero po około 2 sekundach po zakończeniu wykonywania ruchu. W paśmie beta (13-25 Hz) zanik mocy występuje w krótszym sąsiedztwie czasowym ruchu, ale w około sekundę po zakończeniu ruch, moc sygnału wzrasta powyżej poziomu spoczynkowego. Efekt ten znany jest jako *beta-rebound* [23]. W paśmie gamma (>30 Hz) tuż przed rozpoczęciem ruchu obserwuje się silny wzrost mocy. Spadki i wzrosty mocy sygnału w różnych pasmach częstości związane z reakcją na określone zdarzenie można wyrazić jako wielkości względne, mierzone w odniesieniu do pewnego czasu referencyjnego: ERD – Event Related Desynchronization oraz ERS – Event Related Synchronization [21]. ERD i ERS w różnych pasmach częstości wiązane są z różnymi procesami zachodzącymi na poziomie neuronalnym i przypisuje im się różne znaczenia fizjologiczne. Wysoka moc sygnałów w pasmach alfa i beta jest wiązana ze stanem pobudzenia, któremu nie towarzyszy przetwarzanie informacji. Efekt *beta-rebound'u* uważany jest za przejaw aktywnego hamowania obwodów neuronalnych, które ten sygnał wytwarzają [24], [25].

Przeciwne znaczenie jest przypisywane wzrostowi mocy sygnału w paśmie gamma. Wielu badaczy uważa, że wysoka moc w paśmie gamma jest przejawem funkcjonalnej aktywno-

ści obszaru kory, z którego ów sygnał pochodzi [16], [22], [26]. Szczególnie wyraźnie związek ten jest widoczny w danych ECoG [4], [6].

Granice pasm, w których zachodzą interesujące zmiany związane z przetwarzaniem zdarzeń, wykazują dużą zmienność pomiędzy badanymi osobami. Z tego względu pożądane są metody pozwalające na badanie zmian w rozkładach gęstości energii w całej przestrzeni czas-częstość jednocześnie. Podejmowano już próby sformułowania odpowiednich metod opartych na pasmowym filtrowaniu w zachodzących na siebie pasmach częstości [13], spektrogramie [18], skalogramie [27], wygładzanej transformacji Wigner-Villa [17] czy też na dopasowaniu kroczącym (MP) [10].

We wszystkich tych metodach występuje wspólny problem oszacowania istotności statystycznej wykrytych zmian gęstości energii sygnału. Problem ten jest trudny z dwóch względów. Po pierwsze nie ma gwarancji, że analizowane dane pochodzą z rozkładu normalnego. Założenie o normalności rozkładu danych jest kluczowe dla większości parametrycznych metod statystycznych. Po drugie zastosowana analiza statystyczna musi uwzględniać fakt wykonywania testów w wielu punktach przestrzeni czas-częstość. Wymusza to zastosowanie procedur, które nie tracą zbyt szybko mocy wraz ze wzrostem ilości wykonywanych testów.

Rozwiązanie powyższego problemu zaproponowano w [7]. Bazuje ono na testach repróbkowanych – nie wymagających założeń co do rozkładu prawdopodobieństwa danych wejściowych – oraz na uwzględnieniu efektu wielokrotności testów poprzez kontrolę FDR (False Discovery Rate [1]). Ta metoda jest punktem odniesienia dla mniej kosztownych obliczeniowo wariantów opartych na normalizowaniu danych wejściowych i stosowaniu testów parametrycznych wraz z korektą FDR na wielokrotność testów [30].

Kolejny problem pojawiający się przy analizie map ERD/ERS to stwierdzenie, czy obserwowane w różnych pasmach częstości zmiany mocy, widoczne na mapie jako prawie jednoczesne, są ze sobą skorelowane, czy też występują od siebie niezależnie. Do rozwiązania tego problemu zaproponowano metodę średnich korelacji krótkoczasowych.

2. Metoda

2.1. Estymacja gęstości energii w dziedzinie czas-częstość

Poniżej przypomniano krótko dwie metody estymowania rozkładu gęstości energii:

2.1.1. Spektrogram

Spektrogram dzieli sygnał $s(t)$ na zachodzące na siebie odcinki. Każdy z tych odcinków mnożony jest przez funkcję okienkującą $w(t)$, a następnie poddawany transformacie Fouriera. Krótkoczasowa transformata Fouriera (STFT) dana jest wzorem:

$$Fs(u,f)=\int_{-\infty}^{\infty} s(t)w^*(t-u)e^{-2\pi ift}dt \tag{1}$$

W prezentowanych badaniach użyto okna Hanninga $w(t)=\frac{1}{2}\left(1-\cos\left(2\pi\frac{t}{T_{len}}\right)\right)$, gdzie T_{len} jest długością okna. Długość okna jest parametrem, który trzeba ustalić w sposób arbitralny. Najbardziej wyrazisty obraz czasowo-częstotliwościowy otrzymuje się, kiedy długość okna odpowiada długości struktur obecnych w sygnale. Dla prezentowanych danych było to 0.2 s.

Kwadrat wartości bezwzględnej STFT nazywany jest spektrogramem:

$$E^{SP}=\left|Fs(u,f)\right|^2 \tag{2}$$

2.1.2. Dopasowanie Kroczące (Matching Pursuit)

Dopasowanie Kroczące (MP) jest algorytmem, który w sposób iteracyjny dekomponuje sygnał $s(t)$ na składowe g_{γ_n}. Pochodzą one z redundantnego zbioru funkcji zwanego słownikiem D. W każdym kroku ze słownika wybierany jest ten element, który najlepiej pasuje do aktualnego residuum. Procedurę tę można zapisać następująco:

$$\begin{cases} R^0 s &= s \\ R^n s &= \left\langle R^n s, g_{\gamma_n}\right\rangle g_{\gamma_n} + R^{n+1}s \\ g_{\gamma_n} &= \arg\max_{g_{\gamma_i}\in D}\left|\left\langle R^n s, g_{\gamma_i}\right\rangle\right| \end{cases}, \tag{3}$$

gdzie $\arg\max_{g_{\gamma_i}\in D}$ oznacza taki g_{γ_i}, który daje największą wartość iloczynu skalarnego $\left|\left\langle R^n s, g_{\gamma_i}\right\rangle\right|$. Słownik D do analizy sygnałów rzeczywistych jest skonstruowany z rzeczywistych funkcji Gabora:

$$g_\gamma(t)=K(\gamma)e^{-\pi\left(\frac{t-u}{\sigma}\right)^2}\sin\left(2\pi f(t-u)+\phi\right) \tag{4}$$

$K(\gamma)$ jest stała normującą, taką, że $\left\|g_\gamma\right\|=1$, zaś $\gamma=\{u,f,\sigma,\phi\}$ oznacza parametry funkcji w słowniku (u – translacja w czasie, f – częstość, σ – szerokość w czasie, ϕ – faza). Autorzy metody – Mallat i Zhang – wykazali, że procedura dekompozycji jest zbieżna [19]. W praktycznych zastosowaniach używane są skończone sumy, które przybliżają sygnał s w następujący sposób:

$$s\approx\sum_{n=0}^{M}\left\langle R^n s, g_{\gamma_n}\right\rangle g_{\gamma_n} \tag{5}$$

Z opisanej powyżej dekompozycji można uzyskać estymatę gęstości energii w dziedzinie czas-częstość poprzez sumowanie transformat Wigner-Ville'a wszystkich elementów rozwinięcia:

$$E^{MP}=\sum_{n=0}^{M}\left|\left\langle R^n s, g_{\gamma_n}\right\rangle\right|^2 Vg_{\gamma_n}(t,f) \tag{6}$$

gdzie:

$$Vg_{\gamma_n}(t,f)=\int_{-\infty}^{\infty} g_{\gamma_n}\left(t+\tfrac{\tau}{2}\right)g^*_{\gamma_n}\left(t-\tfrac{\tau}{2}\right)e^{-i2\pi f\tau} \tag{7}$$

2.2. Sformułowanie hipotezy zerowej

W typowych eksperymentach dotyczących ERD/ERS analizowanych jest N realizacji sygnału. Każda realizacja zawiera czas przed bodźcem, w którym wyróżniamy okres referencyjny t_{ref} (z założenia nie zaburzony przez bodziec) oraz okres tuż przed i po zaprezentowaniu bodźca t_b, w którym mogą wystąpić interesujące zmiany. Dla każdej realizacji osobno estymowany jest rozkład gęstości energii $E_n(t, f)$ $n = 1,..., N$. Dla potrzeb skonstruowania statystyki trzeba użyć map czasowo-częstościowych mocy sygnału o zredukowanej rozdzielczości. Przemawiają za tym dwa powody. Po pierwsze, rozkład gęstości energii obarczony jest nieoznaczonością $\Delta t \Delta f \geq 1/4\pi$. Po drugie różnego rodzaju szumy występujące w danych wprowadzają wariancję w rozkładach energii. Oba te efekty można zredukować przez wybór odpowiedniej rozdzielczości. Z wcześniejszych badań wynika, że optymalne zachowanie map statystycznych można osiągnąć dla $\Delta t \Delta f = 1/2$ [7]. Dalej element przestrzeni czas-częstość o rozmiarach $\Delta t \times \Delta f$ będzie nazywany reselem (*resolution element*). Tak więc mapa mocy sygnału dana jest wzorem:

$$P_n(i, j) = \int_{i\Delta t}^{(i+1)\Delta t} \int_{j\Delta f}^{(j+1)\Delta f} E_n(t, f)\, dt df \tag{8}$$

Rozkład gęstości prawdopodobieństwa występowania różnych wartości $P_n(i, j)$ jest nieznany. Możliwe są więc dwa wyjścia: zastosowanie testów nieparametrycznych np. bazujących na repróbkowaniu albo dokonanie transformacji danych sprowadzających je do rozkładu normalnego. Niech

$$X_n^{TR}(i, j) = TR(P_n(i, j)) \tag{9}$$

gdzie TR będzie albo identycznością ($TR = NT$ – nie transformowane) albo transformatą Box'a-Cox'a ($TR = BC$). Transformata Box'a-Cox'a [2] jest rodziną transformat potęgowych:

$$BC(x, \lambda) = \begin{cases} \dfrac{x^{\lambda} - 1}{\lambda} & dla \quad \lambda \neq 0 \\ \log(x) & dla \quad \lambda = 0 \end{cases} \tag{10}$$

Dla każdego pasma częstości j parametr λ jest optymalizowany [15] osobno poprzez maksymalizację funkcji loglikelihood na obszarze referencyjnym:

$$\lambda_{opt}^{j} = \max_{\lambda} \left\{ -\tfrac{m}{2} \log \sigma_{BC(x,\lambda)}^{2} + (\lambda - 1) \sum_{k=1}^{m} \log x_k \right\}, \tag{11}$$

gdzie m jest długością wektora danych x w rozważanym przypadku:

$$x_k \in \{P_n(i, j) : i \in t_{ref},\, n = 1,..., N\} \tag{12}$$

Aby znaleźć resele, w których zachodzi istotna zmiana energii, należy zweryfikować zbiór hipotez:

$$H_0^{i,j}: \quad \overline{X^{TR}(i, j)} = \overline{X_{ref}^{TR}(j)} \tag{13}$$

$$H_1^{i,j}: \quad \overline{X^{TR}(i, j)} \neq \overline{X_{ref}^{TR}(j)}, \tag{14}$$

gdzie $\overline{X^{TR}(i, j)}$ jest średnią wartością X^{TR} w reselu o współrzędnych (i, j) wziętą po realizacjach, zaś $\overline{X_{ref}^{TR}(j)}$ jest wartością średnią X^{TR} wziętą po reselach należących do obszaru referencyjnego t_{ref} i po realizacjach.

2.3. Statystyka

Zarówno w testach parametrycznych, jak i nieparametrycznych zastosowano statystykę t, gdyż w przypadku danych o rozkładzie normalnym jest ona kanoniczna, a dla innych rozkładów jest asymptotycznie kanoniczna [29]:

$$t(i, j) = \frac{\overline{X_{ref}^{TR}(j)} - \overline{X^{TR}(i, j)}}{s_{\Delta}}, \tag{15}$$

gdzie s_{Δ} jest wariancją łączną obszaru referencyjnego i badanego resela.

2.3.1. Repróbkowanie

Jeśli nie można założyć, że dane $X^{TR}(i, j)$ pochodzą z rozkładu normalnego, to estymowany jest rozkład statystyki (15) poprzez repróbkowanie danych z obszaru referencyjnego dla każdej częstości j osobno:

1) ze zbioru $X^{TR}(i, j)$, $i \in t_{ref}$ losowane są z powtórzeniami dwie próbki: A o rozmiarze N i B o rozmiarze $N \cdot N_{ref}$

2) obliczana jest wartość t ze wzoru (15): $t_{resamp}(j) = \dfrac{\overline{X_A^{TR}} - \overline{X_B^{TR}}}{s_{AB}}$

Kroki 1 i 2 powtarzane są N_{rep} razy. Zbiór wartości $t_{resamp}(j)$ przybliża rozkład gęstości prawdopodobieństwa $T_{resamp}(j)$ dla częstości j. Następnie wartość statystyki (15) dla resela (i, j) $t(i, j)$ jest porównywana z tym rozkładem:

$$p(i, j) = 2 \min\{P(T_{resamp}(j) \geq t(i, j)), P(T_{resamp}(j) \leq t(i, j))\} \tag{16}$$

Wartość $p(i, j)$ jest dwustronnym prawdopodobieństwem hipotezy zerowej H_0^{ij}.

Względny błąd p zgodnie z [11] dany jest wzorem:

$$err = \frac{\sigma_p}{p} = \sqrt{\frac{1 - p}{pN_{rep}}} \tag{17}$$

W rozważanym problemie wykonywać trzeba test dla każdego resela osobno. Pociąga to za sobą konieczność korygowania otrzymanych wartości p. W analizach opisywanych w dalszych częściach tego artykułu, po korekcie efektywny

poziom istotności często osiągał wartości rzędu 10^{-4}. Aby zapewnić estymację interesujących wartości $p(i, j)$ z błędem poniżej 10% potrzeba rzędu $N_{rep} = 10^{-6}$ powtórzeń.

2.3.2. Testy parametryczne

Jeśli dane pochodzą z rozkładu normalnego to formuła (15) daje zmienną z rozkładu Studenta z $f = N + N \cdot N_{ref} - 2$ stopniami swobody.

2.3.3. Poprawka Welch'a do testu t

Pewnym problemem w stosowaniu standardowego testu t może być fakt, że grupy, których średnie porównujemy, nie mają jednakowych wariancji. Znanym rozwiązaniem tego problemu jest poprawka zaproponowana przez Welch'a [28], dająca efektywną ilość stopni swobody dla testu t:

$$f = \frac{\left(\left(\frac{s_1^2}{n_1}\right)^2 + \left(\frac{s_2^2}{n_2}\right)^2\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}, \tag{18}$$

gdzie s_1 jest odchyleniem standardowym danych dla reseli w częstości j w obszarze referencyjnym ($n_1 = N \cdot N_{ref}$), zaś s_2 jest odchyleniem standardowym dla badanego resela w obszarze bodźca ($n_2 = N$).

2.4. Poprawka na wielokrotność testów

Wspomniane powyżej testy są stosowane do każdego resela osobno. Samo pojęcie poziomu istotności testu kryje jednak w sobie fakt, że na założonym poziomie możemy błędnie odrzucić hipotezę zerową. Jeśli wykonywanych jest wiele testów, to szansa na fałszywe odrzucenie którejś z hipotez zerowych rośnie. Najprostszym rozwiązaniem problemu testów wielokrotnych jest zastosowanie poprawki Benferroniego, polegającej na obniżeniu poziomu istotności pojedynczego testu tylokrotnie, ilokrotnie test taki jest wykonywany. Postępowanie to daje dobre wyniki jedynie dla testów, które są od siebie niezależne. W sytuacji, gdy testy są ze sobą w pewnym stopniu skorelowane, poprawka Benferroniego owocuje testami zbyt konserwatywnymi. W problemie tworzenia map statystycznych testy w sąsiednich reselach są skorelowane, co gorsza – ich struktura korelacji jest *a priori* nieznana. Rozwiązaniem zaproponowanym w [7] może być poprawka na wielokrotność testów oparta na False Discovery Rate (FDR, Benjamini i Yekutieli [1]). FDR jest stosunkiem fałszywie odrzuconych hipotez zerowych (m_0) do wszystkich odrzuconych hipotez zerowych (m). Jeśli zdecydowano się kontrolować FDR na poziomie 5%, to znaczy, że wśród reseli o których orzeczono, że zachodzi w nich istotna zmiana, co najwyżej 5% w istocie nie będzie wykazywać istotnej zmiany. W swojej pracy Benjamini i Yekutieli [1] dowiedli, że następująca procedura kontroluje FDR na zadanym poziomie q: ($\frac{m_0}{m} \le q$):

1. uporządkuj rosnąco oszacowane w pojedynczych testach prawdopodobieństwa hipotez zerowych $p_1 \le p_2 \le \cdots \le p_m$
2. znajdź k takie, że:
$$k = \max\left\{ i : p_i \le \frac{i}{m\sum_{j=1}^{m} \frac{1}{j}} q \right\}$$
3. odrzuć wszystkie hipotezy zerowe, dla których $p \le p_k$ (p_k jest efektywnym poziomem istotności dla wszystkich hipotez).

Wszystkie resele, w których odrzucona została hipoteza zerową zostają oznaczone jako wykazujące istotną zmianę $\overline{X^{TR}}$ w stosunku do obszaru odniesienia. W celu lepszego zobrazowania zmian mocy sygnału w dziedzinie czas-częstość w reselach tych wyrysowywana jest mapa względnych zmian mocy, zwana mapą ERD/ERS:

$$ERD/ERS(t, f) = \frac{\overline{E(t, f)} - \overline{E_{ref}(f)}}{\overline{E_{ref}(f)}} \tag{19}$$

2.5. Metoda średniej korelacji krótkoczasowej

Na średnich mapach ERD/ERS często widoczne są istotne zmiany zachodzące w różnych częstościach równocześnie. Niestety, z map tych nie można wyciągać wniosków co do współzależności pomiędzy obserwowanymi zmianami. Proponowana metoda rozwiązania problemu współzależności zmian w dwóch pasmach oparta jest o szacowanie funkcji korelacji wzajemnej. Dla badanego sygnału s realizuje ją następujący algorytm:

a) wydziel dwa interesujące pasma częstości np. poprzez filtrowanie pasmowe:
$$\begin{aligned} pas_1(t) &= F_1(s(t)) \\ pas_2(t) &= F_2(s(t)) \end{aligned} \tag{20}$$

b) w każdym paśmie częstości oblicz moc chwilową i przy pomocy średniej ruchomej znajdź jej obwiednię:
$$\begin{aligned} o_1(t) &= MA\left((pas_1(t))^2\right) \\ o_2(t) &= MA\left((pas_2(t))^2\right) \end{aligned} \tag{21}$$

c) dla realizacji i w kolejnych, zachodzących na siebie oknach $w(t)$ obliczamy funkcję korelacji wzajemnej obwiedni mocy obu pasm:
$$R^i_{o_1 o_2}(t, \tau) = R(o_1(t) \cdot w(t), \quad o_2(t) \cdot w(t), \quad \tau) \tag{22}$$

d) funkcje korelacji w odpowiadających sobie oknach uśredniamy po realizacjach:
$$\overline{R_{o_1 o_2}(t, \tau)} = \frac{1}{N} \sum_{i=1}^{N} R^i_{o_1 o_2}(t, \tau) \tag{23}$$

Tak uzyskaną średnią korelację krótkoczasową można zobrazować na mapie czas-przesunięcie (M), gdzie dla kolejnych chwil czasu wyrysowywana jest wartość korelacji między obwiedniami mocy pasm przy różnych wzajemnych przesunięciach.

Pozostaje jeszcze do rozstrzygnięcia, które z zaobserwowanych korelacji są istotne statystycznie.

2.5.1. Statystyki ekstremalne

Problem istotności statystycznej dla korelacji na mapach czas-przesunięcie można elegancko rozwiązać przez zastosowanie statystyk ekstremalnych [14]. Zakładamy hipotezę zerową, według której zaobserwowane wartości korelacji są czysto przypadkowe i moglibyśmy je otrzymać dla obwiedni mocy, które są od siebie niezależne. Hipotezą alternatywną jest ta, która mówi, że zaobserwowane wartości korelacji są wyjątkowo duże (co do wartości bezwzględnej) jak dla obwiedni mocy niezależnych od siebie. Powyżej sformułowane hipotezy implikują następujący algorytm oparty o repróbkowanie, kontrolujący istotność korelacji na poziomie α :

a) wygeneruj zbiór (Z) map czas-przesunięcie (MN_i) dla obwiedni mocy, o których *a priori* wiemy, że są od siebie niezależne $Z = \{MN_i, \quad i = 1,..., N_{map}\}$

b) dla każdej z owych map MN_i wyznacz największą i najmniejszą wartość korelacji:

$$c_{\min}^{i} = \min(MN_i) \qquad c_{\max}^{i} = \max(MN_i) \tag{24}$$

c) tworząc zbiór wartości minimalnych i maksymalnych:

$$C_{\min} = \{c_{\min}^{i}\} \qquad C_{\max} = \{c_{\max}^{i}\} \tag{25}$$

d) w zbiorze wartości minimalnych i maksymalnych znajdź wartość percentyla odpowiadającego połowie zadanego poziomu istotności:

$$c_{\min}^{kryt} = P_{\alpha/2}(C_{\min}) \qquad c_{\max}^{kryt} = P_{1-\alpha/2}(C_{\max}) \tag{26}$$

e) wartości korelacji na oryginalnej mapie czas-przesunięcie przekraczające wartości progowe ($c_{\max}^{kryt}$ w górę dla korelacji dodatnich i $c_{\min}^{kryt}$ w dół dla korelacji ujemnych) są istotne statystycznie na zadanym poziomie istotności. Holmes w [14] dowiódł, że procedura taka zapewnia kontrolę poziomu istotności także w przypadku testów wielokrotnych.

3. Dane

Działanie zaproponowanej metody zostanie zaprezentowane na danych ECoG. Dane pochodzą z prac [6], [7].

Pacjent cierpiący na epilepsje, został poddany operacji wszczepienia podczaszkowej siatki elektrod w celu lokalizacji ogniska napadów i dokonania mapowania funkcji motorycznych przed operacją wycięcia ogniska napadów. Elektrody składały się z platynowo-irydowych krążków w miękkim krzemowym pokryciu, ułożonych w macierz 6x8 z 1 cm odstępem między elektrodami. Elektrody zostały umieszczone powyżej lewych przednio-ciemieniowych obszarów kory, zawierających korę sensomotoryczną. Elektrody zostały umieszczone w tym obszarze z przyczyn medycznych.

W odpowiedzi na bodziec wzrokowy pacjent zaciskał swoją prawą pięść. Bodziec wzrokowy był widoczny przez 3 sekundy. Po zniknięciu bodźca pacjent zwalniał zacisk pięści. ECoG było próbkowane z częstością 1000 Hz.

Przed doświadczeniem pacjent poddany został działaniu leków przeciwpadaczkowych. ECoG wolne było od cech typowych dla epilepsji. Została dokonana eliminacja realizacji zawierających artefakty. Analizowane dane odniesiono do średniej ze wszystkich pozostałych kanałów. Dane składają się z 49 wolnych od artefaktów realizacji.

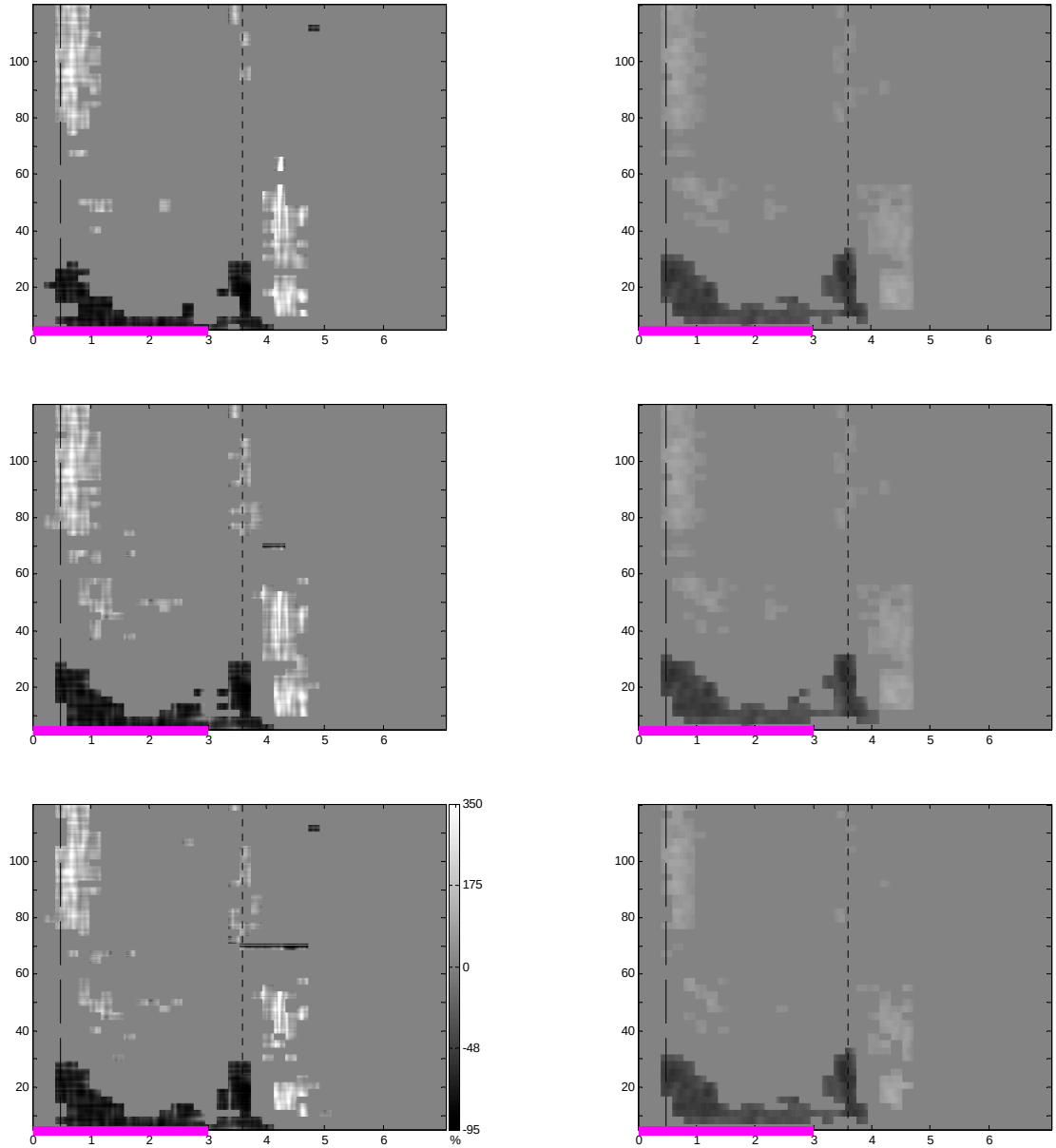
4. Wyniki

Zaprezentowane poniżej wyniki zostały otrzymane dla reseli o wymiarach $0.2\text{s} \times 2.45\,\text{Hz}$. Dane po transformacji Box'a-Cox'a miały rozkład normalny, co zostało potwierdzone testami Lillieforsa (test ten porównuje empiryczną dystrybuantę z dystrybuantą rozkładu normalnego o takiej samej średniej i wariancji, jak zaobserwowane w próbce; jest to test analogiczny do testu Kołmogorowa-Smirnowa z poprawką na fakt, że średnia i wariancja są estymowane z próbki).

Rysunek 1 pokazuje mapy ERD/ERS istotnych statystycznie. Każdy z paneli na rysunku 1 został uzyskany przy innej kombinacji metod estymowania rozkładu gęstości energii (MP albo spektrogram) i testów statystycznych (test repróbkowany, test t albo test t z poprawką Welch'a). Ogólnie, wszystkie zastosowane metody wskazują istotne ERD/ERS w tych samych regionach przestrzeni czas-częstość. Poniżej zostaną omówione szczegóły, którymi warianty proponowanej metody różnią się od siebie.

4.1. Porównanie działania różnych metod estymacji gęstości energii w dziedzinie czas-częstość

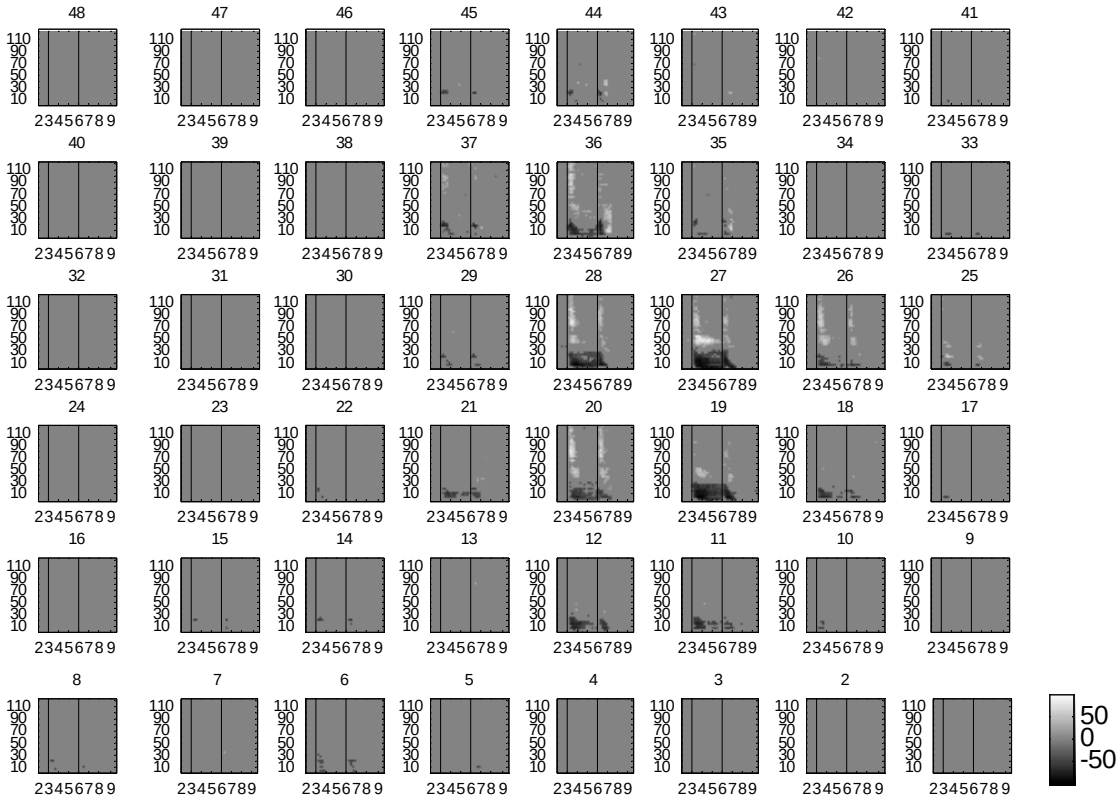
Na mapach otrzymanych przy użyciu metody MP obszary wykazujące zmiany są bardziej zwarte i ERD/ERS wyświetlane w istotnych reselach osiągają większe wartości niż dla odpowiadających im map uzyskanych przy użyciu spektrogramu. Jest to spowodowane dwoma własnościami estymatora MP. Po pierwsze struktury obecne w sygnale są opisywane przez odpowiadające im elementy słownika, wybierane w sposób optymalny dla każdej struktury. W metodzie opartej na spektrogramie jedna, z góry ustalona długość okna jest stosowana do detekcji wszystkich struktur. Po drugie mapa gęstości energii MP konstruowana jest jako suma map pojedynczych elementów słownika. Każdy element słownika ma zwartą reprezentację w przestrzeni czas-częstość. Sumując gęstości energii pojedynczych elementów, możemy w sposób jawny uniknąć reprezentowania na wynikowej mapie wyrazów mieszanych. Wyrazów mieszanych nie da się wyeliminować z map spektrogramowych.



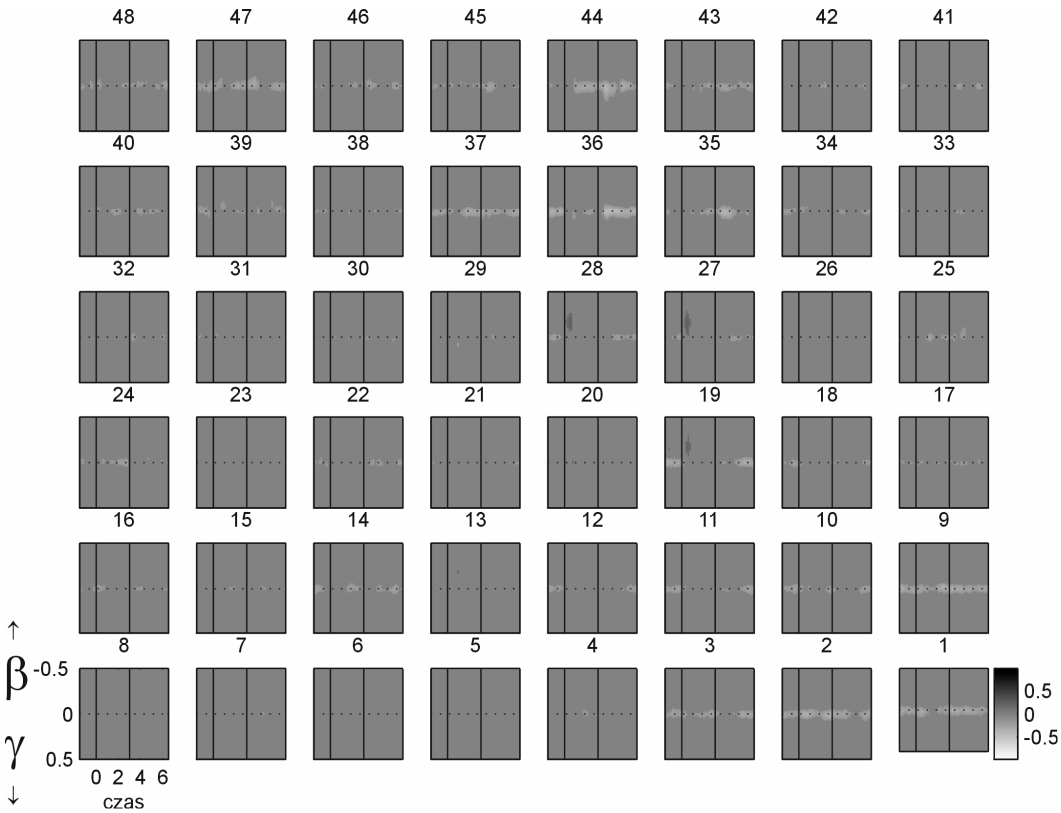
Rys. 1. Porównanie map istotnych statystycznie ERD/ERS uzyskanych przez zastosowanie różnych wariantów proponowanej metody do danych ECoG. Na każdej mapie oś pozioma przedstawia czas w sek.; oś pionowa częstość w Hz odcień szarości wartość ERD (ciemny) lub ERS (jasny). W lewej kolumnie przedstawione są mapy uzyskane z zastosowaniem MP, w prawej — z zastosowaniem spektrogramu do estymacji gęstości energii. W pierwszym wierszu do szacowania istotności statystycznej zastosowano test repróbkowany, w środkowym wierszu test *t* a w dolnym wierszu test *t*, z poprawką Welch'a. Pionowa przerywana linia około sek. 0.5 oznacza medianę czasów zaciśnięcia pięści wyznaczoną na podstawie EMG, druga pionowa przerywana linia (około sek. 3.5) oznacza medianę czasów zwolnienia zacisku pięści. Pozioma gruba linia na osi czasu oznacza czas prezentacji bodźca wzrokowego

Estymator gęstości energii oparty na MP produkuje mapę ciągłą – jest generowany na podstawie dopasowanych elementów słownika, których postać dana jest analitycznie. Spektrogram daje mapę, która jest quasi-ciągłą w dziedzinie czasu (okienko FFT przesuwane jest co jedną próbkę sygnału), natomiast w dziedzinie częstości mapa jest dyskretna, z rozdzielczością określoną przez rozdzielczość częstotliwościową FFT o zadanej długości okna. Rezultatem tego są mapy ERD/ERS gładkie w czasie, lecz ziarniste w częstości.

Wspomniane powyżej różnice między spektrogramem a MP są szczególnie dobrze widoczne w wysokich częstościach (>75 Hz) wkrótce po podaniu bodźca, gdzie widać, że wzrost energii jest największy wokół częstości 90 Hz i około 0.75 s po pojawieniu się bodźca. Widać również, że wzrost mocy po zakończeniu ruchu (pomiędzy 4 a 5 sekundą) w paśmie 10-50 Hz składa się z krótkotrwałych salw aktywności w paśmie beta i gamma.



Rys. 2. Topograficzny układ map pokazujących statystycznie istotne ERD (ciemny) i ERS (jasny) w dziedzinie czas-częstość. Pozioma oś – czas w sek., pionowa oś – częstość w Hz. Pionowe linie wyznaczają czas prezentowania bodźca wzrokowego



Rys. 3. Topograficzny układ map średnich korelacji krótko-czasowych pomiędzy obwiednią mocy w paśmie beta i w paśmie gamma. Czas na osi poziomej w sek., przesunięcie na osi pionowej w sek. Ujemne przesunięcie oznacza, że zmiana mocy w paśmie beta nastąpiła wcześniej niż w paśmie gamma. Wartość korelacji w skali szarości

4.2. Porównanie skuteczności zastosowanych testów statystycznych

Prezentowane mapy otrzymane były przy 5% poziomie tolerancji FDR. Oznacza to, że średnio na dwadzieścia reseli zaznaczonych jako istotnie wykazujące zmianę jeden może być zaznaczony fałszywie. Z tą świadomością wyniki prezentowane na rysunku 1 należy uznać za spójne dla testów repróbkowanych, stosowanych do danych nietransformowanych (X^{NT}) i dla parametrycznego testu t zastosowanego do danych transformowanych (X^{BC}). Od obu tych wyników odbiegają nieco mapy uzyskane przez zastosowanie testu t z poprawką Welch'a (zastosowanych do danych transformowanych X^{BC}) – dolny wiersz na rys. 1. Wyniki otrzymane z zastosowaniem poprawki Welch'a wydają się wskazywać, że stosowanie tej poprawki w przypadku prezentowanych danych daje nieco zbyt konserwatywne testy. Prawdopodobnie wynika to z faktu, że liczebność grupy danych w obszarze referencyjnym jest o rząd wielkości większa niż liczebność grupy danych z badanych reseli.

4.3. Zastosowanie metody średnich korelacji krótkoczasowych do badania współzależności pomiędzy zmianami mocy w paśmie beta i gamma

Na rysunku 2 przedstawiono mapy istotnych statystycznie zmian w mocy sygnału ECoG związanych z zaciskaniem i zwalnianiem pięści. Mapy te ułożone są topograficznie. Widać, że w kanale 27 i w kanałach sąsiednich zachodzi wiele zmian w różnych pasmach częstości. W szczególności można wyróżnić ERD w częstościach od 10 do 30 Hz, krótkotrwałe, szerokopasmowe ERS w częstościach gamma (40-120 Hz) tuż po pojawieniu się i po zaniku bodźca wzrokowego, ERS w paśmie beta po zakończeniu ruchu (kanały 36, 35 i 25) oraz ERS w czasie utrzymywania zacisku pięści w kanale 27. Narzuca się pytanie: czy istnieje zależność pomiędzy ERD w paśmie beta i ERS w paśmie gamma, szczególnie w tzw. niskiej gammie (35-60 Hz). Prezentowane na rys. 2 mapy, choć sugerują taką zależność, to nie mogą jej potwierdzić ani obalić. Można sobie wyobrazić, że w skrajnej sytuacji ERD w paśmie beta występuje w tych realizacjach, w których nie występuje ERS w paśmie gamma, a mapy istotnych statystycznie ERD i ERS w przestrzeni czas-częstość wyglądałyby tak jak na rys. 2. Do rozstrzygnięcia sformułowanego powyżej problemu zastosowano metodę średniej korelacji krótkoczasowej. Znaleziono obwiednie mocy sygnałów w pasmach beta (15-28 Hz) i gamma (40-60 Hz) poprzez filtrowanie pasmowe filtrami Butterwortha. Uzyskane mapy średnich korelacji w przestrzeni czas-przesunięcie przedstawione są na rys. 3. Widać, że w kanałach 28, 27 i 19 zanik mocy w paśmie beta jest związany z następującym po nim wzrostem mocy w paśmie gamma. Zjawiska te przesunięte są względem siebie o około 0.2 s. Na rys. 3 można zaobserwować też, że w wielu kanałach układających się w pas wokół odcinka wyznaczonego przez kanały 47 i 1 występują synchronizacje (korelacje z zerowym przesunięciem) pomiędzy obwiedniami mocy pasm beta i gamma. Synchronizacje te są w wyraźny sposób zaburzane w okresie występowania bodźca wzrokowego, przy czym im dalej od kanału 27, tym owe synchronizacje są mniej zaburzane przez wystąpienie bodźca. Zaobserwowane synchronizacje są najprawdopodobniej efektem występowania krótkich salw aktywności, których pasmo częstości jest na tyle szerokie, że po filtrowaniu pasmowym ich moc częściowo występuje w paśmie beta, a częściowo w paśmie gamma.

5. Podsumowanie

Wszystkie warianty metody estymowania istotnych statystycznie różnic w energii sygnału w przestrzeni czas-częstość dają spójne wyniki. Wnikliwa obserwacja zaprezentowanych rysunków ujawnia jednak pewne różnice pomiędzy proponowanymi wariantami metody. Wersja stosująca MP do estymacji gęstości energii i test repróbkowany do oszacowania istotności statystycznej jest najbardziej wolna od *a priorycznych* założeń i z tego powodu powinna być traktowana jako metoda odniesienia. Wariant ten jest jednak najbardziej kosztowny obliczeniowo. W praktycznych zastosowaniach można z powodzeniem zastąpić test repróbkowany parametrycznym testem t, stosowanym do znormalizowanych danych. Jako metoda normalizacji danych bardzo efektywna i elastyczna okazała się transformacja Box'a-Cox'a. Dalsze znaczące skrócenie czasu obliczeń w badaniach, w których mikrostruktura ERD/ERS nie jest kluczowa, można osiągnąć przez zastąpienie estymacji gęstości energii sygnału opartej o MP przez spektrogram.

Metoda średnich korelacji krótkoczasowych stanowi narzędzie komplementarne do czasowo-częstotliwościowych map ERD/ERS. Pozwala na wnioskowanie o zależności bądź niezależności zmian zachodzących w różnych pasmach częstości widocznych na mapach ERD/ERS.

Podziękowania

Prezentowana praca jest częściowo finansowana z grantu KBN nr 4T11E 028 23.

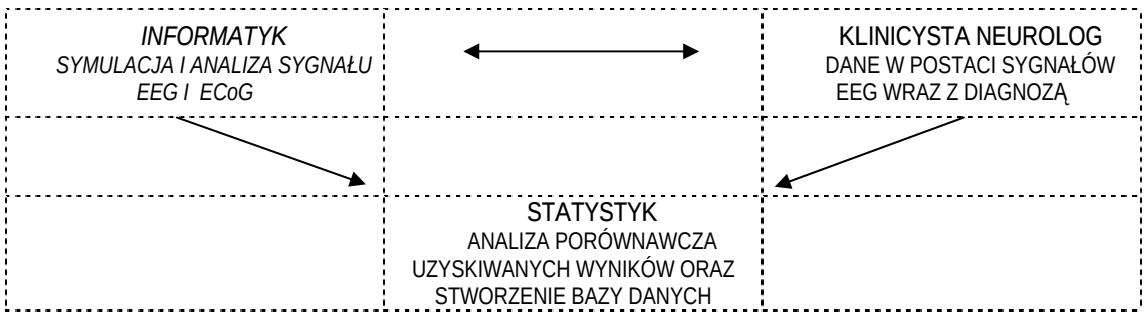
Literatura cytowana

1. Benjamini Y., Yekutieli Y.: The control of the false discovery rate under dependency, Ann. Stat. 29, 1165-1188, 2001.
2. Box G. E. P., Cox D. R.: An analysis of transformations, Journal of the Royal Statistical Society 2, 211-252, 1964.
3. Conover W. J.: Practical Nonparametric Statistics, Wiley, New York, 1980.
4. Crone N. E., Boatman D., Gordon B. Hao L.: Induced electrocorticographic gamma activity during auditory perception, Clinical Neurophysiology 112, 565-82, 2001a.
5. Crone N. E., Hao L., Hart Jr J., Boatman D., Lesser R., Irizarry R.: Electrocorticographic gamma activity during word production in spoken and sign language, Neurology 57, 2045-53, 2001b.
6. Crone N. E., Miglioretti D. L., Gordon B., Lesser R. P.: Functional mapping of human sensorimotor cortex with electrocorticigraphic spectral analysis II. Event-related synchronization in the gamma band, Brain 121, 2301-315, 1998a.
7. Crone, N. E., Miglioretti D. L., Gordon B., Sieracki J. M., Wilson M. T., Uematsu S.: Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis I. Alpha and beta event-related desynchronization, Brain 121, 2271-2299, 1998b.

8. Durka P., Żygierewicz J., Klekowicz H., Ginter J., Blinowska K.: On the statistical significance of event-related EEG desynchronization and synchronization in the time-frequency plane, IEEE Transactions on Biomedical Engineering 51 (7), 1167-1175, 2004.
9. Durka P. J., Ircha D., Blinowska K. J.: Stochastic time-frequency dictionaries for matching pursuit, IEEE Tran Signal Process, 49 (3), 507-510, 2001a.
10. Durka P. J., Ircha D., Neuper C., Pfurtscheller G.: Time-frequency microstructure of event-related desynchronization and synchronization, Med Biol Eng Comput 39 (3), 315-321, 2001b.
11. Efron B., Tibshirani R. J.: An Introduction to the Bootstrap, Chapman& Hall, 1993.
12. Ginter Jr, J., Blinowska K. J., Kamiński M., Durka P. J.: Phase and amplitude analysis in time-frequency space-application to voluntary finger movement, Journal of Neuroscience Methods 110, 113-124, 2001/
13. Graimann B., Huggins J. E., Levine S. P., Pfurtscheller G.: Visualizationof significant ERD/ERS patterns in multichannel EEG and ECoG data, Clinical Neurophysiology 113, 43-47, 2002.
14. Holmes A. P.: Statistical issues in functional brain mapping, Praca Doktorska, Uniwersytet w Glasgow, 1994, dostępna w internecie pod adresem:

    http://www.fil.ion.ucl.ac.uk/spm/doc/theses/andrew/
15. Hyde S.: Likelihood based inference on the Box-Cox family of transformations: Sas and Matlab programs, Tech. rep., Mathematical Sciences, Montana State University, dostępny w internecie pod adresem:

    www.math.montana.edu/hyde/papers/mspape.pdf
16. Kaiser J., Lutzenberger W.: Induced gamma-band activity and human brain function, Neuroscientist 9, 475-84, 2003.
17. Lachaux J.-P., Rodriguez E., Martinerie J., Adam C., Hasboun D., Varela F. J.: A quantitative study of gamma-band activity in human intracranial recordings triggered by visual stimuli, European Journal of Neuroscience 12, 2608-2622, 2000.
18. Makeig S.: Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. Electroencephalogr Clin Neurophysiol 86, 283-293, 1993.
19. Mallat S., Zhang Z.: Matching pursuit with time-frequency dictionaries, IEEE Tran Signal Process, 41, 3397-3415, Dec 1993.
20. Pfurtscheller G.: EEG event-related desynchronization (ERD) and event-related synchronization (ERS), In: Niedermayer, E., Lopes da Silva, F. H. (Eds.), Electroencephalography: Basic Principles, Clinical Applications and Related Fields, 4th Edition. Williams & Wilkins, pp. 958-965, 1999a.
21. Pfurtscheller G.: Quantification of ERD and ERS in the time domain. In: Pfurtscheller G., Lopes da Silva F. H. (Eds.), Event-related desynchronization. Vol. 6. Elsevier, pp. 89-105, 1999b.
22. Pfurtscheller G., Flotzinger D., Neuper C.: Differentiation between finger, toe and tongue movement in man based on 40 Hz EEG, Electroencephalography and Clinical Neurophysiology 90, 456-460, 1994.
23. Pfurtscheller G., Lopes da Silva F. H.: Event-related EEG/MEG synchronization and desynchronization: basic principles, Clinical Neurophysiology 110, 1842-1857, 1999.
24. Pfurtscheller G., Stancak Jr, A. C. N.: Post-movement beta synchronization. A correlate of an idling motor area? Electroencephalography and Clinical Neurophysiology 98, 281-93, 1996a.
25. Pfurtscheller G., Stancak Jr, A., Neuper C.: Event-related synchronization (ERS) in the alpha band–an electrophysiological correlate of cortical idling: a review. International Journal of Psychophysiology 24, 39-46, 1996b.
26. Tallon-Baudry C., Bertrand O.: Oscillatory gamma activity in humans and its role in object representation. Trends Cogn Sci 3, 151-162, 1999.
27. Tallon-Baudry C., Bertrand O., Delpuech C., Pernier J.: Stimulus specificity of phase-locked and non-phase-locked 40Hz visual responses in human, Journal of Neuroscience 16 (13), 4240-4249, 1996.
28. Welch B.: The significance of the difference between two means when the population variances are unequal. Biometrika 29, 350-362, 1938.
29. Westfall P. H., Young S. S.: Resampling-Based Multiple Testing. John Wiley & Sons, 1993.
30. Żygierewicz J., Durka P. J., Klekowicz H., Franaszczuk P. J., Crone N. E.: Computationally efficient approaches to calculating significant ERD/ERS changes in the time-frequency plane, Journal of Neuroscience Methods, przyjęty do druku, 2005.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 349-354.

AUTOMATYCZNY SYSTEM ZDALNEGO NADZORU KARDIOLOGICZNEGO OPARTY NA PROFESJONALNYCH RELACJACH WŚRÓD LEKARZY

AUTOMATIC MONITORING SYSTEM FOR CARDIOLOGY BASED ON PROFESSIONAL RELATIONSHIPS OF HUMAN EXPERTS

PIOTR AUGUSTYNIAK

Katedra Automatyki Akademii Górniczo-Hutniczej, Kraków, august@agh.edu.pl

Streszczenie. Artykuł przedstawia projekt rozproszonego systemu monitorowania czynności serca pacjentów znajdujących się w nieograniczonej odległości od centrali diagnostycznej. Cechą charakterystyczną proponowanego rozwiązania jest optymalizacja użycia łączy bezprzewodowych poprzez adaptację częstotliwości połączeń z monitorami zdalnymi oraz selektywne i priorytetowe przesyłanie parametrów diagnostycznych. Dopasowanie parametrów połączenia do bieżącego stanu pacjenta, celu diagnostycznego oraz warunków technicznych dokonywane jest na podstawie rezultatów automatycznej analizy sygnału. Zastosowane metody adaptacji są wzorowane na utrwalonych w praktyce relacjach w środowisku kardiologicznym oraz uwzględniają standardy postępowania diagnostycznego opublikowane przez stowarzyszenia lekarskie.

Słowa kluczowe: telemedycyna, nadzór w miejscu zamieszkania, kardiologia

Abstract. This paper introduces a concept of a distributed system for heart activity with unconstrained distance between the patient and the supervising center. The proposed approach features an optimum use of wireless channels through the adaptation of connection frequency and through a selective and prioritized transmission of diagnostic data. Adaptation of connection parameters to the current patient status, diagnostic goals and technical conditions in the wireless network is based on the medical interpretation of the ECG performed automatically. Selected adaptation rules are justified by professional relations between human experts in cardiology well established in medical practice, and take account of medical standards promoted by international societies of cardiology.

Key words: telemedicine, home-care, cardiology

1. Wstęp

Dynamiczny rozwój środków łączności o zasięgu globalnym oraz zapotrzebowanie na nieprzerwany monitoring podstawowych funkcji życiowych ze strony starzejących się społeczeństw spowodowały w ostatnich latach wzrost zainteresowania telediagnostyką ze strony naukowców oraz czołowych światowych producentów aparatury medycznej [2], [3], [5], [8]. Wzrost ten zamanifestował się już wdrożeniami kilku systemów nadzoru kardiologicznego i powstaniem ośrodków zdalnej analizy i archiwizacji danych medycznych, a także współpracujących służb interwencyjnych [6], [7], [9].

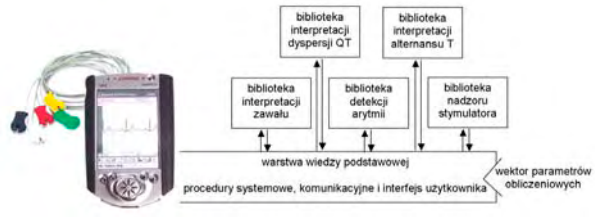
Analiza pracy tych systemów prowadzi do wniosku, że zarówno podział zadań pomiędzy zdalne urządzenia rejestrujące noszone przez pacjenta i serwer centrum diagnostycznego, jak i sposób transmisji realizowano, powielając rozwiązania typowe dla połączeń przewodowych. Charakterystycznymi wadami tych sieci są: sztywne przyporządkowanie procedur interpretacyjnych do jednego z urządzeń oraz nieoptymalne wykorzystanie bezprzewodowego kanału transmisji danych.

Prace badawcze podjęte i przeprowadzone w Laboratorium Biocybernetyki miały za zadanie identyfikację potrzeb i technicznych możliwości realizacji adaptacji algorytmów interpretacji sygnału EKG oraz efektywnego formatu wymiany informacji o zróżnicowanym poziomie przetworzenia opartego na medycznie uzasadnionej hierarchii priorytetów. Podejmując powyższe prace, uwzględniono szereg uwarunkowań natury technicznej i medycznej:

- dążenie do maksymalnej autonomii urządzenia zdalnego,
- minimalizację strumienia informacji, a w konsekwencji kosztów eksploatacji połączenia bezprzewodowego,
- konieczność spełnienia standardów interpretacji obowiązujących w kardiologii,
- niezawodność informacji rozumiana jako minimalne opóźnienie i zarazem minimalne prawdopodobieństwo zniekształceń,
- umożliwienie wspomagania automatyki interpretacji przez eksperta kardiologa.

Każdy system adaptacyjny wymaga uważnego doboru reguł adaptacji, w systemie przeznaczonym do diagnostyki medycznej reguły te warunkują poprawne określenie stanu osoby badanej i właściwy dobór środków terapeutycznych. Wobec tak postawionych bardzo ostrych wymagań ciekawym rozwiązaniem wydaje się zaczerpnięcie wzorca z utrwalonej praktyki medycznej i zbudowanie reguł adaptacji systemu w oparciu o model relacji pomiędzy kardiologami. Model ten jest uogólnieniem zaobserwowanych przepływów **wiedzy** i **informacji** pomiędzy kardiologami o różnym doświadczeniu. Pojęcia te, chociaż z informatycznego punktu widzenia zbliżone, będą używane w tej pracy dla określenia odpowiednio:

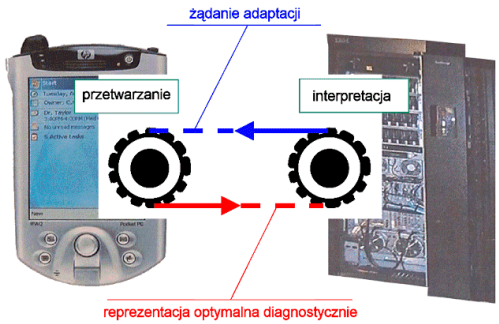
- **wiedza** – zasobu metod interpretacji sygnału i reguł ich przetwarzania w znaczeniu zbliżonym do *pojęcia algorytmu* (rys. 1)
- **informacja** – zasobu sygnałów i innych danych medycznych o różnym stopniu przetworzenia w znaczeniu zbliżonym do *pojęcia danych*.



Rys. 1. Schemat uaktualniania wiedzy rejestratora zdalnego

W zaproponowanym rozwiązaniu adaptacja elementów systemu (rys. 2) na wzór relacji pomiędzy kardiologami przebiega w trzech aspektach następujących:

- początkowa wiedza urządzenia i uzupełniania wiedzy w zależności od cech osobniczych pacjenta i bieżących potrzeb diagnostycznych,
- formatu raportowania w zakresie zawartości i priorytetów informacji przekazywanych do centrum nadzoru,
- częstotliwości raportowania i zasad inicjowania łączności przez oba urządzenia.



Rys. 2. Wymiana informacji pomiędzy serwerem i rejestratorem zdalnym

Inne ewentualnie możliwe wymiary adaptacji także były rozważane, ale ostatecznie nie zostały uwzględnione w raportowanym zakresie prac. W rozdziale drugim zostanie przedstawiona analiza relacji w każdym z wymienionych aspektów oraz ich analogia w rozproszonym zautomatyzowanym systemie nadzoru. Rozdział trzeci jest poświęcony przedstawieniu środków technicznych służących do realizacji adaptacji urządzeń w systemie, a zamykający rozdział czwarty zawiera dyskusję i wnioski końcowe.

2. Model przepływu wiedzy i informacji w elektrodiagnostyce kardiologicznej

Rozdział ten przedstawia generalne wnioski z obserwacji relacji w środowiskach kardiologicznych i koncepcje analogii w systemie automatycznego nadzoru kardiologicznego.

2.1. Wiedza początkowa

Wiedza początkowa to zbiór metod interpretacji sygnału, jakie potrafi stosować początkujący kardiolog. Niezależnie, czy przykładem będzie absolwent medycyny, czy lekarz innej specjalności początkujący w kardiologii, wiedzę początkową charakteryzuje zdolność samodzielnej interpretacji najprostszych patologii. W miarę wzrostu stopnia skomplikowania niepewność diagnozy powiększa się aż do osiągnięcia stanu, w którym konieczna jest pomoc bardziej doświadczonego specjalisty. Należy także zwrócić uwagę na dwie prawidłowości:

- liczba wiarygodnie interpretowanych jednostek chorobowych jest niewielka, ale dzięki skośnemu rozkładowi statystycznemu i poprawnemu doborowi przypadków podczas studiowania ilość elektrokardiogramów interpretowanych poprawnie stanowi znaczny odsetek zapisów,
- objętość informacyjna zapisu surowego jest największa, proces interpretacji prowadzi do zmniejszenia objętości rezultatu diagnostycznego.

Dla zasilanego z baterii urządzenia przenośnego o niewielkiej mocy obliczeniowej z powyższych obserwacji wynikają następujące priorytety dotyczące podstawowej biblioteki procedur interpretacyjnych:

- implementacja procedur, których użycie jest najbardziej prawdopodobne z punktu widzenia spodziewanych patologii,
- implementacja procedur, których działanie redukuje objętość strumienia danych w sposób najbardziej znaczący.

2.2. Uaktualnianie wiedzy

Uaktualnianie wiedzy to długotrwały proces obejmujący praktycznie całą karierę zawodową kardiologa. Proces ten jest związany z niewielkim prawdopodobieństwem napotkania rzadkich przypadków, stałym postępem nauk medycznych oraz z analizą dalszej historii pacjentów uprzednio diagnozowanych. Źródłami dodatkowych umiejętności interpretacyjnych są przeważnie:

- materiały publikowane,
- konsultacje ze specjalistami o większym doświadczeniu, wśród nich szczególne znaczenie ma konsultacja przypadków uznanych początkowo za zbyt trudne, w wyniku której podobne przypadki w przyszłości będą mogły być analizowane samodzielnie.

Uaktualnianie wiedzy, czyli biblioteki procedur interpretacji sygnału, jest bardzo istotną cechą przenośnego urządzenia, w którym ze względów technicznych lub ekonomicznych zaimplementowano jedynie podstawowy podzbiór funkcji. Dzięki uaktualnianiu możliwa jest adaptacja funkcjonalna urządzenia z uwzględnieniem cech osobniczych pacjenta, jego stopnia zagrożenia, a także bieżącego celu diagnostycznego. W rozproszonym systemie automatycznej interpretacji

przewidziano dwa mechanizmy uaktualniania wiedzy rejestratora zdalnego przez serwer centralny:
- uaktualnianie narzucone, inicjowane w przypadku zmiany celu diagnostycznego albo pojawienia się podejrzanych lub sprzecznych raportów i oparte na centralnym rejestrze możliwości funkcjonalnych rejestratorów współpracujących w sieci,
- uaktualnianie na żądanie, inicjowane w przypadku istotnej liczby zapisów raportowanych jako nierozwiązane przez rejestrator zdalny, jeżeli procedura prowadząca do ich rozwiązania jest wykonalna w urządzeniu przenośnym, a koszt uaktualnienia jest niższy niż koszt nadsyłania raportów zawierających sygnał surowy.

Uaktualnianie wiedzy ma dodatkowe zalety wdrożeniowe: przy masowej produkcji gwarantującej niższy koszt pozwala osiągnąć funkcjonalność urządzenia dedykowanego dla określonej osoby, dla określonego zadania, a także urządzenia o przeznaczeniu zmiennym w kontekście historii rezultatów diagnostycznych.

2.3. Częstotliwość raportowania

W codziennej praktyce lekarskiej częstotliwość diagnozowani i raportowania jest zmienna i uzależniona od historii rezultatów diagnostycznych pacjenta. Sytuacje fizjologiczne wpływają na wydłużenie okresu raportowania, a sytuacje patologiczne powodują jego skrócenie aż do ciągłego raportowania na bieżąco.

W automatycznym systemie nadzoru kardiologicznego zastosowano trzy sposoby inicjowania badania, co wpływa bezpośrednio na częstotliwość raportowania:
- badanie planowe, inicjowane przez serwer centralny, wykonywane w interwałach czasu o długości uzależnionej od rezultatów otrzymanych w wyniku interpretacji poprzednich zapisów,
- badanie interwencyjne, inicjowane przez rejestrator zdalny na podstawie przekroczenia zakresu tolerancji parametrów diagnostycznych nadzorowanych w sposób ciągły,
- badanie na żądanie, inicjowane przez pacjenta w sytuacjach złego samopoczucia; dzięki pętli rejestrującej sygnał możliwe jest odtworzenie fragmentu zapisu bezpośrednio poprzedzającego epizod.

Zmienność częstotliwości raportowania ma decydujący wpływ na użycie cyfrowego kanału transmisyjnego, a zatem jest podstawowym sposobem redukcji kosztów eksploatacyjnych. Czynnik ten to także narzędzie preselekcji danych, umożliwiające redukcję ilości komunikatów o poprawnej pracy serca, a jednocześnie utrzymanie stałej gotowości diagnostycznej w przypadkach nagłych.

2.4. Raportowanie sytuacji typowej

Wiedza i umiejętności interpretacji zapisu EKG, jakimi dysponuje kardiolog, są kształtowane przez jego doświadczenie i częstotliwość napotykania przypadków poszczególnych patologii. Zatem w sposób naturalny przypadki, z którymi kardiolog styka się najczęściej, są przez niego interpretowane z największą poprawnością. Zjawisko specjalizacji przypomina pętlę dodatniego sprzężenia zwrotnego, gdyż z jednej strony statystyczny rozkład przypadków kształtuje wiedzę kardiologa, a z drugiej pacjenci z podejrzeniami określonych przypadków kierowani są do specjalistów, którzy najbardziej poprawnie te przypadki interpretują.

Zgodnie z przyjętymi założeniami odnośnie maksymalizacji prawdopodobieństwa wystąpienia przypadków interpretowanych poprawnie przez funkcje wchodzące w skład wiedzy podstawowej rejestratora zdalnego – sytuacja typowa jest przetwarzana autonomicznie. W konsekwencji raport sytuacji typowej zawiera informacje dotyczące podstawowych parametrów diagnostycznych (m.in. rytmu serca) w postaci bezwzględnej albo przyrostowo w odniesieniu do poprzedniego raportu. Priorytet pakietów raportu sytuacji typowej w sieci telekomunikacyjnej może być ustawiony jako normalny albo zmieniany adaptacyjnie w zależności od średniego opóźnienia pakietów. Opóźnienia raportu sytuacji typowej sięgające nawet do 30% okresu raportowania mogą być tolerowane.

Dzięki adaptacyjności funkcjonalnej rejestratora sytuacja typowa nie jest ograniczona do przypadków fizjologicznie prawidłowych (np. rytmu zatokowego miarowego). Niektóre patologie, o ile ich charakter jest utrwalony (np. migotanie przedsionków), również mogą zostać zdefiniowane jako typowe, podlegać autonomicznej interpretacji i generować skróconą formę raportu. Jeżeli sytuacje uznane jako typowe nie mogą być poprawnie zinterpretowane przez procedury wchodzące w zakres wiedzy podstawowej, to ich zdefiniowanie wymaga uprzedniego uzupełnienia wiedzy rejestratora zdalnego. Ograniczeniem zakresu definicji sytuacji typowej jest moc obliczeniowa rejestratora zdalnego, niewystarczająca do poprawnej interpretacji niektórych trudniejszych zapisów.

2.5. Raportowanie sytuacji nieoczekiwanej

Wobec znacznej zmienności osobniczej oraz różnorodności patologii występowanie sytuacji nieoczekiwanych, trudnych do poprawnej interpretacji jest cechą zawodu lekarskiego [4]. Rozwiązaniem wykształconym w praktyce przez lata historii medycyny jest zróżnicowana specjalizacja poszczególnych osób:
- lekarzy pierwszego kontaktu dysponujących szerokim zakresem (w znaczeniu rozmaitości przypadków) stosunkowo płytkiej wiedzy (w znaczeniu zdolności interpretacji przypadków bardzo skomplikowanych),
- lekarzy specjalistów, których zakres wiedzy jest wąski, ale głęboki, co oznacza, że potrafią zinterpretować prawie każdy dowolnie trudny przypadek patologii, ale w zakresie ograniczonym do jednego narządu.

Drugim elementem wykształconym w historii medycyny są relacje lekarzy o zróżnicowanym stopniu specjalizacji:
- wymiana informacji prowadzi do interpretacji przez specjalistę przypadków przekraczających zakres wiedzy lekarza pierwszego kontaktu,
- uaktualnianie wiedzy prowadzi do pogłębienia jej zakresu przez lekarza pierwszego kontaktu i powiększa stopień jego samodzielności, przynajmniej w zakresie kolejnych podobnych przypadków.

Trudno się oprzeć wrażeniu, że przedstawione wymiary wiedzy lekarskiej: szerokość i głębokość są wzajemnie prostopadłe, a wyznaczony przez zakres wiedzy prostokąt ma powierzchnię proporcjonalną do zdolności i doświadczenia, a proporcje boków mogą być określone niemal dowolnie [10]. Analogiczne rozważanie, dotyczące automatycznego systemu interpretacyjnego, prowadzi do wniosku, że ograniczeniem

powierzchni prostokąta wiedzy rejestratora zdalnego są jego zasoby: moc obliczeniowa, wielkość pamięci i maksymalny czas pracy z autonomicznym źródłem zasilania.

Każda sytuacja nieoczekiwana potencjalnie jest manifestacją patologicznej pracy serca, dodatkowo jej interpretacja wykracza poza możliwości interpretacyjne zakresu wiedzy, jaki w chwili jej wystąpienia posiada rejestrator zdalny. W konsekwencji sytuacja nietypowa nie może być poprawnie zinterpretowana w sposób autonomiczny. Poprawna pełna interpretacja wymaga przesłania sygnału surowego lub częściowo przetworzonego (meta dane) do serwera centrali nadzorującej, dysponującego większą wiedzą, większą mocą obliczeniową i wyposażonego w opcję nadzoru ze strony eksperta kardiologa. Niestety, strumień informacji o niższym stopniu przetworzenia ma znacząco większą objętość, w porównaniu z raportem sytuacji typowej.

Prawidłowy dobór zakresu wiedzy rejestratora zdalnego gwarantuje niską częstotliwość sytuacji nieoczekiwanych. Dodatkowo w procesie uzupełniania wiedzy program interpretacyjny rejestratora adaptuje się do specyfiki pacjenta, co dalej zmniejsza ilość takich sytuacji. Ze względu na znaczenie medyczne i potencjalną konieczność interwencji pakiety raportu sytuacji nieoczekiwanej muszą mieć wysoki priorytet w sieci telekomunikacyjnej. Tolerancja opóźnień w stosunku do czasu rzeczywistego rejestracji nie powinna przekraczać wartości typowo tolerowanych w elektrokardiogra fach stacjonarnych, czyli 1-2 sekund.

3. Rozwiązania techniczne

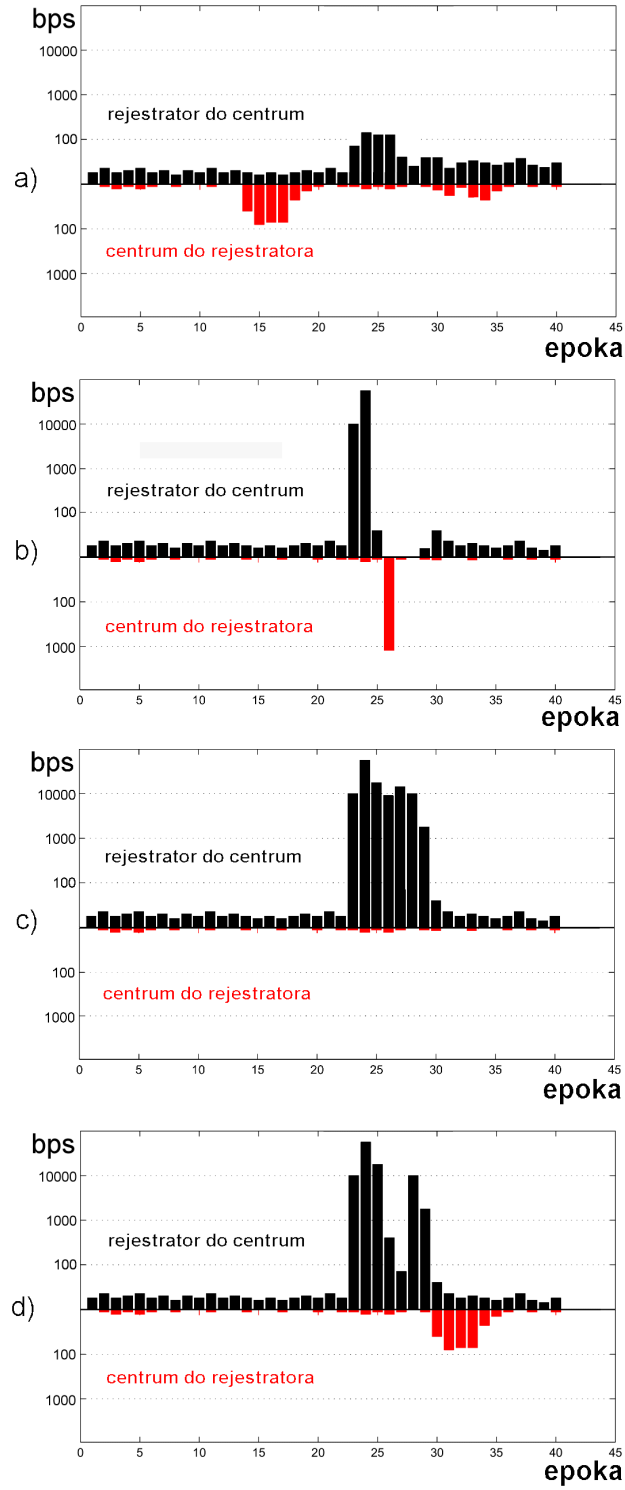
Dyskusja nad zjawiskami i relacjami obserwowanymi w środowisku kardiologów i ich analogii w automatycznym systemie nadzoru kardiologicznego doprowadziła do zaprojektowania prototypu. Rozdział ten jest poświęcony opisowi podstawowych rozwiązań technicznych zastosowanych do zrealizowania głównych funkcji systemu.

3.1. Uaktualnianie oprogramowania i modyfikacja jego działania

Uaktualnianie i modyfikacja oprogramowania interpretacyjnego są głównymi metodami adaptacji rejestratora zdalnego do cech osobniczych pacjenta oraz potrzeb diagnostycznych.

Modyfikacja oprogramowania interpretującego polega na zdalnej zmianie wartości parametrów obliczeniowych (wartości progowe, współczynniki itp.) w dowolnym momencie. Zestaw funkcji wchodzący w skład wiedzy podstawowej rejestratora zawiera 73 parametry obliczeniowe, a dodatkowe biblioteki dołączane w miarę potrzeby powiększają tę liczbę do 168. Dokonywanie zmian jest inicjowane przez serwer nadzorujący na podstawie rezultatów diagnostycznych, wykrytych błędów interpretacji i informacji o aktualnych wartościach parametrów, jaką serwer otrzymuje od rejestratora zdalnego.

Uaktualnianie oprogramowania jest metodą uaktualniania wiedzy rejestratora poprzez wyposażenie go w dodatkowe procedury interpretacyjne w zależności od potrzeb i wcześniejszej historii pacjenta. Uaktualnianie oprogramowania wykorzystuje mechanizm dynamicznie dołączanych bibliotek, które w postaci plików binarnych przesyłane są z serwera nadzorującego do pamięci rejestratora. Ten proces także oparty jest na rezultatach diagnostycznych pacjenta, informacji o aktualnej wiedzy rejestratora oraz analizie sytuacji nieoczekiwanych (rys. 3).

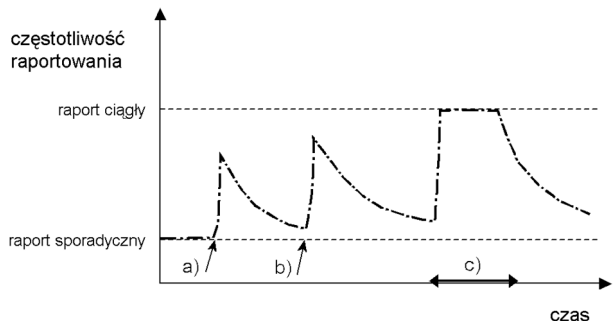


Rys. 3. Przykłady przepływu informacji pomiędzy serwerem centrum diagnostycznego a rejestratorem zdalnym: a) zmiana celu diagnostycznego, b) modyfikacja parametrów obliczeniowych, c) raportowanie sytuacji nieoczekiwanej, d) sytuacja nieoczekiwana i uaktualnienie wiedzy rejestratora

3.2. Sterowanie częstotliwością raportowania

Sterowanie częstotliwością raportowania odbywa się niezależnie w rejestratorze i serwerze nadzorującym (rys. 4). W przypadkach typowych rejestrator pracuje autonomicznie i przesyła raporty z częstotliwością wyliczoną przez siebie na podstawie rezultatów diagnostycznych. Okres raportowania jest także częścią składową raportu, a zatem w sytuacji braku raportu w określonym czasie serwer nadzorujący może zażądać raportu natychmiastowego. Podobne rozwiązanie przyjęto dla obsługi sytuacji nietypowych, interpretowanych przez serwer. W tym przypadku rejestrator zdalny nie może wyliczyć okresu raportowania i przejmuje wartość narzuconą przez serwer.

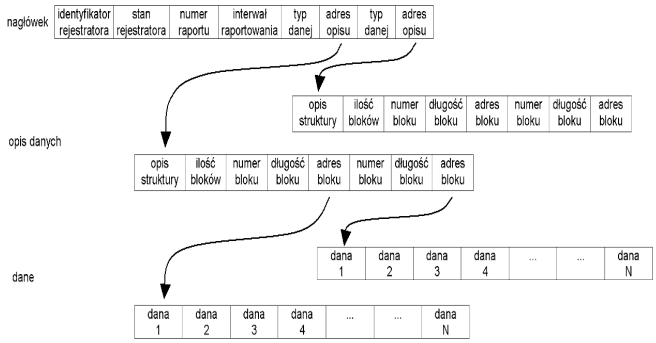
W przypadku raportowania ciągłego przyjęto odmienne rozwiązanie sterowania raportem, nadzorowane wyłącznie ze strony serwera. W tym trybie pracy rejestrator wykonuje tylko podstawowe analizy (m.in. analizę rytmu) w czasie rzeczywistym, albo nie wykonuje żadnych obliczeń interpretacyjnych i przesyła sygnał surowy. Czas trwania raportu ciągłego jest ograniczony przez możliwości obliczeniowe serwera wykonującego prawie wszystkie czynności związane z interpretacją i na podstawie rezultatów decyduje o przejściu w tryb monitorowania ciągłego.



Rys. 4. Przykłady modyfikacji częstotliwości raportowania, a) i b) sytuacje patologiczne zinterpretowane przez rejestrator zdalny, c) sytuacja niespodziewana wymagająca ciągłego raportowania, o zakończeniu decyduje serwer

3.3. Adaptacyjny format danych

Dane (informacje medyczne) podczas interpretacji zapisu EKG istnieją w postaci sygnału, rezultatów diagnostycznych i rezultatów pośrednich (meta dane). Rozproszona architektura oprogramowania interpretacyjnego z adaptacyjnym podziałem ról wymaga przesyłania informacji o zróżnicowanym stopniu przetworzenia [1]. Zadanie to najłatwiej zrealizować, stosując adaptacyjny format danych złożony z niewielkiej objętościowo obowiązkowej części nagłówkowej i zróżnicowanych objętościowo opcjonalnych kontenerów informacyjnych (rys. 5). W części obowiązkowej przesyłane są symbole typu i rozmiaru danych zawartych w poszczególnych kontenerach, a ponadto długość okresu raportowania, identyfikator i podstawowe informacje statusowe rejestratora. Rozszerzone informacje statusowe (np. współrzędne geograficzne położenia, historia obciążenia procesora i poboru energii) mogą się znaleźć w opcjonalnym kontenerze statusowym. Odrębny kontener przewidziano także do komunikacji z pacjentem i może on przechowywać informacje tekstowe, graficzne (np. zdjęcie) lub dźwiękowe.



Rys. 5. Schemat adaptacyjnego formatu raportu

Biorąc pod uwagę założenie optymalnego wykorzystania kanału telekomunikacyjnego, należy zauważyć, że znaczne rozbudowanie formatu raportu łatwo spowoduje wypełnienie całego pasma oferowanego przez kanał (np. GPRS – 56kb/s), ograniczając dostępną częstotliwość raportowania i uniemożliwiając raportowanie ciągłe. Tak więc rozbudowując format danych o kolejne typy kontenerów (zwłaszcza multimedialnych, o dużej objętości), należy zadbać o ograniczenia częstotliwości ich użycia.

4. Dyskusja i wnioski

Projekt rozproszonego systemu nadzoru kardiologicznego jest inspirowany niedoskonałościami dostrzeżonymi w sieciach nadzoru wdrożonych we wczesnych latach XXI wieku. Przetwarzanie i wymiana informacji w tych sieciach odbywa się w sposób jednorodny, co wyklucza adaptację algorytmu interpretacji do pacjenta, a także optymalne użycie kosztownych połączeń. Dodatkowym wyzwaniem była generalizacja i próba przeniesienia na sieć urządzeń telemedycznych relacji i ról wypracowanych wśród kardiologów w długiej historii medycyny.

W początkowym okresie prac projektowych wyłączono niektóre zagadnienia do odrębnego opracowania:

- współpraca serwera nadzorującego z kardiologiem – ekspertem,
- wielowartościowa ocena ryzyka niepoprawnej diagnozy przez rejestrator zdalny,
- oszacowanie zapasu mocy obliczeniowej rejestratora zdalnego przy podejmowaniu decyzji o uaktualnieniu jego wiedzy.

Niektóre inne zagadnienia rozwiązano w sposób uproszczony, zapewniając ograniczoną funkcjonalność sieci. Zagadnienia te wytyczają jednocześnie kierunek prac badawczych w najbliższej przyszłości:

- określanie interwału raportowania na podstawie historii rezultatów diagnostycznych,
- współpraca i wzajemna kompatybilność procedur interpretacyjnych zaimplementowanych w rejestratorze zdalnym i w serwerze,
- automatyczne poszukiwanie przez rejestrator zdalny optymalnego kanału komunikacji.

Obecny stan zaawansowania testów prototypowej wersji zdalnego rejestratora interpretującego i serwera nadzorującego pozwalają stwierdzić, że:

- automatyczny podział ról przy interpretacji zapisu EKG powoduje nieznaczny przyrost objętości przesyłanych

danych w porównaniu do interpretacji tego sygnału przez rejestrator zdalny,
- ograniczenie mocy obliczeniowej ze strony rejestratora zdalnego nieznacznie obniża możliwości i wiarygodność diagnozy, głównie za przyczyną opóźnień,
- w przypadkach wątpliwych możliwa jest weryfikacja diagnozy poprzez powtórną analizę przesłanego na żądanie sygnału EKG przez serwer nadzorujący lub eksperta – kardiologa.

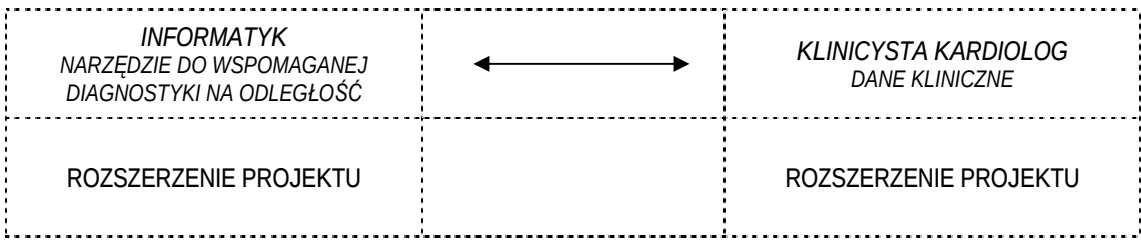
Opisywany system nadzoru kardiologicznego jest obecnie testowany w wersji prototypowej, ale można już jednoznacznie stwierdzić, że interpretacja zapisu EKG w sposób adaptacyjny w systemie rozproszonym łączy najistotniejsze zalety i wyklucza podstawowe mankamenty dotychczas wdrożonych sieci telediagnostycznych.

Podziękowania

Badania naukowe sfinansowane przez Komitet Badań Naukowych w latach 2004-2007 ramach grantu 3T11E 00127.

Literatura cytowana

1. Augustyniak P.: Optimizing the machine description of electrocardiogram, Journal of Medical Informatics and Technology vol. 8, pp. MM-41-MM48, 2004.
2. Bousseljot R. et al.: Telemetric ECG Diagnosis Follow-Up, Computers in Cardiology, 30, pp. 121-124, 2003.
3. Chiarugi F. et al.: Real-time Cardiac Monitoring over a Regional Health Network: Preliminary Results from Initial Field Testing, Computers in Cardiology, 29, pp. 347-350, 2002.
4. Chiarugi F. et al.: Continuous ECG Monitoring in the Management of Pre-Hospital Health Emergencies, Computers in Cardiology, 30, pp. 205-208, 2003.
5. Gouaux F. et al.: Ambient Intelligence and Pervasive Systems for the Monitoring of Citizens at Cardiac Risk: New Solutions from the EPI-MEDICS Project' Computers in Cardiology, 29, pp. 289-292, 2002.
6. Khoór S. et al.: Internet-Based, GPRS, Long-Term ECG Monitoring and Non-Linear Heart-Rate Analysis for Cardiovascular Telemedicine Management, Computers in Cardiology, 30, pp. 209-212, 2003.
7. Maglaveras N. et al.: Using Contact Centers in Telemanagement and Home Care of Congestive Heart Failure Patients: The CHS Experience, Computers in Cardiology, 29, pp. 281-284, 2002.
8. Nelwan S. P., van Dam T. B., Klootwijk P., Meil S. H.: Ubiquitous Mobile Access to Real-time Patient Monitoring Data, Computers in Cardiology, 29, pp. 557-560, 2002.
9. Pinna GD. et al.: Home Telemonitoring of Chronic Heart Failure Patients: Novel System Architecture of the Home or Hospital in Heart Failure Study, Computers in Cardiology 30, pp. 105-108, 2003.
10. Tadeusiewicz R.: Automatic Understanding of Signals. In: Kłopotek M. A., Wierzchoń S. T., Trojanowski K. (Eds.): Intelligent Information Processing and Web Mining. Springer, pp. 577-590, 2004.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 355-359.

USPRAWNIENIE PRZEWLEKŁEJ OPIEKI NAD CHORYMI NA ASTMĘ OSKRZELOWĄ PRZY POMOCY NARZĘDZI TELEINFORMATYCZNYCH

IMPROVEMENT OF CHRONIC CARE FOR PATIENTS WITH BRONCHIAL ASTHMA WITH THE USE OF TELEINFORMATIC TOOLS

MARIUSZ DUPLAGA, GRAŻYNA BOCHENEK, ŁUKASZ KASPER, MAGDALENA TWARDOWSKA, EWA NIŻANKOWSKA-MOGILNICKA, ANDRZEJ SZCZEKLIK

II Katedra Chorób Wewnętrznych Collegium Medicum Uniwersytetu Jagiellońskiego, Kraków, mmduplag@cyf-kr.edu.pl

Streszczenie. Internetowy system wspomagający opiekę nad chorymi na astmę oskrzelową o ciężkim przebiegu został opracowany w ramach projektu finansowanego przez Ministerstwo Zdrowia i Komitet Badań Naukowych w latach 2001-2003. Podstawowym celem projektu było usprawnienie opieki nad tą grupą chorych na astmę, u których przebieg choroby jest szczególnie ciężki i trudny do leczenia. Wybór grupy chorych z ciężką astmą jako docelowej populacji objętej projektem wynikał także z faktu znacznych nakładów kierowanych na opiekę nad nią we współczesnych systemach ochrony zdrowia, przekraczających nawet 70% wszystkich środków przeznaczonych na opiekę nad chorymi na astmę (pomimo stosunkowo małej liczebności grupy chorych na astmę, nie przekraczającej 5-10% populacji wszystkich chorych na astmę). Proces usprawnienia opieki nad chorymi na ciężką astmę zrealizowany w ramach projektu opierał się na utworzeniu sieci ośrodków referencyjnych zapewniających kompetentną opiekę chorym na astmę o ciężkim przebiegu, opracowaniu i wdrożeniu wytycznych określających organizacyjne i medyczne aspekty opieki oraz przygotowaniu internetowego systemu przyczyniającego się do zapewnienia ciągłości opieki i włączenia chorych w proces leczenia i monitorowania przebiegu choroby. System umożliwia rejestrację chorych spełniających kryteria ciężkiej astmy; wprowadzenie wywiadu dotyczącego przebiegu choroby, informacji o leczeniu i stanie pacjenta w trakcie kolejnych wizyt, a także wyników samoobserwacji prowadzonych przez chorych. Funkcja samodzielnego monitorowania objawów choroby jest realizowana poprzez indywidualizowany dostęp pacjenta do systemu, z opcjami wprowadzania wyników samoobserwacji i samodzielnych pomiarów poprzez formularze dostępne w standardowej przeglądarce internetowej lub w postaci SMSów przesyłanych z telefonu komórkowego używanego przez pacjenta. Dane wpisywane regularnie do systemu (codziennie, co tydzień) stają się dostępne dla lekarzy zaangażowanych w opiekę nad chorym w instytucjach pozostających na różnych poziomach referencyjności (opieka podstawowa, poradnia alergologiczna lub pulmonologiczna, ośrodek referencyjny). W nagłych przypadkach pacjent może zapewnić dostęp do swoich danych w systemie lekarzowi, do którego trafi z powodu zaostrzenia dolegliwości. Wdrożenie systemu było realizowane w 2004 roku, a obecnie planowane jest jego rozszerzenie na kolejne placówki.

Słowa kluczowe: astma oskrzelowa, opieka przewlekła, telemonitorowanie, medycyna mobilna

Abstract. An Internet-based system supporting care delivered to patients with severe bronchial asthma was developed within a project financed by Ministry of Health and State Committee for Scientific Research in years 2001-2003. The main objective of the project was the improvement of care in the population of severe asthma patients with. The focus on patients with severe asthma resulted from the fact that this relatively small group of patients (not more than 5-10% of all patients with bronchial asthma) consumes even 70% of health care resources assigned to asthma care. The project envisaged the improvement of severe asthma care through establishment of a network of referential-centre focused, implementations of guidelines related to organizational and medical aspects of care in severe asthma as well as development of the Internet-based system supporting asthma care with mechanisms maintaining shared care and patient's empowerment in management plan. The system enables registration of the patients fulfilling the criteria for severe asthma, input of patient's medical history, information on medications and patient's status during control visit as well as self-monitoring data. Self-monitoring option was implemented through the individualized patient's access to the system and options for input of self-observations and self-measurements carried out by the patients through Internet forms available through standard Internet browser or text messages (SMS) sent from cellular phone. Data entered by the patients on regular basis (daily, weekly) are available for physicians involved in patient care at various levels of referentiality (primary care, asthma policlinics, referential centre). In case of emergency state, the patient may grant access to his or her data contained in the database of the online system to the health professional at the point of care. The system was implemented throughout 2004 and currently its extension into further referential and other asthma centres is continued.

Key words: bronchial asthma, chronic care, telemonitoring, mobile medicine

1. Problem skutecznego leczenia i monitorowania chorób przewlekłych

Zapewnienie adekwatnej opieki medycznej pacjentom z chorobami przewlekłymi jest jednym z największych wyzwań stojących przed współczesnymi systemami ochrony zdrowia [8]. Liczba chorych cierpiących na przewlekłe schorzenia ciągle się zwiększa, rośnie też presja, aby zagwarantować im pełny udział w aktywnościach społecznych. Prowadzi to do poszukiwania takich rozwiązań organizacyjnych i technicznych, które przyczynią się do poprawy sprawności systemu ochrony zdrowia. Z dotychczasowych doświadczeń wynika, że istotnym elementem sprzyjającym skuteczności działań medycznych jest aktywny udział pacjenta z przewlekłymi dolegliwościami w procesie monitorowania objawów i leczenia [7]. Ponadto wymóg zapewnienia ciągłości opieki, stawiany przed systemami ochrony zdrowia, nabiera w przypadku chorych z przewlekłymi schorzeniami szczególnie ważnego znaczenia.

Przewlekłe choroby układu oddechowego prowadzące do obturacji w drogach oddechowych, głównie astma oskrzelowa i przewlekła obturacyjna choroba płuc, przyczyniają się znacząco do rosnących wydatków na opiekę zdrowotną. Charakter zjawisk patofizjologicznych występujących u chorych na te schorzenia doprowadza do nawracających lub uporczywych objawów klinicznych i konieczności planowania długotrwałego postępowania leczniczego.

Zapadalność na astmę oskrzelową we współczesnych społeczeństwach jest coraz większa. Potrzeba wszechstronnego podejścia do diagnostyki i leczenia astmy oskrzelowej z uwzględnieniem jej skutków społecznych i ekonomicznych znalazła odzwierciedlenie w międzynarodowych wytycznych opublikowanych przez Global Initiative for Asthma (GINA) w 1993 r. i regularnie uaktualnianych [2]. W wytycznych tych zwrócono szczególną uwagę na potrzebę zaangażowania pacjenta w długotrwały proces monitorowania dolegliwości i modyfikacji leczenia. Z dostępnych wyników badań wnioskujemy bowiem, że regularne pomiary czynności płuc oraz samodzielna ocena nasilenia objawów znacząco przyczyniają się do skuteczności leczenia u chorych na astmę oskrzelową. W tym kontekście wykorzystanie narzędzi teleinformatycznych może przyczynić się do usprawnienia procesu monitorowania choroby poprzez zagwarantowanie lepszej współpracy ze strony chorego.

2. Wykorzystanie technologii teleinformatycznych w opiece przewlekłej

Znaczący postęp, jaki dokonał się w zakresie telematyki medycznej, zwłaszcza w kontekście dążenia do poprawy jakości usług medycznych i usprawnienia monitorowania stanu pacjenta, otworzył także nowe możliwości dla rozwoju strategii postępowania w chorobach przewlekłych [1], [4]. Aplikacje internetowe wspomagające leczenie chorób przewlekłych mogą być wykorzystywane przez różnych użytkowników i w różnych scenariuszach. Istnieją rozwiązania adresowane do pacjentów, wspomagające lekarzy albo też zapewniające platformę komunikacji pomiędzy pacjentem a pracownikiem opieki zdrowotnej [5].

Zastosowanie aplikacji opartych o wykorzystanie Internetu w opiece przewlekłej może nieść ze sobą wiele korzyści [3]. Okazuje się bowiem, że stosunkowo proste działania mogą poprawiać skuteczność opieki nad chorym, zwłaszcza, jeśli są one realizowane na szeroką skalę. Regularne prowadzenie samodzielnych obserwacji przez pacjenta i samodzielne pomiary czynności płuc przyczyniają się do lepszego opanowania objawów astmy. Wiadomo także, iż zastosowanie Internetu lub komunikacja przy pomocy modemu dla codziennej transmisji danych do ośrodka monitorującego zwiększa stosowanie się pacjenta do ustalonego trybu leczenia i przestrzegania zaleceń co do trybu życia. Automatyczna analiza danych przesyłanych przez chorych umożliwia wczesne wykrywanie niekorzystnych trendów w przebiegu choroby i generowanie alarmów, kierowanych następnie do pacjenta i jego lekarza. Istotnym aspektem jest także możliwość zagwarantowania dostępu do elektronicznej dokumentacji pacjenta każdemu lekarzowi sprawującemu nad chorym opiekę praktycznie w dowolnym miejscu. Oczywiste są także możliwości edukacji pacjenta za pośrednictwem informacji udostępnianych na stronach internetowych, dostępne dla niego w dowolnym czasie, bez ograniczeń czasowych, jakie charakteryzują kontakt z lekarzem.

Coraz szersze wykorzystywanie technologii teleinformatycznych w ochronie zdrowia przyczyni się być może w przyszłości do wytworzenia uniwersalnego środowiska wspierającego opiekę medyczną poprzez automatyczne wykrywanie niekorzystnych trendów u chorych, zapewnienie wszechstronnych interakcji pomiędzy pacjentami i pracownikami ochrony zdrowia i wreszcie zanurzenie pacjenta w wirtualnej wspierającej go na co dzień społeczności.

3. System opieki nad chorymi na ciężką astmę

3.1. Zakres projektu

Projekt celowy mający na celu usprawnienie opieki nad chorymi na ciężką astmę był koordynowany przez II Katedrę Chorób Wewnętrznych CMUJ i realizowany na zlecenie Ministerstwa Zdrowia i Komitetu Badań Naukowych w latach 2001 – 2003. W realizacji zadań projektu uczestniczyła sieć 6 ośrodków referencyjnych zlokalizowanych w głównych aglomeracjach miejskich w Polsce. Od 2004 r. wdrażane są wyniki tego projektu. Podstawowe zadania zrealizowane w jego trakcie obejmowały: przygotowanie wytycznych dotyczących postępowania z chorymi na ciężką astmę z uwzględnieniem hierarchii powiązań pomiędzy placówkami leczącymi chorych na astmę, doposażenie ośrodków referencyjnych uczestniczących w projekcie w sprzęt diagnostyczny i terapeutyczny oraz opracowanie aplikacji internetowej umożliwiającej dostęp do danych pacjenta na różnych poziomach opieki z opcją długotrwałego monitorowania nasilenia choroby u chorego.

Główną przesłanką dla objęcia projektem populacji chorych na ciężką astmę był fakt, że ta stosunkowo nieduża grupa pacjentów wykorzystuje większą część zasobów systemu ochrony zdrowia przeznaczonych na leczenie astmy oskrzelowej. Mimo to jakość opieki gwarantowana chorym na astmę o ciężkim i trudnym do leczenia przebiegu bywa często niewystarczająca, a pacjenci trafiają niejednokrotnie do placówek ochrony zdrowia w stanie ciężkiego zaostrzenia dolegliwości, a nawet zagrożenia życia.

Podstawowym wymogiem dla systemu informacyjnego przygotowywanego dla potrzeb projektu było zagwarantowanie modelu współdzielonej opieki medycznej poprzez dostęp do elektronicznej dokumentacji chorego na różnych poziomach referencyjnych. Ponadto, ponieważ pacjenci z ciężkimi postaciami astmy wymagają starannego kontrolowania ciężkości choroby, jedną z istotnych funkcji systemu miała być możliwość monitorowania objawów choroby poprzez wykorzystanie formularzy internetowych lub urządzeń mobilnych dostępnych choremu.

Wybór narzędzi informatycznych wykorzystanych do stworzenia systemu zakładał zachowanie jak najlepszego stosunku kosztu do uzyskanej funkcjonalności zarówno na etapie pozyskiwania technologii, jak i następnie utrzymania systemu. Pozostawało to w oczywistym związku z finansowymi dylematami placówek medycznych.

W czasie prac projektowych zdefiniowany został zakres informacji gromadzonych w bazie danych o pacjentach z ciężką astmą. Po określeniu zakresu bazy danych oraz funkcjonalności systemu przygotowany został prototyp aplikacji internetowej. Został on poddany testom w ośrodkach referencyjnych współpracujących w projekcie. Robocza wersja systemu była używana w trakcie projektu celowego i przed jego zakończeniem zostały wprowadzone kolejne poprawki wynikające z doświadczeń użytkowników. W fazie pilotażowej do grupy użytkowników systemu włączono także pierwszych pacjentów: część z nich używała opcji telemonitorowania dostępnej poprzez formularze internetowe [6], natomiast druga grupa testowała przydatność SMSów do komunikacji z systemem.

Od 2004 roku rozpoczął się program wdrożenia wyników projektu celowego. Rozszerzona została liczba ośrodków współpracujących z ośrodkami referencyjnymi. Prowadzono rejestrację chorych, część z nich korzystała z funkcji telemonitorowania dostępnych w systemie przez kilkanaście miesięcy.

Rys. 1. Strona główna internetowego systemu dla chorych na ciężką astmę

3.2. Struktura systemu

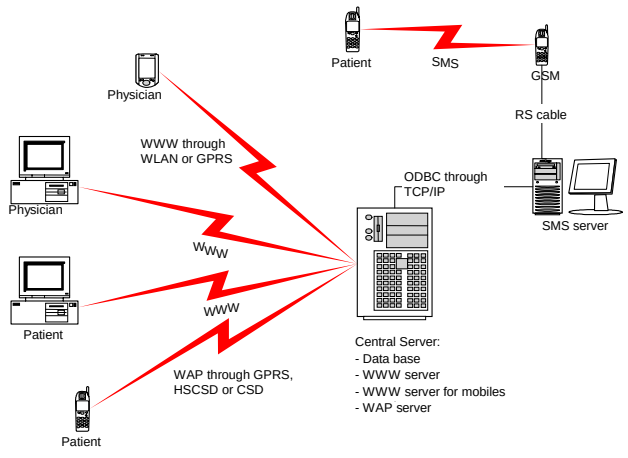
Jądro systemu ciężkiej astmy stanowi baza danych z DBMS (systemem zarządzania bazą) Postgres. Baza danych zawiera wszystkie informacje niezbędne do funkcjonowania systemu. Dostęp do bazy danych systemu jest realizowany przez:

- Serwis WWW przeznaczony dla komputerów stacjonarnych – serwer WWW, umożliwiający użytkownikom systemu dostęp do zasobów bazy danych,
- Serwis WWW dla komputerów przenośnych i palmtopów,
- Serwis WAP – wydzielona część serwisu WWW jest przeznaczona wyłącznie dla pacjentów wysyłających dane za pomocą przeglądarek WAP wbudowanych w telefon GSM,
- Usługę SMS – dostęp za pomocą krótkich wiadomości tekstowych jest możliwy dzięki zewnętrznemu systemowi przetwarzania SMS, działającemu na innej platformie sprzętowej i programowej.

Rdzeń systemu stanowią:

- baza danych oparta na systemie zarządzania bazą danych Postgres,
- serwer WWW Apache funkcjonujący w systemie operacyjnym Debian (http://www.debian.org/). System operacyjny został wybrany ze względu na stosunkowo dużą niezawodność, możliwość obsługi wybranej architektury sprzętowej, bardzo dobrą integrację pakietów oprogramowania oraz bardzo wysoki poziom bezpieczeństwa.

Serwis WWW Systemu zbudowano na bazie serwera WWW rodziny Apache z wbudowaną obsługą języka skryptowego PHP, wykorzystując dodatkowo moduły programowe SMARTY, PEAR oraz AGORA udostępniane na zasadach GPL (GNU Public Licence).



Rys. 2. Struktura systemu internetowego dla celów projektu usprawnienia opieki nad chorymi na astmę o ciężkim przebiegu

Podstawowym elementem serwisu, łączącym w całość poszczególne moduły jest biblioteka klas PMLIB, stworzona specjalnie na potrzeby Systemu Ciężkiej Astmy. Dzięki wykorzystaniu generatora układu stron (*template engine*) SMARTY dokonano rozdzielenia między kodem PHP (realizującym dostęp do bazy danych i odpowiadającym za mechanizmy autoryzacji) a kodem HTML (odpowiadającym za szatę graficzną stron WWW). Dostęp do serwera bazy danych Postgres SQL realizowany jest za pomocą biblioteki klas PEAR, w znaczny sposób ułatwiającej dostęp do danych. Forum dyskusyjne lekarzy i pacjentów zrealizowano w oparciu o oprogramowanie firmy AGORA.

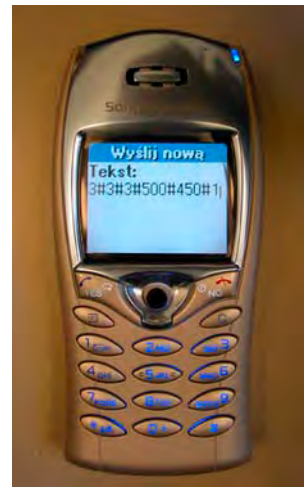
Dostęp do systemu jest realizowany za pomocą protokołu HTTP (Hypertext Transfer Protocol). Do używania systemu wystarczy zwykła przeglądarka WWW, np. Internet Explorer, Netscape lub Mozilla Firefox. Ponieważ protokół HTTP nie zapewnia wystarczającego poziomu zabezpieczeń i szyfrowania przesyłanych danych, dane są szyfrowane za pomocą protokołu SSL (Secure Sockets Layer).

3.3. Funkcjonalność systemu

Informacje gromadzone w bazie danych systemu ciężkiej astmy obejmują wywiad dotyczący objawów astmy, czasu ich wystąpienia, charakteru, nasilenia, czynników zaostrzających, występowania atopii, a także wywiad dotyczący występowania astmy i chorób alergicznych w rodzinie. Ponadto system umożliwia wprowadzanie danych o leczeniu i wcześniejszych kontaktach z placówkami ochrony zdrowia. Istotnym elementem gromadzonych informacji są wyniki samodzielnie prowadzonych przez chorych obserwacji nasilenia objawów oraz samodzielnych pomiarów szczytowego przepływu wydechowego. Po wstępnej rejestracji pacjenta w systemie, która musi nastąpić w ośrodku referencyjnym, elektroniczna dokumentacja pacjenta może być udostępniona innym pracownikom ochrony zdrowia sprawującym opiekę nad chorym, o ile wyrazi on na to zgodę.

Dostęp do systemu jest możliwy tylko dla tych użytkowników, którzy są w stanie wprowadzić właściwy, indywidualny login i hasło. Pracownicy ochrony zdrowia, którzy uzyskali dostęp do systemu, mają zróżnicowany zakres dostępu. Lekarz-konsultant może dokonać rejestracji nowego chorego z ciężką astmą (zgodnie z kryteriami opracowanymi dla potrzeb projektu w fazie pilotażowej) i wprowadzić informacje o lekarzu prowadzącym. Lekarze-konsultanci są zwykle wyznaczani przez ośrodki referencyjne uczestniczące w projekcie i rejestrowani w systemie przez Ośrodek Koordynujący.

W czasie kolejnych wizyt w gabinecie lekarskim lekarz prowadzący może wprowadzić informacje o stanie pacjenta i zleconych lekach. Lekarz może także przeglądać dane z wywiadu wprowadzone w czasie pierwszej wizyty oraz wyniki samodzielnych obserwacji i pomiarów dokonywanych przez pacjenta pomiędzy wizytami.

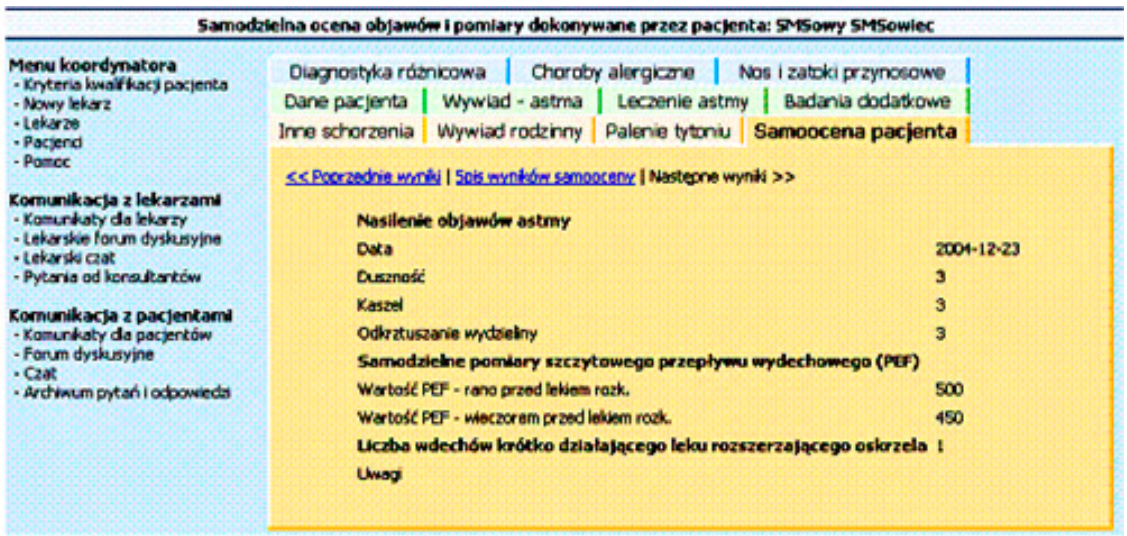


Rys. 3. Format SMSu z danymi z samoobserwacji przesyłanymi przez pacjenta do systemu

Pacjenci zarejestrowani w systemie mogą wprowadzać wyniki samodzielnych obserwacji i pomiarów PEFR codziennie lub rzadziej, zgodnie z planem ustalonym z lekarzem prowadzącym. Dane wprowadzane w panelu samoobserwacji obejmują (rys. 4) ocenę nasilenia 3 typów objawów wyrażoną w 10-punktowej skali Borga, wyniki pomiarów szczytowego przepływu wydechowego oraz liczbę wdechów krótko działającego beta-mimetyku.

Pacjent może wprowadzać te dane poprzez wypełnienie formularza internetowego (dostępnego w przeglądarce internetowej po zalogowaniu do systemu) albo jako SMS z telefonu komórkowego (rys. 3).

System informatyczny przygotowany dla potrzeb projektu może być także traktowany jako platforma komunikacji zarówno dla lekarzy pozostających w placówkach na różnych poziomach referencyjnych, opiekujących się pacjentami z astmą, jak i pomiędzy chorymi z ciężką astmą i ich lekarzami prowadzącymi.



Rys. 4. Formularz dostępny w systemie prezentujący wyniki z samoobserwacji wprowadzone przy pomocy SMSu (patrz rys. 3)

3.4. Wdrożenie

Obecnie w systemie zarejestrowanych jest ok. 120 pacjentów. Wśród nich 15 korzysta z dostępnych w systemie funkcji wprowadzania wyników samoobserwacji i samodzielnych pomiarów. Planowane jest rozszerzenie liczby ośrodków referencyjnych uczestniczących w programie. Większość chorych prowadzących samoobserwacje przy pomocy systemu, wykorzystuje do tego celu formularze internetowe. Kilku chorych bierze udział w testowaniu SMSów do przesyłania wyników samoobserwacji.

Na podstawie dotychczasowych doświadczeń ogólnie stwierdzić można, że pacjenci z ciężką astmą wykazują duże zainteresowanie udziałem w prowadzeniu samodzielnych obserwacji. Problemem jest jednak ograniczona dostępność Internetu w polskiej populacji, co wpływa na zjawisko „cyfrowego wykluczenia" znaczącej grupy chorych z możliwości używania systemu.

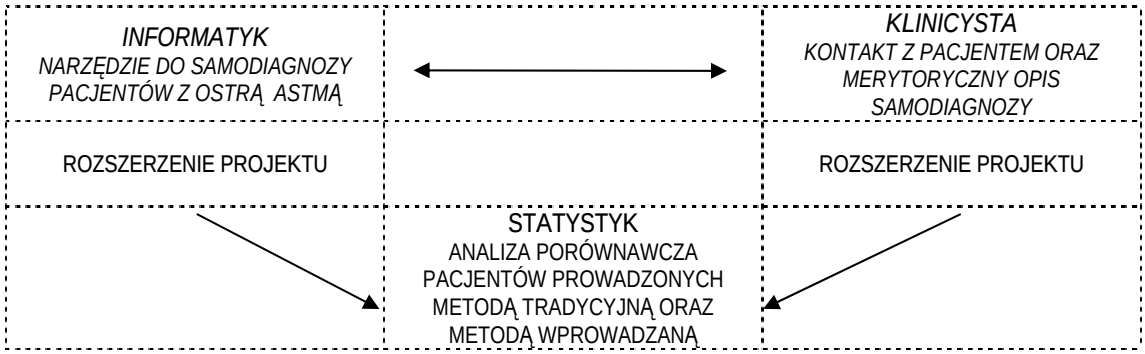
4. Podsumowanie

Skuteczne leczenie chorych na astmę oskrzelową wymaga ich regularnych kontaktów z systemem ochrony zdrowia. Istotnym elementem strategii terapeutycznej jest zaangażowanie pacjenta w proces monitorowania objawów. Ponieważ choroby przewlekłe, w tym także astma oskrzelowa, mają istotne skutki społeczne i finansowe, podejmowane są szczególne wysiłki w celu wypracowania spójnych strategii zapobiegania tymże chorobom, rozpoznawania i leczenia. Wytyczne publikowane przez międzynarodowe grupy ekspertów mogą być podstawą opracowania planów dla indywidualnego pacjenta. Zastosowanie rozwiązań teleinformatycznych dla realizacji tych strategii może przyczynić się do ich większej skuteczności.

W programie usprawnienia opieki nad chorymi na ciężkie postaci astmy oskrzelowej realizowanym na zlecenie Ministerstwa Zdrowia i Komitetu Badań Naukowych uwzględniono wymogi istotne z punktu widzenia nowoczesnego modelu opieki w chorobach przewlekłych, tj. współudział w opiece nad chorymi pracowników ochrony zdrowia na różnych poziomach referencyjnych oraz włączenie pacjenta w proces monitorowania i leczenia choroby.

Literatura cytowana

1. Gillespie G.: Deploying an I. T. cure for chronic disease, Health Data Manag, 8(7): 68-74, 2000.
2. Global Initiative for Asthma GINA Website: http://www.ginasthma.com/
3. Gomaa W. H., Morrow T., Muntendam P., Smith G.: Technology-based disease management: A low-cost, high-value solution for the management of chronic disease, Dis Manag Health Outcomes, 9(10): 577-588, 2001.
4. Joch A.: Can the Web Save Disease Management? Healthcare Informatics March 2000 [http://www.healthcare-informatics.com/issues/2000/03_00/cover.htm]
5. Regan M.: Utilizing an Intranet in Disease Management. Dis Manag Health Outcomes, 10(3): 147-154, 2002.
6. Strona głowna systemu ciężkiej astmy http://www.astma.med.pl]
7. Tattersall R.: The expert patient: a new approach to chronic disease management for the twenty-first century, Clin Med, 2(3): 227-229, 2002.
8. World Health Organization "Innovative Care for Chronic Conditions: Building Blocks for Action. Global Report." 2002.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 361-364.

TELEKONSULTACJE CHIRURGICZNE NA OBECNYM ETAPIE ROZWOJU TELEINFORMATYKI MEDYCZNEJ W POLSCE NA PRZYKŁADZIE REGIONU MAŁOPOLSKI

SURGICAL TELECONSULTING IN POLAND ANALYZED FOR THE MAŁOPOLSKA REGION

JANUSZ LEGUTKO, EWA WOŹNIAKIEWICZ, STANISŁAW GRZESIK, MIROSŁAW SZURA

I Katedra Chirurgii Ogólnej i Klinika Chirurgii Gastroenetrologicznej Collegium Medicum Uniwersytetu Jagiellońskiego, Kraków, ewozniak@cm-uj.krakow.pl
Kierownik: Prof. dr hab. med. Jan Kulig

Streszczenie. Światowy rozwój telematyki i jej zastosowanie w wielu dziedzinach medycyny powoduje konieczność tworzenia w Polsce systemów telekonsultacji medycznych ukierunkowanych na diagnozowanie i leczenie chorób nowotworowych. W takcie prac badawczych stwierdzono, że istnieje zapotrzebowanie wśród polskich chirurgów na dostęp do Internetu i jego wykorzystanie w praktyce chirurgicznej. Nie mniej jednak wiedza o telemedycynie i telechirurgii w Polsce jest jeszcze mała. W dużym stopniu spowodowane jest to ograniczeniami wywołanymi brakiem odpowiedniej infrastruktury, niezbędnej do przeprowadzania telekonsultacji w prawie wszystkich naszych szpitalach.

Słowa kluczowe: chirurgia przewodu pokarmowego, telemedycyna, telekonsultacje

Abstract. The worldwide development of telematics and of its applications in various fields of medicine makes it necessary to create systems of medical teleconsultations for diagnosing and treating cancer in Poland. The research showed that there is a need among Polish surgeons for access to the internet and its use in their work. However, the knowledge telemedicine and telesurgery is still limited in Poland. It is mostly due to the lack of appropriate infrastructure required for teleconsultations in almost all our hospitals.

Key words: GI-tract surgery, telemedicine, teleconsulting

1. Wstęp

Rozwój telekomunikacji, jaki dokonał się w ostatniej dekadzie, otwiera nowe możliwości przesyłania danych i obrazów na odległość. Są one już wykorzystywane we współczesnej medycynie, co znacznie skraca czas od postawienia rozpoznania do rozpoczęcia optymalnego leczenia. W ostatnich latach rośnie częstość zachorowań na choroby przewodu pokarmowego, ze szczególnym uwzględnieniem nowotworów, wzrasta więc ilość przypadków wymagających doraźnej konsultacji specjalistycznej. Szczególnie w chirurgii konsultacje telemedyczne to skrócenie czasu, w jakim podjęta zostanie decyzja o rodzaju zastosowanego leczenia i rodzaju wykonywanego zabiegu operacyjnego. Czas ten ma ogromne znaczenie dla rokowania i ostatecznych wyników leczenia raka żołądka i innych nowotworowych i nienowotworowych chorób przewodu pokarmowego Można go znacznie skrócić, przesyłając dane o chorym i wyniki badań (w tym zdjęcia z badań obrazowych) do ośrodka referencyjnego dla potrzeb konsultacji. Istotne znaczenie ma także bezpośredni przekaz obrazu z zabiegu operacyjnego w celu zasięgnięcia opinii wybitnego specjalisty. W ten sposób dokonuje się wyboru najbardziej odpowiedniego postępowania chirurgicznego niezależnie od ośrodka, w którym chory jest operowany. Nie można też sobie wyobrazić rozwoju robotyki bez telechirurgii.

Systemy telekonsultacji medycznych są rozwijane w świecie już od początku lat 90-tych ubiegłego wieku. W roku 1995 w Stanach Zjednoczonych wprowadzono system telemedyczny w zakresie onkologii. W momencie zakończenia programu wdrożeniowego dokonano oceny kosztów przeprowadzanych konsultacji w roku 1995 i w roku 2000. Uwzględniono wszystkie koszty aparatury i technologii łącznie. W wyniku wprowadzenia systemu telemedycznego uzyskano redukcję kosztów ogólnych z 812$ w 1995 roku do 410$ w roku 2000. Koszty te są obecnie jeszcze mniejsze, ponieważ aparatura jest coraz tańsza, a ponadto znamiennie wzrosła liczba telekonsultacji [1]. Na Ukrainie w latach 2000-2003 wdrożono program telemedyczny dla konsultacji chorych z mnogimi urazami. Wykonano 210 telekonsultacji dotyczących mnogich urazów, przypadków ortopedycznych z zakresu onkologii kości i neurochirurgii. Aż u 80% chorych telekonsultacje spowodowały modyfikację postępowania leczniczego [2]. Podobny międzynarodowy program telemedyczny stworzono w 1998 roku w basenie Dunaju – od Adriatyku do Morza Czarnego [3]. Programy konsultacji telemedycznych wdrażano również w Nepalu [4], w daleko-

wschodniej Rosji [5], Bangladeszu [61], Południowej Afryce [7], Palestynie [8] i w Ekwadorze [9], a także w Indiach, gdzie dla celów telekonsultacji wykorzystywano połączenia satelitarne [10]. W Stanach Zjednoczonych czynione są starania o stworzenie światowej sieci telemedycznej. Powstało specjalne konsorcjum MITAC (Medical Informatics and Technology Applications Consortium), którego pierwszym dziełem był Yale-NASA Mount Everest Telemedicine Project, w ramach którego himalaiści wspinający się na szczyt Sagarmatha w Himalajach w zimie 1998 i 1999 roku byli pod stałym nadzorem centrum medycznego Uniwersytetu Yale [11].

Wykorzystanie telemedycyny w ochronie zdrowia w Polsce jest jeszcze sporadyczne i dotyczy przede wszystkim zadań z zakresu diagnostyki kardiologicznej i pulmunologicznej [12,13], a także interpretacji wyników tomografii komputerowej [14] oraz rezonansu magnetycznego [15].

2. Cel

Ocena możliwości przeprowadzenia telekonsultacji chirurgicznych śródoperacyjnych za pośrednictwem różnych mediów transmisyjnych.

Ustalenie potrzeb i możliwości technicznych dla przeprowadzania telekonsultacji w obecnym stanie infrastruktury teleinformatycznej i świadomości pracowników ochrony zdrowia.

3. Materiał i metodyka

Ocenę potrzeb i możliwości przeprowadzania telekonsultacji chirurgicznych dotyczących rozpoznawania i leczenia chirurgicznego chorób nowotworowych przewodu pokarmowego dokonano poprzez badania ankietowe. Badaniami tymi objęto 33 oddziały chirurgiczne Małopolski. Pytania ankietowe skierowane do chirurgów dotyczyły wykorzystywania Internetu przez lekarzy chirurgów, ich potrzeb w zakresie szkoleń i konsultacji oraz dostępności Internetu na oddziałach chirurgicznych.

Analiza ankiet obejmowała standardowe procedury statystyczne dla tego typu badań, a wyniki przedstawiono w formie łatwych do interpretacji zestawień procentowych.

Teletransmisje śródoperacyjne przeprowadzono w różnych odstępach czasowych z wykorzystaniem różnych mediów transmisyjnych: ISDN, połączenia internetowe, połączenia modemowe.

Po każdej teletransmisji przeprowadzano cenę jej jakości i przydatności do celów rutynowych konsultacji telechirurgicznych.

4. Wyniki

Pierwszą teletransmisję przeprowadzono w październiku 2001 r. Transmitowano wtedy zabieg operacyjny, podczas którego prof. dr hab. med. Jan Kulig dokonał wycięcia raka trzustki. Odbiorcami byli uczestnicy Targów Medycznych w Łodzi. Dwustronna łączność umożliwiła nie tylko przekaz obrazu, ale operator mógł wyjaśnić przebieg operacji oraz odpowiadać na zadawane pytania.

W styczniu 2002 roku przeprowadzono w oparciu o 4 dzierżawione linie ISDN międzynarodową telekonferencję pomiędzy I Katedrą Chirurgii CM UJ i ośrodkiem wiedeńskim, gdzie prof. Harald Rosen komentował zabieg operacyjny i służył jednocześnie radą w trudnych fragmentach operacji. W przerwie między dwoma operacjami założenia elektrod do sakralnej stymulacji u chorych z nietrzymaniem stolca prof. Harald Rosen wygłosił wykład poświęcony tej tematyce. Są to operacje wykonywane tylko w kilku ośrodkach w Europie, a Profesor jest uznanym autorytetem w tej dziedzinie w świecie. Po raz pierwszy nie był potrzebny osobisty jego udział w operacji w Krakowie.

W dniu 17.10.2002 r. poprzez złącza internetowe, wykorzystując protokół TCP/IP, przeprowadzono teletransmisję z I Katedry Chirurgii Ogólnej CM UJ na salę obrad VI Konferencji Internetu i Telematyki Medycznej z maksymalną prędkością 10 Mb/s. Nie była to wprawdzie jedyna transmisja, ale inni prelegenci demonstrowali tylko przekaz obrazu z badań kardiologicznych.

W dniu 27.02.2004 r. udostępniono poprzez Internet możliwość uczestniczenia w pokazach na żywo zabiegów operacyjnych wykonywanych w I Katedrze Chirurgii Ogólnej CM UJ przez zagranicznych profesorów w czasie 6-tego Międzynarodowego Sympozjum Ludwika Rydygiera.

Transmitowane w tym dniu przez Internet zabiegi to:

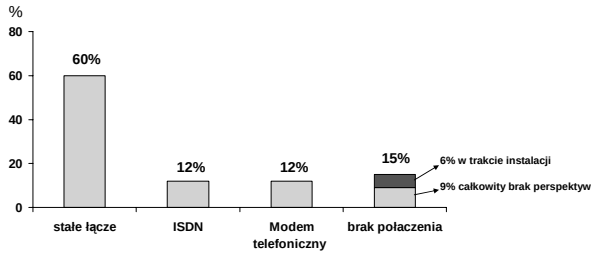
- Laparoscopic gastric cancer resection – operator prof. V. C. Huscher, asysta: dr Cz. Osuch, dr W. Milanowski,
- Laparoscopic TEM in rectal cancer – operator prof. V. J. Schmidt, asysta: dr hab. P. Richter, dr P. Szybiński,
- Laparoscopic adrenalectomy – operator prof. I. N. Basso, asysta: dr hab. P. Richter, dr P. Szybiński,
- Laparoscopic Collectomy – operator prof. V. C. Huscher, asysta: dr Cz. Osuch, dr W. Milanowski.

Ponadto w czasie wykonywania zabiegów na żywo prowadzone były konferencje pomiędzy Krakowem a AM w Szczecinie, w czasie których uczestnicy mogli nie tylko wysłuchać bieżącego komentarza operatora, ale również zadawać pytania. Telekonferencje przeprowadzono po złączach internetowych.

W dniu 13 maja 2004 r. przeprowadzono teletransmisję z I Katedry Chirurgii Ogólnej CM UJ na salę obrad III Krakowskich Warsztatów Inżynierii Medycznej, wykorzystując radiowe połączenia internetowe.

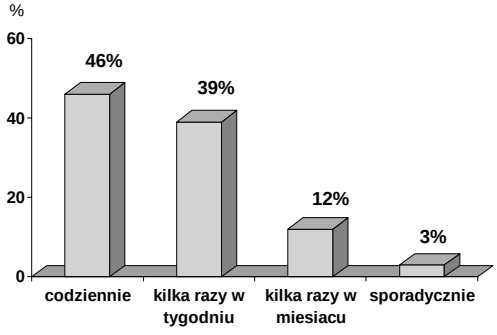
Nabyte doświadczenia dowodzą, że najlepszą jakość video i audio transmisji uzyskano poprzez łącza ISDN oraz łącza internetowe z transmisją o maksymalnej szybkości 100 Mb/s, przy czym wykorzystanie połączeń radiowych nie pogarszało jakości transmisji. Jedynym problemem do rozwiązania pozostaje jakość dźwięku, bowiem standardowe mikrofony, nawet o dużej czułości, w środowisku sali operacyjnej, gdzie wiele urządzeń wytwarza fale elektromagnetyczne, przewodzą szumy zniekształcające przekaz dźwięku.

W wyniku przeprowadzonej analizy ankiet stwierdzono, że jedynie 60% oddziałów chirurgicznych szpitali regionu Małopolski posiada stałe łącza z siecią internetową i tym samym możliwość szybkiej transmisji poprzez Internet. Wprawdzie za pomocą modemu ISDN lub telefonicznego może się łączyć z Internetem 24% szpitali objętych badaniem ankietowym, nie można jednak tych rozwiązań uważać za wystarczające dla potrzeb telekonferencji lub bezpośredniej transmisji zabiegu w celu konsultacji. Informacja o tym, że połączenia internetowe są w trakcie instalacji, dotyczyła 6% oddziałów chirurgicznych. Natomiast aż 9% ankietowanych palcówek nie posiada w ogóle połączenia z Internetem (rys.1).

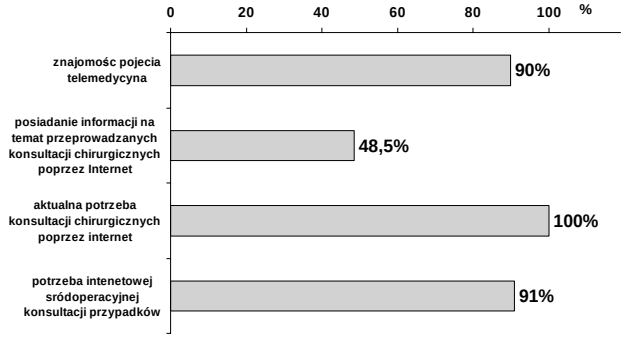


Rys. 1. Rodzaje połączenia z Internetem w 33 ośrodkach chirurgicznych Małopolski

Zapotrzebowanie na korzystanie z Internetu przez lekarzy chirurgów obrazują odpowiedzi dotyczące częstości z jego korzystania (rys. 2). Codziennie z Internetu korzysta aż 46% ankietowanych, a kilka razy w tygodniu 39%. Jedynie 15% ankietowanych łączy się z Internetem kilka razy w miesiącu lub rzadziej.



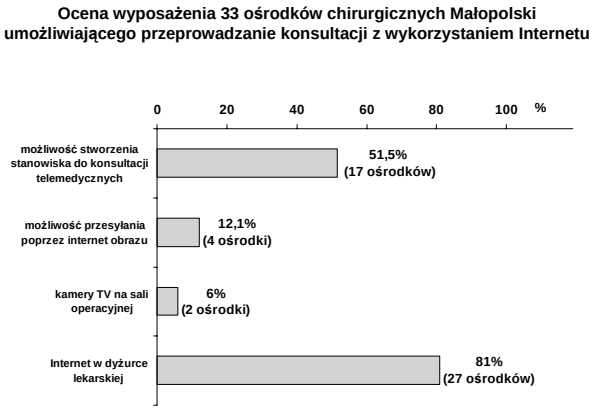
Rys. 2. Częstość korzystania z Internetu w środowisku lekarzy 33 badanych ośrodków chirurgicznych Małopolski



Rys. 3. Świadomość i potrzeby zastosowania telekonsultacji chirurgicznych w codziennej praktyce w środowisku lekarzy 33 badanych ośrodków chirurgicznych Małopolski

Kolejnym zagadnieniem, którego dotyczyły pytania, była telemedycyna. Odpowiedzi przedstawiono na rys. 3. Wprawdzie 90% ankietowanych chirurgów słyszało o telemedycynie, ale tylko 48.5% spośród nich posiada wiedzę na temat przeprowadzanych przez Internet konsultacji chirurgicznych. Wszyscy ankietowani widzą potrzebę konsultacji na odległość poprzez Internet, a aż 91% widzi zastosowanie konsultacji internetowych w czasie wykonywania trudnych zabiegów.

27 z 33 (81,4%) oddziałów chirurgicznych, z których lekarze byli ankietowani, posiada Internet w dyżurce lekarskiej. Jedynie 2 oddziały chirurgiczne (6%) posiada możliwość monitorowania zabiegu poprzez kamery telewizyjne. Sprzęt zapewniający możliwość przesyłania obrazów poprzez Internet posiadają jedynie 4 ośrodki chirurgiczne (12,1%). W 17 ośrodkach (51,5%) lekarze liczą na możliwość stworzenia stanowiska do konsultacji telemedycznych (rys.4).



Rys. 4. Ocena wyposażenia 33 ośrodków chirurgicznych Małopolski umożliwiającego przeprowadzenie telekonsultacji chirurgicznych z wykorzystaniem Internetu

Ankietowani lekarze najczęściej (w 81%) wykorzystują Internet do prowadzenia korespondencji elektronicznej. Równie liczna grupa lekarzy chirurgów na bieżąco śledzi serwisy medyczne (81% odpowiedzi). Do najczęściej wybieranych serwisów medycznych w kolejności należą: Medycyna Praktyczna, Medline, cmkp, gastrologia, medscape, onkologia, hernia, młody chirurg, trauma i Vesalius.

5. Podsumowanie

Rozwój Internetu spowodował olbrzymie zmiany w wielu sferach działalności ludzi, w tym także chirurgii [16]. Stanberry z Walii dokonał oceny telemedycyny w pierwszym dziesięcioleciu III-go millenium w aspekcie moralnym, etycznym i kulturowym oraz jej wpływu na lekarzy, pacjentów i społeczeństwo [17]. Niewątpliwie odgrywa ona coraz większą rolę w postępowaniu lekarzy z chorymi, ale z drugiej strony autor obawia się, że nowe technologie zmienią tradycyjne relacje między pacjentem i lekarzem. Decyzja o wyborze metod badania, stawianiu rozpoznania i wyborze metody leczenia będzie podejmowana w miejscu odległym od chorego.

Uzyskane w badaniu ankietowym wyniki wskazują na potrzebę stworzenia sieci telemedycznej dla potrzeb chirurgii w regionie małopolskim. Z dużym zainteresowaniem spotkał się pomysł stworzenia portalu chirurgicznego umożliwiającego szybką konsultację przypadków wymagających natychmiastowego leczenia operacyjnego u chorych, których ogólny stan zdrowia nie pozawala na transport do ośrodka specjalistycznego. Konsultujący chirurg mógłby szybko otrzymać komplet informacji i wyników badań, w tym badań obrazowych, a następnie tą samą drogą przesłać wynik konsultacji z proponowanym postępowaniem leczniczym.

Niewątpliwie najlepszym rozwiązaniem byłaby możliwość przesyłania dynamicznych obrazów z pola operacyjnego w trakcie trwania zabiegu operacyjnego, bez konieczności znacznego przedłużania czasu trwania zabiegu. Oznacza to także uniknięcie w wielu przypadkach konieczności późniejszej reoperacji po przekazaniu do ośrodka specjalistycznego.

Wielu lekarzy w trakcie specjalizacji z zakresu chirurgii ogólnej zainteresowanych jest również szkoleniami specjalistycznymi organizowanymi *on-line*, bez konieczności wyjeżdżania z miejsca codziennej pracy na kursy teoretyczne.

Portal dla specjalizujących się w chirurgii może stanowić znakomite źródło wiedzy teoretycznej i praktycznej, gdy prezentowane wiadomości zobrazowane są filmami zarejestrowanymi w autentycznych przypadkach, pokazującymi drogę do ustalenia definitywnego rozpoznania i zastosowania optymalnej dla danego chorego metody leczenia operacyjnego. Demartines i wsp. ze Szwajcarii przeprowadzili 70 telekonferencji (50 wykładów i 271 konsultacji) pomiędzy ośrodkami chirurgicznymi w odstępach tygodniowych w ciągu 2 lat [18]. Uczestniczyli w nich lekarze z 6 szpitali uniwersyteckich w 4 krajach europejskich. Oceniono przydatność tego typu konferencji w porównaniu z bezpośrednimi spotkaniami na 95% w skali 100%.

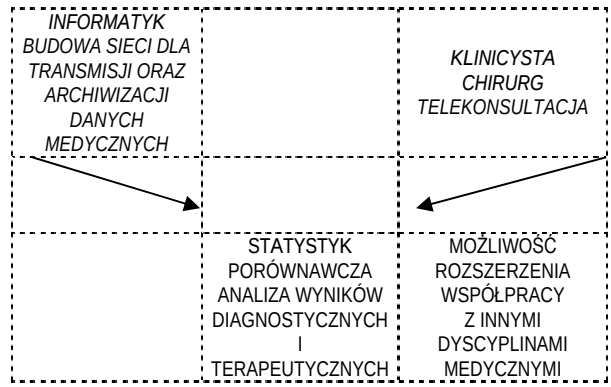
Należy pamiętać, że możliwość stworzenia sieci telemedycznej zależy nie tylko od zaangażowania poszczególnych ośrodków, ale również od rozwoju infrastruktury telekomunikacyjnej w regionie. Szczególnie dotyczy to możliwości taniego przyłączenia do sieci przy użyciu łączy dających możliwość szybkiego przesłania dużej ilości danych. Ciekawy sposób rozwiązania problemów z brakiem infrastruktury telekomunikacyjnej przedstawili Rosser i wsp. z Uniwersytetu Yale w USA [19]. Zaproponowali oni sposób rozwiązania problemów z odpowiednim wyposażeniem wymaganym w telemedycynie. Docierali oni do wiejskich terenów Ekwadoru z salą operacyjną na samochodzie wyposażonym w komputer i oprogramowanie do teletransmisji. Łączyli się następnie z Szpitalem Uniwersytetu Yale liniami telefonicznymi. Dzięki temu możliwe było przeprowadzanie telekonferencji i konsultacji przed- i śródoperacyjnych.

Przyszłość chirurgii jest uzależniona od wielu czynników, ale już teraz wiadomo, że zależeć będzie od rozwoju komputeryzacji, sieci komputerowych i robotyki. Schlag i wsp. z Berlina zwrócili uwagę na coraz większą liczbę specjalistów różnych dziedzin zaangażowanych w proces diagnozowania i leczenia chirurgicznego chorych [20]. Będzie to szczególnie istotne przy planowaniu w przyszłości leczenia chirurgicznego poprzez planowanie i przeprowadzanie wirtualnych symulacji zabiegu operacyjnego oraz konsultacji śródoperacyjnych. Maresceaux i wsp. z Francji opracowali komputerowy program integracyjny nazwany Computer Assisted Surgery (CAS) [21]. Umożliwia on tworzenie trójwymiarowych obrazów narządów z wirtualnej kolonoskopii i cholangiografii. Obrazy te służą do celów diagnostycznych i planowania zabiegu operacyjnego. W czasie operacji możliwa była konsultacja specjalisty z danej dziedziny. Natomiast po operacji, w celu monitorowania wyników leczenia i generalnie ewolucji choroby proponują stosowanie obrazów czterowymiarowych, czyli z czasem jako czwartym wymiarem, bowiem uzyskiwane w badaniach trójwymiarowych obrazy mogą być nakładane na siebie.

Wykorzystanie telemedycyny w chirurgii jest także możliwe w nowym zastosowaniu, jakim jest robotyka [21], [22]. Ta dziedzina co prawda jest dopiero na samym początku jej rozwoju, ale przykładowo Melvin i wsp. z Ohio w USA wykonali do 2002 roku 35 zabiegów operacyjnych z zakresu chirurgii gastroenterologicznych, w tym obwodową pankreatektomię i resekcję przełyku z zespoleniem w obrębie klatki piersiowej [22].

Literatura cytowana

1. Doolittle G. C., Williams A. R., Spaulding A., Spaulding R. J., Cook D. J.: A cost analysis of tele-oncology practice in the United States, Journal of Telemedicine and Telecare, 10, 27, 2004.
2. Vladzymyslavskyy A. V., Klymovytskyy V. G.: Successes of Ukrainian telemedicine, Journal of Telemedicine and Telecare, 10, 103, 2004.
3. Anogeianaki A., Ilonidis G., Anogianakis G., Lianguris J. i wsp.: A training network for introducing telemedicine, telecare and hospital informatics in the Adriatic-Danube-Black Sea region, Journal of Telemedicine and Telecare, 10, 3, 2004.
4. Graham L. E., Zimmerman M., Vassallo D. J. i wsp.:Telemedicine – the Way Ahead for Medicine in the Developing Word, Trop Doct., 33, 36-38, 2003.
5. Schulman E. D., Burton D. C., Domatov E. R.: Establishing a Low-Cost Telemedicine Link in Far-Eastern Russia, Journal of Telemedicine and Telecare, 8, (suppl 2), 74-76, 2002;
6. Linderoth H. C.: Managing telemedicine: from noble ideas to action. J. Telemed. Telecare 8, 143, 2002.
7. Gulube S. M., Wynchank S.: Telemedicine in South Africa: Success or Failure?, J Telemed Telecare. 7 (suppl 2), 47-49, 2001.
8. Zatari D.I.: Design of a Centralized Telemedicine Model in Palestine. Journal of Telemedicine and Telecare, 8 (suppl 2), 96-97, 2002.
9. Doarn C. R., Fitzgerald S., Rodas E., i wsp.: Telemedicine to Integrate Intermittent Surgical Services Into Primary Care. Telemed J E Health. 8, s. 131-137, 2002.
10. Ganapathy K.: Telemedicine and Neurosciences in Developing Countries, Surg Neurol. 58, 88-394, 2002.
11. Angood P. B.: Telemedicine, the Internet, and World Wide Web: Overview, Current Status, and Relevance to Surgeons. World J. Surg., 25, 1449-1457, 2001.
12. Cieśla-Dul M., Drop D., Pfitzner R., Koprowski A., Swat M., Rogóż R., Pasowicz M., Kapelak B., Plicner D., Sadowski J.: Telekonsultacje i telekonferencje w kardiochirurgii - program Krakowskiego Centrum Telemedycyny i Medycyny Zapobiegawczej. Terapia 11 [3 z. 1] s.32-34, 2003.
13. Grzesik S., Milanowski W., Legutko J., Woźniakiewicz E., Kulig J.: The Usefrulness of Telemedical Consultation and Technical Possibilities In Surgical Wards of Małopolska Region – Poland. Procedings of the Interbational Joint Meeting EuroMISE s. 135, 2004.
14. Głowacki M., Czernicki Z., Mierzejewski M.: The first in Poland system of transmission of imaging and neurosurgical telecommunication. Neurol. Neurochir. Pol., 36 (2), 363-369, 2002.
15. Chinzei K., Miller K.: Towards MRI guided surgical manipulator, Med. Sci. Monitor 7 (1), 153-163, 2001.
16. Hall J. C.: The internet: from basics to telesurgery, ANZ J. Surg. 72, 35, 2002.
17. Stanberry B.: Telemedicine: barriers and opportunities in the 21st century, J. Intern. Med., 247, 615, 2000.
18. Demartines N., Mutter D., Vix M., Leroy J. et al.: Assessment of telemedicine in surgical education and patient care, Ann Surg, 231, 282, 2000.
19. Rosser J. C., Bell R. L., Harnett B., Rodas E. et al.: Use of mobile low-bandwith telemedical techniques for extreme telemedicine applications, J. Am. Coll. Surg., 189, 397, 1999.
20. Schlag P. M., Hunerbein M., Ulmer C., Graschew G.: Operative Interdisziplinaritat und Telematik, Chirurg, 75, 411, 2004.
21. Maresceaux J., Soler L., Ceulemans R., Garcia A.: Bildfusion, virtuelle Realitat, Robotik und Navigation, Einfluss auf die chirurgische Praxis, Chirurg, 73, 422, 2002.
22. Melvin W. S., Needleman B. J., Krause K. R, Schneider C.: Computer-enhanced robotic telesurgery, Initial experience in foregut surgery, Surg. Endosc., 16, 1790, 2002.



BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 365-368.

ZASTOSOWANIE KRYPTOGRAFII W TELEKONSULTACJI KARDIOLOGICZNEJ

APPLICATION OF CRYPTOGRAPHY IN CARDIOLOGIC TELECONSULTATION

KRZYSZTOF SARAPATA, GRZEGORZ BILO, WIESŁAW PYRCZAK, ANDRZEJ KONONOWICZ, IRENA ROTERMAN-KONIECZNA, KALINA KAWECKA-JASZCZ

Zakład Bioinformatyki i Telemedycyny, Collegium Medicum Uniwersytetu Jagiellońskiego, bit@cm-uj.krakow.pl
I Klinika Kardiologii, Collegium Medicum Uniwersytetu Jagiellońskiego, Szpital Uniwersytecki

Streszczenie. W niniejszej pracy przedstawiamy prototyp sieciowego systemu umożliwiającego prowadzenie bezpiecznych i poufnych telekonsultacji kardiologicznych.
Zaletą tego systemu jest możliwość kontroli stanu pacjenta niezależnie od miejsca jego przebywania. Ponieważ lekarz nie może ustalać rozpoznania nie widząc i nie badając pacjenta, dlatego nasz system służy jedynie do monitorowania stanu pacjenta z już ustalonym rozpoznaniem i leczeniem.
Jako zastosowanie takiego systemu wybrano monitorowanie leczenia przeciwkrzepliwego. Podstawową informację przekazywaną lekarzowi jest tzw. wskaźnik INR (Czas protrombinowy), który jest miarą skuteczności leczenia. (Pacjent jest zobligowany do regularnego oznaczania INR w laboratorium). Na jego podstawie lekarz modyfikuje dawkowanie.
Komunikacja pomiędzy lekarzem i pacjentem dokonuje się przez medyczny serwer w sieci Internet. Każda ze stron posiada parę kluczy kryptograficznych. Pacjent na przypisanym sobie terminalu wypełnia formularz, który zawiera pytania stawiane pacjentowi w czasie wywiadu lekarskiego (tutaj będzie to wskaźnik INR oraz informacje dodatkowe). Lekarz odbiera wprowadzone informacje od pacjenta i wpisuje pod nimi swoje wnioski, które podpisuje elektronicznie i przesyła pacjentowi.
Pacjent i lekarz posiadają certyfikaty podpisane przez niezależne centrum autoryzacji, które umożliwiają dostęp na specjalnie do tego celu dedykowany serwer z określonym przydziałem ról (pacjenci lub lekarze).
Zdobyte przez nas doświadczenie może być wykorzystane przy realizacji koncepcji e-recepty (elektroniczne recepty). Nasza prezentacja ma na celu przedstawienie możliwości technicznych i prawnych bezpiecznego monitorowania stanu pacjenta w kardiologii.

Słowa kluczowe: telemedycyna, kryptografia, telekonsultacje

Abstract. A prototype of a web-based system enabling secure confidential teleconsultation in cardiology is presented. Advantage of our system is the possibility to control the patient's state regardless of his or her current location. Since the visual contact of doctors with patients is necessary to make a diagnosis, our system is ocused on monitoring patient's state on condition of earlier recognised diagnosis.
Anticoagulant treatment has been selected as the field of application of our system. The basic information transmitted to the physician is the INR value (prothrombin time), which estimates treatment efficiency. (A patient is obliged to test INR regularly in a laboratory). On the basis of INR, the physician is able to modify patient's drug dosage.
Communication between physician and patient is implemented through a medical server connected to the Internet. Each pair-pair. The patient fills out a form on his computer, which contains questions to be asked to the patient during the anamnesis (in our case it is INR and additional information). The physician receives information from patient, adds his findings, puts the digital signature and sends everything back to the patient.
Both the patient and the physician are equipped with digital certificates verified by an independent certification authority, which enables access to a dedicated medical server with identified roles of the participants (patient, physician).
Experience gained may be useful in accomplishment of the eprescription project. The goal of our presentation is to show the technical and legal significance of secure monitoring of patient's state in cardiology.

Key words: telemedicine, cryptography, teleconsultation

Wprowadzenie

Telekonsultacja jako forma konsultacji lekarskiej znalazła już zastosowanie w leczeniu przeciwkrzepliwym [1], [2]. Telekonsultacja lekarska (przez nas zrealizowana) jest formą tekstowej konsultacji przy użyciu medium połączeniowego, (Internetu) spełniającego: odpowiednie kryteria, gwarantującego autoryzację dwóch podmiotów konsultacji (pacjent i lekarz), niedostępność dla osób trzecich treści przekazy-

wanej w konsultacji oraz zdalnie przekazywane dokumentów opatrzonych podpisem elektronicznym.

Zastosowanie odpowiedniego poziomu bezpieczeństwa przesyłanych danych medycznych wraz z zastosowaniem prototypu recepty elektronicznej stanowi podstawę do rozwoju projektu oraz jego praktycznego wykorzystania.

Praca przedstawia efekt adaptacji tradycyjnie pojmowanej konsultacji medycznej na grunt telekonsultacji oraz informatycznych zmagań dla zapewnienia odpowiedniego poziomu technicznego telekonsultacji. Stworzone aplikacje informatyczne bazują na oprogramowaniu typu Open Source na licencji GNU [3].

Pacjent i lekarz posiadają klucze (dokładnie – certyfikaty podpisane przez niezależne centrum autoryzacji), które umożliwiają połączenie z serwerem i automatyczne zakwalifikowanie jako pacjent lub lekarz (może tutaj występować różnicowanie lekarzy wg ich specjalizacji). Te same certyfikaty służą także do opatrywania podpisem elektronicznym przesłanych dokumentów. Lekarz po zalogowaniu się (na podstawie swojego klucza) monitoruje stan „poczekalni", a momencie zalogowania się pacjenta, oczekuje na wypełnienie formularza i przesłanie go na serwer. Na jego podstawie lekarz dokonuje diagnozy oraz przepisania odpowiednich leków lub skierowania na inne badania. Pacjent ma możliwość po telekonsultacji zachować dokument lekarski zabezpieczony podpisem elektronicznym.

Nasza praca jest formą samodzielnej organizacji środowiska informatycznego dla systemu telekonsultacji.

1. Rozwinięcie problematyki medycznej

Serwis internetowy „Monitorowanie chorego z migotaniem przedsionków" poświęcony jest monitorowaniu leczenia przeciwkrzepliwego.

Podstawową informacją przekazywaną lekarzowi jest tzw. wskaźnik INR (czas protrombinowy), który jest miarą skuteczności leczenia. Pacjent jest zobligowany do regularnego oznaczania INR w laboratorium.

Komunikacja pomiędzy lekarzem i pacjentem dokonuje się przez medyczny serwer w sieci Internet. Pacjent na przypisanym sobie terminalu wypełnia formularz, który zawiera pytania dotyczące jego stanu zdrowia (tutaj będzie to wskaźnik INR oraz informacje dodatkowe: zażywane leki, dawkowanie, ciśnienie krwi itd.).

Lekarz odbiera wprowadzone informacje od pacjenta i wpisuje pod nimi swoje wnioski, które podpisuje elektronicznie i przesyła pacjentowi.

Krzepliwość krwi w tym konkretnym przypadku jest obniżona w stosunku do normalnych wartości krzepliwości (a więc podwyższony INR). Ta zmiana wynika z leczenia i jest zjawiskiem pożądanym. Niemniej jest szereg chorób cechujących się zaburzeniami krzepliwości w różnych mechanizmach, czyli tzw. skaz krwotocznych. Najczęściej wynikają one np. z niedoboru lub nieprawidłowego funkcjonowania płytek krwi (np. w białaczkach, po chemioterapii czy napromienianiu). Zmiany w INR występują tylko w niektórych typach skaz, najczęściej w chorobach wątroby (np. marskości) i przy niedoborze witaminy K. Grupa leków przeciwkrzepliwych, które nas interesują (acenokumarol i warfaryna), działają właśnie na zasadzie blokowania działania witaminy K.

W tym przypadku interesują nas chorzy z migotaniem przedsionków jest to stan, w którym przedsionki się nie kurczą, przez co dochodzi w nich do zwolnionego przepływu krwi, co sprzyja jej wykrzepianiu. Jeśli w przedsionku powstanie zakrzep, może on się oderwać i powędrować z krwią do istotnych organów (mózgu, płuc, nerek) i zatkać tętnice doprowadzające tam krew, co określamy jako zator. Dlatego też u niektórych chorych z migotaniem przedsionków (tych ze szczególnie wysokim ryzykiem zakrzepowo-zatorowym) stosuje się profilaktycznie leczenie przeciwkrzepliwe. Oczywiście, za mało intensywne leczenie jest nieskuteczne, a zbyt intensywne może spowodować krwawienia nawet zagrażające życiu. Niestety, uzyskanie odpowiedniego efektu wymaga indywidualnego dobrania dawki leku u każdego chorego, co więcej, u danego chorego ta dawka może wymagać okresowej korekty. Dlatego konieczne jest regularne oznaczanie INR w czasie leczenia. Norma INR u zdrowego człowieka wynosi 0.8-1.2, natomiast u osób w trakcie leczenia przeciwkrzepliwego celem jest utrzymywanie go w granicach 2.0-3.0. W zależności od laboratorium wynik przychodzi z dokładnością do 1 lub 2 miejsc po przecinku. Jeśli w stanie pacjenta nie zaszły istotne zmiany (o które pytamy), to lekarzowi sam wynik INR (bez konieczności badania pacjenta) wystarcza, by zdecydować, czy utrzymać dotychczasowe dawkowanie leku, czy je zmienić. Oczywiście, w przypadku konieczności innych działań lekarskich (np. zmiana w przebiegu samego migotania przedsionków) konieczna jest „tradycyjna" wizyta.

2. Wybrane zagadnienia rozwiązania technicznego

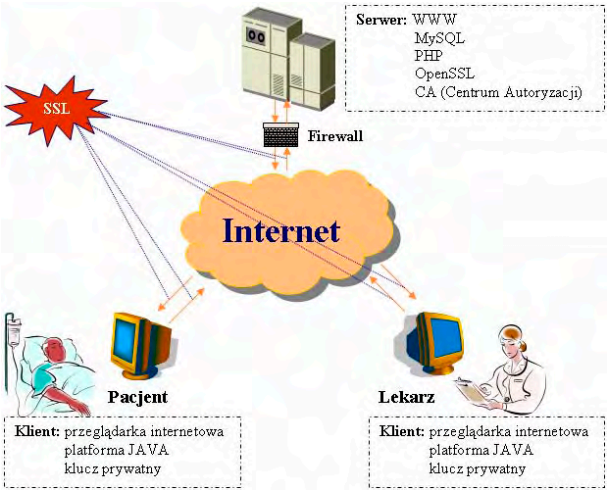
Przechowywanie danych medycznych, często także danych osobowych, ze względu na prawo do prywatności oraz ustawę o ochronie danych osobowych wymusiło zastosowanie odpowiednich środków bezpieczeństwa wobec zagrożeń pracy systemu w sieci publicznej, dostępu niepowołanych osób itd. System serwera (pełniący rolę pośrednika oraz rejestrujący i archiwizujący treść konsultacji) został wyposażony w wiele różnych elementów „bezpiecznych systemów". Część z nich, oczywiście, pozostaje poza sferą jawności. Jednym z aktualnych i powszechnie stosowanych rozwiązań kryptograficznych jest infrastruktura klucza publicznego (PKI Public Key Infrastructure). Wymaga ona istnienia centrum autoryzacji (CA), które służy do generacji certyfikatu centrum oraz kluczy użytkowników (co odbywa się w kolejności: generacja klucza użytkownika, podpisanie go certyfikatem CA, konwersja do formatu akceptowanego przez przeglądarki).

Charakterystyczne elementy zastosowane w projekcie to: sesje użytkownika, automatyczne odświeżanie stron, realizacja podpisu elektronicznego. Skrypty zarządzające konsultacją (aplikacje programowe) powstały w języku PHP, baza danych przechowująca treść konsultacji oraz dokumenty podpisane elektronicznie to MySQL, natomiast oprogramowanie klienckie oprócz przeglądarki internetowej zostało dodatkowo powiększone o aplet JAVY, którego zadaniem jest realizacja podpisu elektronicznego po stronie klienckiej (klucz prywatny użytkownika nie jest przesyłany na serwer!).

Sesje użytkownika to opcja serwera, która redukuje niedogodności sieci Internet (bezstanowy protokół HTTP) dla

ciągłej kontroli podmiotów konsultacji. Sesje opierają się na własnych funkcjach i bazie danych MySQL.

Kolejnym problemem wynikającym z wykorzystania protokołu HTTP była konieczność ciągłego odświeżania (zastosowano ramkowy układ strony WWW) dla celu aktualizacji treści wyświetlanych wg danych wprowadzonych przez użytkowników.



Rys. 1. Schemat poglądowy dla przyjętego rozwiązania technicznego telekonsultacji

I. Serwer

Serwer systemu to trójwarstwowa struktura: Baza danych(MySQL), Serwer aplikacji (PHP/Apache), Serwer WWW (Apache).

Warunki, jakie musi spełniać serwer na przewidziany użytek: szyfrowane połączenie, autoryzacja użytkowników, niezaprzeczalność przesłania dokumentów. W celu spełnienia tych założeń powołano lub wykorzystano następujące byty:

1. Centrum Autoryzacji (CA) – poświadczające swoim autorytetem (podpisem elektronicznym) respektowane przez serwer klucze użytkowników i serwera oraz umożliwiające spełnienie warunku niezaprzeczalności przesłanych dokumentów.
2. Klucze pacjentów i lekarzy – wykorzystane do autoryzacji podpisywanych dokumentów.
3. SSL (Secure Socket Layer) z autoryzacją użytkownika i serwera – szyfrowanie przesyłanych danych pomiędzy serwerem a klientem (tutaj między lekarzem a pacjentem z pośrednictwem serwera).

Ad. 1. Centrum Autoryzacji (CA, *certificate authority* lub *certification authority*) – jednostka wystawiająca cyfrowe certyfikaty sygnowane własnym podpisem elektronicznym; stanowi więc odpowiednik powszechnie zaufanego podmiotu. Na swoim wyposażeniu posiada tzw. klucze publiczne, służące innym do weryfikacji certyfikatów.

Certyfikaty są zbiorem informacji potwierdzającym tożsamość danej osoby (lub serwera, CA).

Wspomniana już wcześniej infrastruktura klucza publicznego jest umową typującą osobę trzecią do weryfikacji i poręczenia użytkownika. W jednoznaczny sposób przypisuje ona użytkownikowi klucz publiczny. Klucze publiczne są zwykle zawarte w certyfikatach.

CA wydaje certyfikat klucza publicznego, który potwierdza, że CA sprawdziło, że klucz publiczny wmontowany w certyfikat należy do osoby, organizacji, serwera lub innej jednostki ujętej w certyfikacie. CA jest zobligowane do weryfikacji kandydata, więc użytkownicy mogą polegać na informacji zawartej w certyfikatach. Ogólna idea polega na tym, iż użytkownik ufa CA i może zweryfikować podpis, wtedy CA wtedy może także zweryfikować, że klucz publiczny należy do tego, kogo zidentyfikowano w certyfikacie.

Ad. 2. Najważniejszym kluczem, gdyż jednoznacznie przypisanym danej osobie fizycznej, jest tzw. klucz prywatny, który powinien być chroniony przed utratą. Jest on generowany przez mechanizm generujący liczby pseudolosowe oraz w odpowiednim formacie przechowywany wraz z certyfikatem użytkownika (PKCS#12).

Ad. 3. SSL [4] wraz z autoryzacją użytkowników stanowią na dzień dzisiejszy bezpieczny system transmisji danych poprzez sieć publiczną. Mechanizm szyfruje wszystkie dane wypływające z aplikacji i deszyfruje dane wpływające do aplikacji. Na początku nawiązania połączenia SSL korzysta z odmiany szyfrowania asymetrycznego (szyfrowanie z kluczem publicznym), aby wynegocjować klucz, który będzie potem wykorzystany do szyfrowania symetrycznego.

CA wydaje certyfikat klucza publicznego po sprawdzeniu prawdziwości danych w nim zawartych: nazwy właściciela, organizacji, przydzielonego serwera i innych. CA dokonuje tego w sposób obligatoryjny.

X.509 jest standardem (ITU-T) infrastruktury klucza publicznego (PKI). X.509 zawiera min. specyfikacje, format PKI, wskazanie na użyty algorytm kodujący. Szczegóły można znaleźć pod adresem:
http://en.wikipedia.org/wiki/Certificate_authority

Przeglądarki internetowe zawierają domyślnie zainstalowane certyfikaty podstawowych zaufanych centrów autoryzacji. W naszym przypadku, w celu ominięcia kosztownej procedury akceptacji naszego serwera stworzono własne centrum autoryzacji, którego certyfikat użytkownicy muszą jednorazowo zainstalować na odpowiedniej liście w swojej przeglądarce internetowej.

II. Stanowisko klienckie

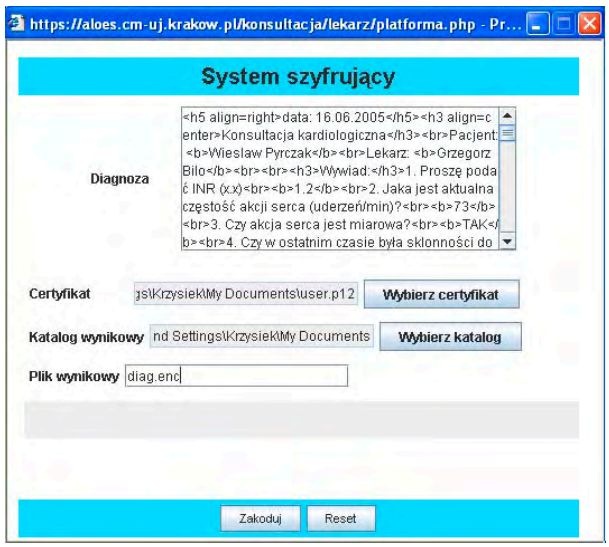
Stanowisko klienckie to tzw. *thin client*, którego idea polega na maksymalnym ograniczeniu wymagań sprzętowo-programowych stawianych po stronie klienta. Są to: komputer z dostępem do Internetu oraz przeglądarka internetowa.

W przeglądarce należy zainstalować klucz (podpisany przez nasze Centrum Autoryzacji). Klucz ten jest wystawiony personalnie, jest w nim zawarta min. informacja rozróżniająca lekarza i pacjenta. Instalujemy go w przeglądarce internetowej (instalacja zależna od rodzaju przeglądarki). Następnie ma miejsce instalacja platformy Javy, która jest dostępna w Internecie: (http://www.sun.java.com).

Oprócz klucza prywatnego klient otrzymuje plik, który nadaje uprawnienia do operacji na lokalnych plikach z poziomu apletów Javy. Powinien on być skopiowany do katalogu domowego użytkownika.

Rolą apletu szyfrującego jest podpisanie kluczem prywatnym tekstu diagnozy oraz zakodowanie i zapisanie go na dysku twardym lekarza (rys. 1). Do apletu przekazywany jest

jako parametr wejściowy tekst diagnozy. Tekst ten przesyłany jest w formacie HTML i nie może być modyfikowany z poziomu apletu. Użytkownik wybiera plik z certyfikatem oraz ścieżkę dostępu i nazwę pliku, do jakiego podpisany i zakodowany tekst diagnozy ma być zapisany. Aplet akceptuje pliki z kluczami w formacie PKCS#12 [5].



Rys. 1. Aplet szyfrujący

Aplet został zrealizowany w technologii Java (JDK 1.4) [6]. Funkcje kryptograficzne zostały zaczerpnięte z bezpłatnej biblioteki Bouncy Castle (wersja 1.28) [7].

By móc korzystać z apletu, lekarz musi dodatkowo skonfigurować zainstalowaną na własnym komputerze wirtualną maszynę Javy, gdyż standardowe ustawienia bezpieczeństwa nie pozwalają na dostęp do lokalnego dysku z poziomu apletu. Nie jest to jednak skomplikowana czynność i może być wykonana przy użyciu dostarczanego wraz z dystrybucją Javy narzędzia o nazwie Policy Tool.

Ze względów bezpieczeństwa zdecydowaliśmy się na implementację mechanizmu podpisującego diagnozę po stronie klienta, a nie serwera. Lekarz może nie ufać oprogramowaniu zainstalowanemu po stronie serwera i może nie chcieć przesyłać (nawet szyfrowanym kanałem transmisyjnym) własnego klucza prywatnego na serwer. Działanie apletu po stronie klienta może być w pełni kontrolowane przez użytkownika.

System telemedyczny wykorzystuje sieć Internet, w której ulokowany został serwer pełniący rolę pośrednika.

Podsumowanie

Wersja prototypowa systemu telekonsultacji wyposażona jest jedynie w możliwość asynchronicznego przesyłania danych medycznych, ich interpretacji oraz wskazówek dla pacjenta, który nie wymaga bezpośredniej konsultacji lekarskiej. Rozwiązanie to nie uwzględnia aspektów psychologicznych procesu konsultacji medycznych. Pacjent pozbawiony możliwości synchronicznej konsultacji z lekarzem (np. poprzez czat lub wideokonferencję) może odczuwać lęk przed przeoczeniem przez lekarza ważnych sygnałów dotyczących jego stanu zdrowia. Rozbudowa systemu o te mechanizmy powinna być ważnym elementem w pracach nad dalszym rozwojem aplikacji.

Otwarty problem, który nie został rozwiązany, to zabezpieczenia komputera klienckiego (przed atakami hakerskimi). Autorzy systemu w małym stopniu mają wpływ na poziom informatycznego zabezpieczenia komputera użytkownika, na którym dokonywana jest procedura podpisu elektronicznego z użyciem klucza prywatnego. Wydaje się, że wyjściem byłoby zastosowanie samodzielnego urządzenia realizującego podpis elektroniczny i przesyłającego wynik na lokalny komputer.

Osobnym zagadnieniem pozostają prawne regulacje e-recepty, które oczekują na prawne usankcjonowanie.

Możliwości przyszłego rozszerzenia systemu z punktu widzenia medycznego to podobne do obecnego monitorowanie wskaźnika INR prowadzonego u chorych po wszczepieniu sztucznych zastawek serca (w celu zapobiegnięcia wykrzepiania krwi na sztucznej zastawce).

Literatura cytowana

1. Finkelstein J., Khare R., Ansell J.: Feasibility and patients' acceptance of Home Automated Telemanagement of oral anticoagulation therap,. AMIA Annu Symp Proc., 230-4, 2003. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=14728168
2. Waterman A. D., Banet G., Milligan P. E., Frazier A., Verzino E., Walton B., Gage B. F.: Patient and physician satisfaction with a telephone-based anticoagulation service, J. Gen. Intern. Med., 16 (7), 460-3, 2001. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=pubmed&dopt=Abstract&list_uids=11520383
3. http://www.gnu.org.pl
4. www.ssl.org
5. PKCS #12: Personal Information Exchange Syntax Standard, http://www.rsasecurity.com/rsalabs/node.asp?id=2138
6. Helton J., Helton R.: Java Security Solutions, Indianapolis, Ind., Wiley, 2002.
7. The Legion of the Bouncy Castle http://www.bouncycastle.org

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 369-372.

CYBERNETYCZNE ZWIĄZKI NAUKI ZE SZTUKĄ

CYBERNETIC CONNECTIONS BETWEEN SCIENCE AND ART

JAN TRĄBKA, PAWEŁ TRĄBKA-ZAWICKI

Szpital Uniwersytecki Collegium Medicum Uniwersytetu Jagiellońskiego, Kraków, mytrabka@cyf-kr.edu.krakow.pl

Streszczenie. W pracy przedstawiono cybernetyczne podejście do zagadnienia związków pomiędzy działalnością artystyczną a działalnością naukową. Tytuł wydanej przed prawie pół wiekiem książki Beveridge'a *Sztuka badań naukowych* wydanej przed prawie pół wiekiem dziś zyskuje na aktualności dzięki najnowszym odkryciom fizyki i neurologii. Przykładem tego może być Teoria Topologicznej Geometro-Dynamiki (TGD) M. Pitkänena, wyrosła na gruncie fizyki teoretycznej. Takie połączenie wyobraźni z pragmatyką suponowane było już przez wienerowską cybernetykę. Dynamicznie rozwijająca się cyberkultura, będąca odpowiedzią społeczeństwa na coraz większą technicyzację codziennego życia, stanowi również tęsknotę nauki za wyobrażeniowo-intuicyjną formą poznania. Wydaje się, że nie sposób być twórczym w nauce i pomijać jednocześnie elementy subiektywnego przeżycia podmiotu. Jeśli nauka ma służyć człowiekowi, wzbogacać go i pomagać mu, to musi również posiadać humanistyczne atrybuty. Antropoiczne pryncypium nie oznacza ciasnego antropocentryzmu, a wręcz przeciwnie – jest otwarciem człowieka na siebie samego, na swoje własne doznania i artystyczne uniesienia. Sztuka nauki jest postulatem powrotu do piękna koncepcji naukowych jako pochodnej artystycznej kreacji.

Słowa kluczowe: sztuka, nauka, cybernetyka, cyberkultura, antropoizm

Abstract. The paper presents a cybernetic approach to the relation between artistic activity and scientific work. "Art of Scientific Investigation" is the title of Beveridge's book published almost half a century ago. Now it increases in value because of recent discoveries in physics and neurology. An example of that may be the Theory of Topological Geometrodynamics (TGD) by Matti Pitkänen based on the grounds of theoretical physics. Such connection of imagination and pragmatics was assumed in Wiener's cybernetics. Dynamically developing cybercultrure, which is an answer of the society to a continuous spreading of technical knowledge in everyday life, is also a longing of science for intuitive and imaginational form of cognition. It seems that it is not possible to be creative in science and to omit elements of subjective cognition. If science serves the human, enriches and helps him or her, it should also possess humanistic attributes. Anthropoic principle does not mean tight anthropocentrism, on the contrary, it is an opening on yourself, on your experiences and artistic passion. The art of science is a postulate of returning to the beauty science concepts being a derivative of artistic creation.

Key words: art, science, cybernetics, cyberculture, antropoism

Pomimo, iż książka Sztuka badań naukowych wydana została przez W. I. B. Beveridge'a przed prawie pół wiekiem, nie straciła nic swojej aktualności – nadal inspiruje, a nawet prowokuje myślenie współczesnych badaczy [1]. Już sam tytuł na pierwszy rzut oka wydaje się być paradoksalny, ponieważ nauka rozwija się raczej pod wpływem racjonalnym, skoro wymaga prawdy, która przekłada się na precyzję, niesprzeczność i klarowność. Obecnie przyjmuje się, że prawda to nie tylko nauka, ale w równej mierze należy także do głównych przejawów życia, bo dysponuje cechami wirtualnymi i bytami możliwymi. To szersze określenie prawdy zbliża naukę do sztuki, która bardziej grawituje ku emocjonalności, wyobraźni i intencjonalności. Jakkolwiek nie może wyrzec się całkowicie racjonalności. Wynika stąd, że radykalne przeciwstawianie sztuki nauce ulega stopniowemu spłaszczeniu i „zblednięciu". Sztuka, zwłaszcza w fazie powstawania, czyli swoich narodzin, tzn. właśnie w trakcie badań, szczególnie doświadczalnych, wielokrotnie musi odwoływać się do materiału wyobrażeniowego, czyli do hipotez wyjściowych, których racje doświadczalne mają dopiero być udowodnione lub obalone. Także i ostatecznie uzyskane wyniki z badań muszą być poddane rozumowym osądom, chociaż ich ostateczny efekt powinien być oceniany głównie pod auspicjami futurystycznego celu. Z autopsji znam ów trudny i złożony problem, ponieważ sam prze wiele lat przeprowadzałem eksperymenty w instytutach klinicznych.

Widać stąd, że nauka i sztuka przenikają się wzajemnie, a przepaść między dwoma owymi dziedzinami – wykopywana przez metodologię, czyli przez różne środki badawcze – prawie znika, zwłaszcza, gdy zagadnienie przerzutuje się z gruntu klasycznego do cyberprzestrzeni. Cybersztuka łatwo zespala się z cybernetyką jako nauką, ponieważ obydwie uzupełniają się i stają się komplementarne. I już jako dziedziny zespolone muszą zintegrować się z wydzieloną i bujnie

ostatnio rozwijającą się, choć na marginesie głównego nurtu cybernetyki, gałęzią nauk technicznych: cybersieciami („Kybernettes"), którym obecnie poświęcono osobne czasopismo (*Kybernettes. The International Journal of System and Cybernetic*). Obecnie więc nie chodzi ani o podkreślenie izolacjonizmu techniki i humanistyki, ani zawłaszczenie jednej dziedziny przez drugą, ale o postęp i dalszy wszechstronny rozwój, którego dopatruje się aktualnie w uwzględnieniu roli sieci tworzących się w mikroświecie kwantowo-falowym. Są to sieci funkcjonujące w wymiarach nanoskopowych i stanowiące substrat dla obiektów psychicznych – psychonów reprezentujących pojęcia lub idee, czyli materiał duchowy R. Descartesa (tzw. *mind's stuff*). Nic „cieńszego" niż świadomość kosmiczna nie istnieje nawet w sennych marzeniach. Przytoczyć tu należy badania sieci protonowych, które powstają z rozpadu jądra atomowego. Pod wpływem bombardowania wysokoenergetycznym, magnetycznym promieniowaniem kosmicznym jądra atomowe ulegają destabilizacji i rozpadają się na gluony i neutrina. Owe neutrina są cząsteczkami prawie bezmasowymi, ale dysponują ogromną energią ($W^{+/-}$ 89 GeV, Z^0 93 GeV), warunkującą niewyobrażalną przenikliwość. Ze względu na swe własności fizyczne i brak wszelkiego powinowactwa chemicznego, strugi neutrinowe upodobniają się do legendarnego eteru, który jako kategoria wirtualna dobija się o konkretny status. Dzięki cybersieciom protonowym zintegrowane mogą być odległe od siebie miejsca w mózgu, w układzie słonecznym, a nawet kosmicznym.

Oczywiście, sieci protonowe nie są i długo nie będą technicznymi obiektami operacyjnymi, ale już teraz powinny indukować nowe podejście badawcze. Są pochodnymi nowej fizyki, czyli Topologicznej Geometro-Dynamiki (TGD), która reinterpretuje wszystko, co dotąd poznaliśmy (z wyjątkiem cybernetyki) [2]. Chaos naturalny, narzucony świadomości człowieka został przyjęty i gdyby nie busola cybernetyczna, byłby nie do opanowania. Analogiczną rolę spełnia myślenie cybernetyczne, gdy od makroświata, kondensującego na polu Bosego przechodzimy na dno bytu, czyli do mikroświata neutrinowego [3]. Nie pora teraz ani nie miejsce na to, aby zapuszczać się w dalekie horyzonty nowej biofizyki, która, jak na razie – ze względu na techniczną niewydolność – przez dłuższy czas jeszcze będzie pędzić żywot cieni platońskich.

Wróćmy więc do lektury W. I. B. Beveridge, a zwłaszcza do jej motywu przewodniego, głoszącego wprost, że „*badanie naukowe samo w sobie nie stanowi nauki, ale raczej sztukę lub kunszt*" [1]. To zdanie pozostaje w pełnej zgodzie z cyberkulturą, bo faktycznie tworzenie sztuki nie może odbywać się bez asystencji racjonalności, najbezpieczniejszej drogi do prawdy; przecież sztuka nie może być „bezmyślna". Z tego stwierdzenia wcale jednak nie wynika, że badania naukowe nie mogą zawierać ani szczypty sztuczności. A z kolei sztuka nie może posługiwać się wyłącznie teksturą intelektualną, czyli być okryta racjonalną pajęczyną. Rozpatrując związki cybernetyki z nauką i ze sztuką obserwuje się jedność i spójność działalności intelektualnej zgodnie z założeniem, że człowiek, poznając – przeżywa, a przeżywając – poznaje [4]. Niemniej badania naukowe powstają w pełnym nurcie świadomości oraz przy całkowitym zaangażowaniu reflektora uwagi. Natomiast sztuka powstaje raczej na peryferiach, na okrainach pola świadomości i nie wymaga najpełniejszego jej strumienia światła. Sztuka sadowi się bliżej wirtualno-potencjalnej rzeczywistości, czyli bliżej życia; intuicja opiera się bardziej na przeczuciach, preferuje raczej atmosferę emocji i intencjonalizmu niż logicznego sztywnego rozumu. Można więc przypuszczać, że mechanizmy sądzenia, ferowania wyroków rządzą się innymi przesłankami niż poznanie przez sztukę, czyli przez obcowanie z uczuciami i woluntaryzmem.

Poruszając różne aspekty cybersieci, staramy się zrozumieć, jak człowiek rozwija sensowną i satysfakcjonującą współpracę między humanistyką a techniką. Niestety, nadal pozostaje intrygujący problem naukowy, czy funkcjonuje jedna wspólna dynamika organizmów biologicznych niezależnych od obiektów technicznych – partnerstwo, czyli symbioza techników i sztukmistrza.

Modele robią obecnie zawrotną karierę, a współczesne modelowanie to nic innego, jak eksperymentowanie myślowe wywodzące się z heideggerowskiego doświadczenia myśli (*Aus der Erfarung des Denkes*). W dalekowschodniej filozofii pojawiła się mistyczna sugestia „wchodzenia do środka myśli", dzięki niej udało się odkryć, że myślenie to nie tylko racjonacja (*rationatio*), ale także wyobraźnia (*immaginatio*), ukierunkowywanie motywacjami (*intentio*) w celu sięgania w jak najbardziej odległy los w przyszłości oraz zakotwiczanie w najgłębszej historycznej przeszłości. Stąd wniosek, że modelowanie doświadczane w czasie teraźniejszym staje się w istocie rzeczy medytacją hermeneutyczną, lansującą różne wybory lub wręcz przechodzi w fazę kontemplacji. Modelowanie więc jako eksperymentowanie na różnych zestawach myślowych, czyli modelach, obejmuje wszystkie „agendy" psychiczne intelektu ludzkiego oraz wszystkie naturalne sfery życia duchowo-fizycznego człowieka. Gwoli sprawiedliwości trzeba przyznać, że cybernetyka jako pierwsza z nauk zaczęła metodycznie i programowo propagować modele, a modelowanie wciągnęła na piedestał rygorystycznej działalności naukowej.

Świadectwem tego stanu rzeczy stały się Krajowe Konferencje pod tytułem „Modelowanie Cybernetyczne Systemów Biologicznych" (MCSB) organizowane od 1975 roku co pięć lat przez dwie krakowskie uczelnie wyższe: techniczną i medyczną. Wydawane pokonferencyjne materiały zawierały referaty i dyskusje – można tu już mówić o pewnej tradycji. Przyszła konferencja MCSB odbędzie się w terminie: maj 2005.

W trakcie rozkwitu tendencji modelarskich okazało się, że wiennerowska cybernetyka, zajmująca się w pierwszym okresie tylko makrosystemami, nadaje się świetnie do obsługi zarówno astronomicznych gigasystemów, jak też kwantofalowych mikrosystemów. A modele stały się jedynym spolegliwym naukowym narzędziem badawczym. Szkoda tylko, że intelektualnym i przeważnie czysto teoretycznym, bo przecież możliwości eksperymentatorskie w dziedzinie astrofizyki galaktycznej oraz na polu wyimaginowanej mechaniki kwantowej są znikomo małe, ze znaczną komponentą iluzji. Ponadto opierają się prawie wyłącznie na życzeniowo-tendencyjnych obserwacjach, czyli na „świadectwach-relacjach" ułomnych i na nie do końca jeszcze rozpoznanych, znikomych zdolnościach zmysłowych człowieka. Do dnia dzisiejszego nie wiemy, iloma zmysłami dysponujemy. Od czasów Arystotelesa, który uznał tylko pięć zmysłów, ich lista wydłużyła się do około dwudziestu modalności. Obecnie żaden szanujący się neuropsycholog nie ośmieliłby się ich wyliczyć, skoro brane są pod uwagę nie tylko ich proste, fizyczne sparametryzowane obiektywne charakterystyki, ale głównie ich subiektywnie przeżywane złożone moduły i odcienie. Artysta malarz wyróż-

nia podobno aż ponad sześćdziesiąt odcieni w zakresie jednej barwy. Wiemy to na podstawie zachowania odruchów warunkowych, a na podstawie eksperymentów psychofizycznych udowodniliśmy, że barwa jako jedna prosta modalność nie istnieje, ale każdą percepcję – wrażenie – warunkują trzy parametry fizyczne, z których jednego może brakować, a umysł – mózg – może sobie go uzupełnić: iluzyjnie dośpiewać lub dopatrzeć. Zachodzi więc uzasadnione pytanie: jak i czym w istocie operuje percepcja, czyli akwizycja danych o otoczeniu zewnętrznym?

Pomimo ogromnych mankamentów psychoneurologii świadczących o głębokiej ignorancji tego, co faktycznie dzieje się w mózgu, metafizyczna nanotechnologia wdziera się bezpardonowo w kulturę i zaczyna przynosić pożyteczne rezultaty w społeczeństwie. Wydaje się, że zespolony układ mózg-umysł dysponuje różnymi stopniami zdolności rozdzielczej i w zależności od tego percypuje tyle szczegółów z otoczenia, na ile filtr separacyjny zostanie nastawiony. Ów stan rzeczy zmusza badaczy do zmiany dotychczasowego stanowiska wobec braku realnego wsparcia, a nawet błędnych wskazań biologicznych aparatów recepcyjno-pomiarowych. A ponadto neurobiologowi wypada zmienić pogląd na matematykę i technikę jako na nauki czysto formalne. Okazuje się bowiem, że matematyka operuje nie tylko pustymi znakami formalnymi, a technika to nie tylko rękodzieło i rzemiosło, ale dziedziny z bogatym zapleczem inspiracji metafizycznych, które nie są wysuwane na pierwszy plan, choć skutecznie interweniują zza kulis sceny życia. Źródłem owych inspiracji, z których rodzą się owe interwencje są egzoterestialne (pozaziemskie, pozamózgowe), a nawet astralne związki semantyczne. Niestety, ich zależności nie dostrzegła nauka klasyczna i traktowała je jako objaw mistycyzmu, fantazji i światłoćmiącej fikcji.

Procedury pomiarowe, od których na obecnym etapie badawczym wymaga się maksymalnej dokładności teoretycznej, bo niewyobrażalnej, a zwłaszcza konieczność wprowadzenia skalowania, stawiają człowieka wobec całkiem nowych, dotychczas prawie niezauważalnych problemów.

Po obaleniu paradygmatu o absolutnych: czasie i przestrzeni, zostaliśmy „wyprowadzeni w pole" wielowymiarowości przestrzeni i zmienności – nawet kierunkowej – czasu. Do nie tak dawna jeszcze hołdowaliśmy niewzruszonym kanonom trójwymiarowej przestrzeni Euklidesa oraz niewzruszonego poczucia czasu opatrzonego strzałką, czyli jednokierunkowym wektorem. Obecnie jesteśmy skłonni przyznać, że determinacja neurofizjologiczna zamyka ludzką percepcję w dwuwymiarowej, płaskiej powierzchni, a trzeci wymiar przestrzeni to kompozycja wirtualna dokonana wysiłkiem intelektu. Oprócz tego, że sami – choć nieświadomie – odcięliśmy się od bezpośredniego fizycznego kontaktu z naturą, zamieniając go na metafizyczny w dosłownym, bardziej filozoficznym znaczeniu. A więc nie tylko w sensie Andonikosa z Rodos, co należy rozumieć tylko jako formalne następstwo „po", ale także poznanie zawierające tchnienie duchowości naturalnej oraz obejmujące żywioł semantyczny, merytoryczny, sygnalizacyjny.

Na obecnym etapie skazani jesteśmy na drugą niekorzystną okoliczność rozmycia się definicji poznania. Zgodnie z filozofią procesową A. N. Whiteheada nie ma tylko obiektywnego, tylko konkretnego, egzystencyjnego procesu poznawczego, bez równoczesnego, subiektywnego, wirtualnego przeżywania go w sposób prywatny w *selfie*. Racjonalnym, sensualistycznym zasadom poznania towarzyszyć musi emocjonalnie podbarwiona intelektualna komponenta. I *vice versa*: nie można niczego przeżywać bez jednoczesnego zmysłowego, antropoicznego, społecznie uzgodnionego rozpoznawania zewnętrznych cech świata otaczającego.

Oprócz tego należy przyznać, że bez włączenia nanotechnologii komputerowej trudno byłoby wyobrazić sobie dalszy postęp, zwłaszcza, że wielu badaczy nie liczy się z możliwością, że mogą być wygłaszane tezy prawdziwe bez możliwości ich eksperymentalnego potwierdzenia. Obecnie uczony powinien wiedzieć, że w puli mądrości kosmicznej mogą ciągle tkwić myśli zakryte, których nie może i których nie powinien znać. Mędrzec Sokrates chełpił się, że „*tylko jedną zna pewną myśl, że wszystkiego nie może wiedzieć*". To znacznie głębsza dewiza: „*wiem, że nic nie wiem*" przypisywana M. Sokratesowi.

Neurofizjologia dostarcza dowodu, że są głębie, do dna których nie można dotrzeć żadną sondą. Weźmy dla przykładu najważniejszy analizator wzrokowy, którego sama natura obarczyła szeregiem ograniczeń percepcyjnych, wynikających choćby z funkcji znikomych rozmiarów anatomicznych kamery oka. Pomimo tego przyzwyczailiśmy się widzieć nie tylko to, co przyroda faktycznie eksponuje, ale to, co chcemy widzieć. Przyzwyczailiśmy się oglądać świat trójwymiarowo w reżymie geometrii Euklidesa, podczas gdy mózg odbiera otoczenie płaszczyznowo, w przybliżeniu – dwuwymiarowo. Jesteśmy urodzonymi „płaszczakami", a operowanie przez mózg ludzki tylko powierzchniami, obrazami, wprowadza człowieka w zdumienie, a potem osłupienie. Konstelacja fal elektromagnetycznych pasma widzialnego projektuje się na siatkówkę jako twór płaski, choć warstwowy, czyli nadal w sposób płaski przedostaje się do ośrodków wzrokowych na wszystkich piętrach układu nerwowego. Czystą fantazją antropoiczną, zasugerowaną przez technikę informacyjną, pozostaje koncepcja kodowania obrazów w domenie czasu i przestrzeni. Na każdym końcu drogi wzrokowej kod czasowy można zastąpić kodem przestrzennym, dyktowanym teorią miejsca – jak na błonie podstawowej ślimaka analizatora akustycznego. Wyniki badań neurofizjologicznych niedwuznacznie świadczą, że trójwymiarowe parametry przestrzenne oraz metryka czasowa są psychologicznymi fantasmarami. W naturze nie ma miejsca na falsyfikaty i wszystko tam dzieje się w sposób przyrodniczy, to znaczy jednolity, bez potrzeby uwzględniania schizoidnej zasady niepewności (lub niedokładności) W. Heisenberga, która musi momenty pędu (czasu) rozważać niezależnie i osobno od momentów umiejscowienia (czyli przestrzeni). Niestety, w tajemnice owego rozszczepienia nie należy wchodzić „z butami", ale można do niej przybliżyć się asymptotycznie, na najbardziej minimalny dystans. Między człowiekiem „zmysłowym" a naturą powstaje zbudowana przez intelekt jakby szyba, która pozwala jedynie na prześwitywanie – przybliżanie się (heideggerowskie *gelassenheit*). Oczywiście, natura dąży do tego, aby „prześwitujące" kontury naturalne stawały się coraz wyraźniejsze i wierniejsze, aby nieustannie przybliżać się do bytu rzeczywistego, jaki istnieje sam w sobie, bez retuszu intelektualnego.

Inna zagadka, która tkwi w enigmie poznawania i przeżywania to fakt, iż dysponując tylko operacjami zmysłowymi, czyli wyobrażeniami płaszczyznowymi, dwuwymiarowymi, potrafimy interpretować obrazy z giga-, mega-, makro-, i mikroświata. Wszak domyślamy się, że ze zmianą stanowiska obserwacyjnego percepcja musi zmieniać swój charakter

i mechanizmy abstrakcji, ponieważ posługuje się odmienną zdolnością rozdzielczą detali. Pojawia się więc pytanie, dlaczego, wznosząc się na coraz wyższe pułapy wysokości lotów kosmicznych, pilotom dokonującym obserwacji ze statków ukazują się szersze horyzonty, pełne różnych szczegółów, a giną krajobrazy dawne. To wskazuje jednoznacznie, że muszą obowiązywać odmienne zasady abstrakcji i generalizacji. Tu trzeci wymiar wysokości staje się balastem, a trójwymiarowość – fikcją. Zapuszczając się w głąb mikroświata, operacyjnie i spektakularnie również anuluje się trzeci wymiar. Tu także wyobrażenia muszą ulec spłaszczeniu i redukcji do dwóch wymiarów, a nawet do bezwymiarowego punktu. Struny T. Kałuży, a zwłaszcza hiperstruny, są obiektami punktowymi, jakkolwiek – według teoretycznych założeń – rozciągliwymi jak guma.

Współcześnie przyszedł stosowny moment, aby zastanowić się, co pozostało po obecnej, najbardziej radykalnej rewolucji naturalnego ducha; co ocalało z ogromnej zawieruchy duchowego (intelektualnego) żywiołu; jakie ramy uogólnień; jakie reguły abstrakcji; jakie drogowskazy i jakie wsporniki pozostały człowiekowi na dzień dzisiejszy? Prawie wszystko, co zostało wymyślone i co misternie nauka stworzyła i przedstawia „do wierzenia", okazuje się mistyfikacją i alienacją. Nie powinniśmy się więc obecnie dziwić, że wpadają w ręce dzieła Seid Lloyda [6], E. Bobuli [7], z których wynika niedwuznacznie sugestia, że najlepiej byłoby zapomnieć to, czego w mozole nauczył się człowiek w ciągu minionych wieków. Ja z całą pokorą swego ducha ośmieliłbym się dorzucić jedną, skromną radę: nauki nie zapominać, klasyczną wiedzą nie gardzić, ale przeciągnąć ją za kulisy sceny życia, pozostawiając im status wirtualności i potencji.

Chęć zmiany poglądów na świat w celu dotarcia do istoty natury musiała przyczynić się zburzenia gmachu dotychczasowych półprawd. Warto także w ramach refleksji porewolucyjnej, po opadnięciu erystycznych emocji i ustaniu gorących dyskusji zastanowić się, czy nie bardziej komfortowo żyłoby się, gdybyśmy nadal wierzyli, że atomy są niepodzielnymi cegiełkami budowy kosmosu, gdybyśmy byli niewolnikami spętanymi pseudogenetycznymi, pozornie dziedzicznymi paradygmatami trójwymiarowej przestrzeni i nieskończonego, jednokierunkowego liniowego czasu fizycznego? Ale komfort to nie kryterium badaczy dążących do sedna sprawy i do pełnej prawdy. A kompromis konformistyczny gasi w człowieku wrodzony instynkt do umiłowania dobra i piękna oraz moralny imperatyw sprawiedliwości.

Pod koniec XVIII wieku pojawił się autentyczny geniusz, wspaniała genetycznie odległa krzyżówka Włocha i Niemki: Bernard Placyd Jan Neponucen Bolzano (tylko trzecie imię „Jana Nepomucka" zdradzało związek z Czechami, jako że mieszkał w Pradze, w której się kształcił i wykładał w seminariach duchowych). Genialne dzieło B. Bolzana *Paradoksy nieskończoności*, pomimo iż powstało na przełomie XVIII i XIX wieku, nie straciło nic ze swojej aktualności i na całe przyszłe czasy – jak wolno domniemywać – będzie źródłem inspiracji i natchnienia. Nie zgadzam się z T. Kotarbińskim [8], że autor wyniki „swych dociekań wyłożył w sposób niezwykle jasny i systematyczny". Właśnie bolzanowski język i własna oryginalna terminologia stanowi najtrudniejszą przeszkodę lingwistyczną do przyswojenia sobie geniuszu tego giganta myśli. Dla wielu współczesnych badaczy geniusz B. Bolzana to bardziej przedmiot do przeżywania niż do pojmowania. Ja na przykład nie wiem, jak rozumieć „tezę" (*satz on sich*) oraz „ideę" (*vorstellung on sich*) lub pojęcie „jako wyobrażenie samo w sobie". Męczę się z tekstem, ale się zbytnio nie dziwię, ponieważ ze współczesnym E. Husserlem mam podobny kłopot ze zrozumieniem. Trafiają się intelektualiści „sobki", którzy tworzą własny, pełen min lingwistycznych system filozoficzny, który wymagałby tłumacza, aby od licznych „efektów karmelowych" czytelnikowi nie zawróciło się w głowie. Ironiści mówią, że „autorzy tworzą auto-kod", albo lepiej „auto-szyfr" na swój własny użytek. Mądrzejszym okazał się I. Kant, który wprawdzie także na swym koncie ma *Ding an sich*, ale, wprowadzając szerokim strumieniem naturalną duchowość do swego „oględowego – transcendentnego" myślenia, starał się zachowywać raczej milczenie, bo nie ufał instytucji języka. Wracając do intuicji B. Bolzana, chcę podkreślić, że tam znalazłem „złote runo" – kamień filozoficzny, opisany w sposób mniej lub bardziej bezpośredni i dosłowny, że „od założenia świata" funkcjonuje ogromna pula kosmicznej mądrości, gromadzonej nie tylko przez wszystkie pokolenia *homo sapiens*, ale wszystkie istoty żywe, których relikty znajdować się mogą na Tytanie Jowisza. Z owej prastarej skarbnicy mądrości cała ludzkość wszechczasów nie tylko czerpie, ale którą współtworzy i wzbogaca antropoiczna świadomość. *Nihil novi sub sole*. B. Bolzano – jeden z nielicznych potomków, którzy znaleźli się lub będą znajdować się na planecie – pozostawił po sobie w spuściźnie genialną syntezę owej wszechświatowej, związkowej, spójnej esencyjnej mądrości [9]. Wprawdzie „nieskończoność" to nie tylko zaprzeczenie „skończoności", czyli – jakby się wydawało – próżnia Toricell'ego, ale także pełna bogactwa pustka

Literatura cytowana

1. Beveridge W. I. B.: Sztuka badań naukowych, tłum. L. Żebrowski, PZWL, Warszawa 1963.
2. Pitkänen M.: TGD inspired model for Nerve Pulse, Esselte 4T200, cz. I, 1-68.
3. Morese M., Perry P.: Przemienieni przez światło, tłum. K. Chudziński, Limbus, Bydgoszcz 1993.
4. Trąbka J.: Przeżywanie i poznanie – źródła świadomości, Konspekt 16/A, s.93-102.
5. Ascott R. (red.): Consciousness Reframed: Art and Consciousness in the postbiological era, Conference Proceedings, University of Wales, Newport 1997.
6. Lloyd P. B.: Everything you know is wrong. www.Lloydeye.com
7. Bobula E.: Przeformowanie świata, Kraków 2003
8. Kotarbiński T., Gnoseology: The Scientific Approach to the Theory of Knowledge. [tłum.] Olgierd Wojtasiewicz, (Elementy teorii poznania. Logika formalna i metodologia nauk.) 1929; 1961.
9. Bolzano B., Paradoksy nieskończoności, PWN, Warszawa 1966.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 373-374.

GENETIC IMMORTALITY IN THE LIGHT OF HOLISTIC PEDAGOGY

NIEŚMIERTELNOŚĆ GENETYCZNA W ŚWIETLE PEDAGOGIKI HOLISTYCZNEJ

ANDRZEJ SZYSZKO-BOHUSZ

Krakowska Szkoła Wyższa im. Andrzeja Frycza Modrzewskiego, Kraków

Abstract. The author devised the Theory of Genetic Immortality and a new branch of pedagogy – The Holistic Pedagogy. He presents the new holistic multidisciplinary view of several fundamental notions, as life, consciousness, personal identity, heredity, genetic memory, etc. In Accordance with to the above, the revision of to-date terminology is proposed. The author claims the notion of heredity can be understood as continuity of the parents' life in children. Therefore, the chain of life and consciousness has not been interrupted despite the death of an individual. Methods applied by the author during the investigation were: introspection, logic (logistics), observation, psychometric and physiometric tests, questionnaires, as well as spectrum of the photon emission (in cooperation with other researchers). The program of investigation including the state of relxation and the peculiar state of consciousness named by the author "the state undergoing no change" or "universal state of consciousness" was extended to about 1000 schoolboys and schoolgirls, and about 5000 university students.

Key words: Holistic Pedagogy, Genetic Immortality, consciousness, personal identity, genetic memory

Streszczenie. Autor jest twórcą Teorii Nieśmiertelności Genetycznej i nowego kierunku w pedagogice – Pedagogiki Holisycznej. Proponuje on nowy pogląd multidyscyplinarny, dotyczący szeregu podstawowych terminów: życia, świadomości, tożsamości osobniczej, dziedziczności, pamięci genetycznej itp. Zdaniem autora, pojęcie dziedziczności winno być interpretowane jako kontynuacja życia rodziców przez potomstwo. Dlatego łańcuch życia i świadomości nie zostaje przerwany mimo śmierci danego indywiduum. Metody zastosowane przez autora w trakcie badań były następujące: introspekcja, logika (logistyka), obserwacja, testy psychologiczne i fizjologiczne, ankieta, jak również badanie emisji fotonowej (we współpracy z innymi badaczami). Program badań, uwzględniający stan relaksacji oraz szczególny stan świadomości nazwany przez autora „stanem nie podlegającym zmianom" lub „uniwersalnym stanem świadomości, objął około 1000 młodzieży szkolnej obojga płci oraz około 5000 studentów wyższych uczelni.

Słowa kluczowe: Pedagogika Holistyczna, Nieśmiertelność Genetyczna, świadomość, tożsamość osobnicza, pamięć genetyczna

Introduction – principal hypothesis – methodological strategy

The author is an inventor of a new branch of pedagogy: the Holistic Pedagogy. According to it he presents the new holistic multidisciplinary view on several fundamental notions, as life, consciousness, personal identity, personality, heredity, memory, genetic memory etc. Therefore, the revision of to-date terminology is proposed.

The author's idea is an attempt to promote a seemingly incredible hypothesis which appears to be derived from science-fiction but, in fact, is simple, rational and clear. The author claims that the of notion "heredity" can be understood as continuity of the parents' life in children. According to the author's thesis the primordial reaction to the stimulus converts gradually into human consciousness, so the chain of life and consciousness has not been interrupted despite an individual's death.

According to the author's proposal, the essential definition of consciousness is "Consciousness is a perception of existence".

1. For a man this perception of existence is reflective (because of the work of the brain) and not reflective (because of the reaction of senses).
2. For animals the perception of existence takes a form of excitability and has exclusively not reflective character. The perception of existence is inseparably connected with LIFE.
3. The origin of consciousness and life is ONE.
4. Consciousness and excitability inseparably coexist with the living body and its genotype.
5. The individual (personal) identity of the living organism should not be considered analogously to the exact science, such as mathematics or mathematical logistics. In this disciplines the term "identity" is equal to the rule of constancy, unchangeability, reversibility: a-b b-a. On the contrary, in every organism, in each fraction of time, numerous changes of biomolecular, chemical, bioelectric, cybernetic character in the process of metabolism take place. Despite these changes, the identity of consciousness ("the personal identity") determining the individuality of the organism has not been changed. Therefore, the consciousness has not been destroyed despite the metabolism of the organism.
6. In the moment of sexual multiplicity the genotype of parents has been transmitted wholly, in 100% to their children. Therefore, the central hypothesis of the author reads "The consciousness of parents has been transmitted to their children". In other words "The parents, even after the death of their bodies, live in their children". The chain of life, excitability and consciousness from the immemorial time has not been destroyed.

The author's investigations have been carried out since 1961 to the present moment among persons of both sexes, mainly students. Among the methods applied by the author the following were the most important: logic, mathematical logic, introspection, observation,

psychometric and physiometric tests, inquiries, as well as a spectrum of the photon emission (in cooperation with other researchers). The program of investigation including the state of relaxation and the peculiar state of consciousness named by the author "The State Not Liable to Change" or "The Universal State of Consciousness" was extended to approx 1000 schoolboys and girls, and approx 5000 students of an university level school.

Selected results of investigation

In 1988-1990 the author made research concerning students' ideas about the essence of heredity. Students answered the following question (demanding some introspection): "Do life and consciousness of parents find continuity in children?"

Rather unexpected answers given by 135 students of pedagogy and rehabilitation of the Academy of Physical Education in Kraków were as follows:

Table I. Students' opinion on inheritance of parents' life and consciousness in children

Yes	%	No	%	No opinion	%
109	80,75	7	5,11	19	10,06

The verified small sample should be confirmed by more precise, wide, multidisciplinary investigations. Nevertheless, the author's hypothesis of genetic immortality and inheritance of life and consciousness seems to be very popular among students. It is worth noting that author's hypothesis has been accepted by representatives of different branches of science, medicine and philosophy. Professor Julian Aleksandrowicz M.D. author's close friend and colleague, a world famous scientist in hematology, used to say: "your hypothesis of inheritance of parents life and consciousness in children is absolutely *true*, but you will see, people will not understand it for a long time!"

Discussion

The question arises, whether the individual history of the species can be recapitulated only in the brain, since only the latter has the exclusive monopoly to recapitulate the past events? According to the "Biogenetic Law" formulated by E. Haeckel and F. Muller, the development of ontogenesis is equal to the shortened phylogenesis. Then it is *not* only the brain but the organism *as a whole* which reflects (copies) in the every movement (motion, reaction) the physiological process of the passing events from the genetic history of its species. This process of the recapitulation of the events from the history of the species, which manifests itself in reactions of one's body is called "genetic memory". The characteristic trait of this particular "memory", contrary to the common use of the term, consists of the fact that "genetic memory" is absolutely reliable, immune to destruction, mistakes and deformation. It manifests itself also in *atavism*.

"Genetic memory" proves convincingly that life and consciousness is continual, incessant and really *immortal*. If the memory of past events is the most credible criterion of one's identity, coded in the brain, then even more reliable criterion seems to be the genetic memory, embracing the *whole organism*.

Summary – Conclusions

The "Holistic Pedagogy" creates new multidisciplinary view on the entity of life, death and consciousness of man: a question of genetic immortality in organisms of common origin.

According to the author's idea, the revision of to-date terminology of several fundamental notions as life, consciousness, personality, personal identity, memory, genetic memory, heredity is absolutely necessary.

In his works the author has invented a new original definition of consciousness. Namely, consciousness is a reflective and non-reflective perception of existence (according to the person) and merely a non-reflective perception of existence (according to the animal). This perception, determining the identity of an organism, is inseparably existing with its genotype. Therefore, the parents' consciousness finds its continuity in the life of their children. This co-existence enabled the author to formulate the hypothesis of "genetic immortality", namely the continuity from generation to generation of the consciousness of parents' life in children. This idea suggests the immortality of all organisms of the common descent.

The author considers the genetic memory (also atavism) as a proof of continuity of forefathers' life and their consciousness in progenies.

According to the author's "Holistic Pedagogy": "All organisms of common genetic origin (including Mankind) are a *unity, "a single body"* equipped with an integrated system of communication, based on biomolecular, biochemical, bioelectric and cybernetic chains". Examples of a profound comprehension of the true unity of life can be tound in the scriptures of all major religions, the same can be shown in ancient and modern philosophical systems and it inspires the magnificent works of science, medicine, literature, poetry, paintings, sculpture, architecture and specially "*Holistic Pedagogy*".

The inspired Science Philosophy Art and Metaphysics seem to proclaim: LIFE IS ONE OMNIPRESENT IMMORTAL.

References

1. Benson H.: The Mind Body Effect, Berkeley, New York 1980.
2. Berg P., Singer M.: Dealing with Genes. Language of Heredity, University Science Books 1992.
3. Gnitecki J.: Globalne przewartościowania technologii, eksploatacji i użytkowania energii. Konstrukcja urządzeń zasilanych energią kwantową i subkwantową, Wydawnictwo Naukowe Polskiego Towarzystwa Pedagogicznego, Poznań 2004.
4. Green E., Green A.: Biofeedback and State of Consciousness, New York 1985.
5. Hubert J.: Creativity (A definition Based on the Concept of Negentropy), Dialectics and Humanism, 2, Warsaw 1985.
6. Sheldrake R.: A New Science of Life. The Hypothesis of Formative Causation, London, Anthony Blond 1981.
7. Szyszko-Bohusz A.: The Identity of the Personality as an Interdisciplinary Question, Commission Pedagogical Sciences, Vol. XXII/2, pp. 372-373, Wrocław, Polish Academy of Sciences, 1978.
8. Szyszko-Bohusz A.: Biological Immortality of the Multicellular Organisms. An Essay on the Border of Science and Philosophy, Cracow, AWF 1985.
9. Szyszko-Bohusz A.: The State Not Liable to Change. A Way for Humanity of Tomorrow, S. M. Dinsdale International Conference in Rehabilitation. Ottawa 1986.
10. Szyszko-Bohusz A.: The State Not Liable to Change, 6th International Conference on Psychotronic Research, Psychotronic Science Nowadays New Technology, Zagreb 1986.
11. Szyszko-Bohusz A.: Holistic Pedagogy, Wrocław, Polish Academy of Science, 1989.
12. Szyszko-Bohusz A.: A State of Awareness Not Affected by Change, Commission on Educational Sciences, Vol. XXXIII/2, pp. 269-270, Wrocław, Polish Academy of Sciences, 1989.
13. Szyszko-Bohusz A.: Is Mankind a Single Organism? Research on the Universal State of Consciousness. Mental Health in a Changing World, Polish Society for Mental Health, Warsaw, 1990.
14. Szyszko-Bohusz A.: States of Consciousness. Problems and Results, Kraków, AWF, 1992.
15. Szyszko-Bohusz A.: The Genetic Immortality. Is Consciousness Inherited?, Kraków, 1996.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 375-380.

ZNACZENIE MODELI W HIPOTEZIE GAI

MEANING OF MODELS IN THE HYPOTHESIS OF GAIA

HONORATA KORPIKIEWICZ

Zakład Filozofii Nauki, Instytut Filozofii, Uniwersytet Adama Mickiewicza w Poznaniu

Streszczenie. Hipoteza Gai przyjmuje, że jest ona homeostatem, który funkcjonuje na zasadzie ujemnego sprzężenia zwrotnego, pozwalającym jej na powracanie do stanu sprzed zakłóceń. Jednak wyniki badań składu atmosfery w dawnych epokach poddają to założenie w wątpliwość, sugerując możliwość występowania dodatniego sprzężenia zwrotnego, co skutkowałoby istotnymi zmianami klimatycznymi pod wpływem czynników antropogenicznych, jak również podważałoby samą hipotezę Gai.
Modele związków pomiędzy gatunkami (wypadających nitów, gatunków zwornikowych itd.) nie oddają właściwie zależności pomiędzy nimi, a co najważniejsze – nie potrafią odpowiedzieć na pytanie o wartość bioróżnorodności i konsekwencje jej zagrożenia. Opierając się na cybernetycznej zasadzie, że układ złożony jest bardziej stabilny, proponuję model elastycznej sieci związków pomiędzy gatunkami. Można oczekiwać, że model elastycznej sieci odpowie także na pytanie o wartość pewnych działań proekologicznych (np. prób introdukcji osobników ginących gatunków).

Słowa kluczowe: Gaja, bioróżnorodność, ekologia, homeostat, modele atmosfery, związki międzygatunkowe

Abstract. The hypothesis of Gaia assumes its being a homeostat that works according to a negative feedback, allowing the return to the state before disturbances. Nevertheless, research results of atmosphere composition in earlier times call this assumption into a question and suggesta possibility for positive feedback appearing, which would result in essential climate changes under human factors and the challenging of the Gaia hypothesis as well.
The models of relationships between species (prociduous rivets, keystone species, etc.) neither properly render the dependences among them, nor answer the question about the value of biodiversity and consequences of threats to it. I propose a model of an Hexlible net of relationships among species, based on the cybernetic rule of complex system being more stable. There are expectations that the model would also answer the question about the value of certain ecological activities (e.g. attempts of introduction of dying species).

Key words: Gaia, biodiversity, ecology, homeostat, models of atmosphere, relationships between species

1. Zależności atmosfera-klimat

W egzobiologii do niedawna panowało przekonanie, że życie można odnaleźć w podobnych warunkach fizykochemicznych jak na Ziemi, a więc w Układzie Słonecznym powinno zaistnieć na Marsie i Wenus. Te bowiem planety zawierały się w *ekosferze Strugholda*, w takiej odległości od Słońca, która zapewniała temperaturę wewnątrz sfery zawartą pomiędzy -70° a +80° C.

Sytuacja okazała się jednak bardziej skomplikowana. Dla zaistnienia życia właściwa odległość od gwiazdy centralnej nie jest wystarczająca, potrzebne jest jeszcze spełnienie szeregu innych warunków, np. stabilność parametrów orbity, co zapewnia masywny satelita. Co więcej, okazało się także, że bynajmniej nie szczególny skład atmosfery ziemskiej zapewnił na niej powstanie życia. Było to odwracaniem zagadnienia: życie, organizując się, stwarzało środowisko coraz bardziej sobie przyjazne. Zrozumienie tego procesu stało się podstawą do stworzonej przez Lynn Margulis i Jamesa Lovelocka hipotezy Gai [11]. Jednocześnie Lovelock wysunął wtedy sugestię odnośnie badań egzobiologicznych, żeby poszukiwać globalnych planetarnych oznak życia. Taką cechą miał być stan atmosfery, której skład i niestabilność były podtrzymywane przez życie.

Do podobnych wniosków na temat biologicznego pochodzenia gazów atmosferycznych doszedł niezależnie i znacznie wcześniej W. I. Wiernadski: biosfera zawiera nie tylko wszystkie żywe istoty, ale także ich naturalne środowiska wraz z materią nieożywioną: skałami, wodami, atmosferą [19]. Tym samym jego biosfera byłą bardziej bliska temu, co dziś rozumiemy pod pojęciem Gai niż homeostatyczny twór samego Lovelocka i Margulis. Zbliżone poglądy wygłaszali także inni badacze. Artur Redfield wysunął w 1958 roku hipotezę, że skład chemiczny hydrosfery i atmosfery jest kontrolowany przez życie. Natomiast James Hutton już w 1785 roku ogłosił, że uważa Ziemię za żywą istotę, której badaniem powinna się zajmować fizjologia.

Lovelock traktował Gaję jako *homeostat* i uzasadniał, że skoro radziła sobie ona z tak globalnymi kataklizmami, jakimi są epoki lodowe, to również obecne działania człowieka – zanieczyszczenia przemysłowe czy szkody wyrządzane przez rolnictwo – nie przywiodą jej do katastrofy. Także rejony zurbanizowane otoczone są, według Lovelocka, roślinnością, czekającą tylko na sposobność, aby „wziąć je znów w posiadanie", jak to się działo z ruinami miast po II wojnie światowej [12].

W czasach narodzin hipotezy Gai ruch ekologiczny dopiero się rodził i można odnieść wrażenie, że sam autor hipotezy znajdował się poza nim, nie dostrzegając pewnych, już wtedy istotnych, problemów. Nawet zmniejszanie się warstwy ozonowej z przyczyn antropogenicznych Lovelock bagatelizuje, zwracając uwagę, że taki proces ma miejsce również pod wpływem zjawisk naturalnych, poza tym w początkach życia Ziemi w atmosferze nie było w ogóle ozonu (bo nie było również tlenu), a pierwsze organizmy dawały sobie w takich warunkach doskonale radę.

Pełna wiary w możliwości autoregulacyjne Gai jest także współautorka hipotezy Gai, Lynn Margulis: *„To my znajdujemy się pod kontrolą planety nie odwrotnie. Nasz przerośnięty zmysł moralny, każący nam podejmować próby kierowania niesforną planetą lub leczyć ją z jej przypadłości, świadczy tylko o fantastycznej zdolności do samooszukiwania się. Jeśli już – powinniśmy chronić się przed samymi sobą"* [13].

Również współcześnie istnieją optymistyczne scenariusze wpływu zanieczyszczeń na środowisko. Np. lobby węglowe dowodzi, że wzrastająca ilość gazów cieplarnianych w atmosferze, głównie CO_2, wpłynie dobroczynnie na produkcję roślinną, co przyczyni się do zmniejszenia głodu na świecie (!). Rośliny wszak przyswajają dwutlenek węgla pod wpływem światła.

Gazowa otoczka Ziemi w początkach jej istnienia była beztlenowa, a uwalnianie się tlenu wiązało się z działalnością pierwszych *fotosyntetyzujących bakterii* sprzed ok. 3,5 miliarda lat. Tlen stał się istotnym składnikiem ziemskiej atmosfery prawdopodobnie ok. 2 miliardów lat temu, co stanowiło czynnik katastroficzny dla *anaerobów* – bakterii beztlenowych, które odtąd zmuszone były zmienić swoje środowisko życia na przewód pokarmowy rozwijających się na Ziemi gatunków. Fotosyntetyzujące bakterie i glony były najprawdopodobniej pierwszymi organizmami, które na globalną skalę wpływały na ziemskie środowisko i zdołały je zmienić w diametralny sposób.

Obecnie atmosfera Ziemi jest szczególną mieszaniną gazów, których współistnienie jest pogwałceniem równowagi chemicznej. W warunkach laboratoryjnych niektóre gazy występujące w atmosferze, jak np. tlen i metan, amoniak i tlenek azotu oraz inne – gwałtownie ze sobą reagują. To, że występują w stanie wolnym, wskazuje na ich nieustanne powstawanie na powierzchni Ziemi i uwalnianie się do atmosfery.

W wyniku fotosyntezy rośliny wytwarzają tlen atmosferyczny, natomiast dwutlenek węgla powstaje w procesie oddychania zwierząt i to w bezpośredniej zależności od ilości produkowanego przez rośliny tlenu. Koncentracja tlenu wynosi obecnie 21% i, jak to wykazał Andrew Watson, jej wzrost o 1% zwiększa zagrożenie spontanicznym pożarem lasu aż o 70%. Przy stężeniu tlenu 25% dzisiejsza szata roślinna bardzo szybko uległaby zniszczeniu [11].

Azot natomiast wytwarzany jest z azotanu i amoniaku przez mikroorganizmy gleby. Metan, produkowany w dużej ilości przez bakterie, przy tak dużej zawartości tlenu jest chemicznie niestabilny i łatwo utlenia się, a w wyniku tego procesu powstaje dwutlenek węgla i woda. W efekcie otrzymujemy atmosferę wysoce niestabilną, o niewielkiej *entropii*, która nie podtrzymywana przez życie przeszłaby bardzo prędko w stan stabilny o większej entropii. Tylko życie jest zdolne do zmniejszania entropii. Poszukiwanie innych form życia to szukanie *globalnych oznak malenia entropii*. Co ważne dla egzobiologii, można to zaobserwować z odległości kosmicznych, podczas gdy stwierdzenie istnienia poszczególnych organizmów wymaga ich bezpośredniej obecności.

Ziemskie życie uczestniczy w obiegu materii w litosferze i atmosferze Ziemi *(biotyczne cykle geochemiczne)*. Obieg ten rozpoczął się jeszcze bez udziału życia *(abiotyczne cykle geochemiczne)*, głównie pod wpływem energii promieniowania słonecznego oraz energii wnętrza Ziemi. Ta ostatnia wywołuje gradient temperatury pomiędzy wnętrzem planety a jej powierzchnią, ale na jego wartość składa się przy powierzchni Ziemi także energia słoneczna, której udziela gruntowi Słońce podczas nasłonecznienia. Najbardziej znanym obiegiem jest cykl wody; istniał on zanim pojawiło się życie, ale od czasu jego powstania życie w nim wydatnie uczestniczy.

Abiotyczne cykle obiegu pierwiastków istnieć muszą na wszystkich planetach wszystkich gwiazd, bo podstawowy warunek ich istnienia – dostarczanie energii gwiazdy do powierzchni planety – jest spełniony. Planeta posiadająca życie charakteryzować się będzie dodatkowo biotycznym obiegiem pierwiastków. W przypadku Ziemi na obieg biotyczny składają się przede wszystkim: *obiegi węgla, azotu i siarki* – i od początku istnienia życia na Ziemi obieg ten jest realizowany głównie przez drobnoustroje [18].

Ze względu na znaczenie dla stabilności klimatu, a w konsekwencji dla warunków życia większości organizmów biosfery, najważniejszy jest *obieg węgla*. Pierwotna atmosfera Ziemi, podobnie jak atmosfery Wenus i Marsa, nie zawierała tlenu, natomiast występowały w niej duże ilości dwutlenku węgla. Powstanie życia gwałtownie zmieniło sytuację – na powierzchni Ziemi zaczęły tworzyć się coraz większe ilości biomasy z zawartością węgla, natomiast ubywało węgla w atmosferze na koszt produkowanych coraz większych ilości tlenu w procesie fotosyntezy. Bilans ten „nie domykał się", tzn. pozostawała nadwyżka węgla w postaci biomasy na powierzchni Ziemi, która utworzyła eksploatowane dziś złoża paliw kopalnych.

Ocenia się, że współcześnie rośliny lądowe asymilują 75 miliardów ton węgla rocznie, a rośliny oceanów – 40 miliardów; prawie tyle samo dwutlenku węgla uwalnia się do atmosfery w procesie oddychania. Mniej niż ok. 1% asymilowanej ilości węgla jest deponowane w osadach [18]. Niewielkie „niedomknięcie" bilansu tworzy nowe złoża i nie zakłóciłoby stabilności atmosfery, gdyby nie fakt, że dawne złoża paliw kopalnych są zużywane w szybkim tempie. Wprawdzie spalanie drewna czy węgla uwalnia tyle CO_2, ile go przedtem związały rośliny, należy jednak pamiętać, że one czyniły to przez długie miliony lat, natomiast człowiek chce je zużyć w ciągu kilkuset lat (!). To, jak również wycinka lasów, stanowił przyczynę szybkiego wzrostu ilości dwutlenku węgla w atmosferze, który jako *gaz cieplarniany* zwiększa *efekt cieplarniany* i wybitnie wpływa na zmianę (ocieplanie się) klimatu.

Oprócz dwutlenku węgla w atmosferze wzrasta także ilość innych gazów cieplarnianych antropogenicznego pochodzenia. Jak się ocenia, w tym procesie udział procentowy dwutlenku węgla jest największy – 48%, kolejne miejsce zajmują freony – 20% oraz metan – 13 %. Ten ostatni, jak wykazał G. E. Hutchinson, produkują bakterie żyjące w bagnach oraz w przewodzie pokarmowym termitów i przeżuwaczy. Hodowla bydła, owiec i kóz wybitnie przyczynia się więc do zmiany stanu atmosfery [18].

Według obliczeń, o ile ludzkość nie powstrzyma emisji gazów cieplarnianych, ich ilość produkowana przez człowieka w 2030 roku osiągnie równowartość dwutlenku węgla produkowanego przez biosferę [2].

Nadmierna ilość gazów cieplarnianych w atmosferze grozi zachwianiem stabilności procesów klimatycznych i ociepleniem klimatu. Istnieje jednak czynnik działający przeciwnie, hamujący dochodzące do powierzchni Ziemi promieniowanie Słońca, a więc ochładzający klimat – *zapylenie atmosfery*. Może ono zmniejszać katastrofalne skutki emisji gazów cieplarnianych, może je nawet przewyższyć, działając

katastroficznie – wywołując gwałtowne ochłodzenie klimatu i zlodowacenia.

Rozstrzygnięcie problemu, w którą stronę zdążają globalne procesy wynikające z zanieczyszczania atmosfery, nie jest wcale łatwe i zależy od ustalenia charakteru *sprzężenia zwrotnego* pomiędzy zmianami klimatu a zmianami stężenia CO_2 oraz metanu w atmosferze, a także zawartością pyłów (naturalnego czy antropogenicznego pochodzenia).

Hipoteza Gai traktuje Ziemię wraz z biosferą i atmosferą jako *homeostat* – układ *stabilny,* zdolny do zachowania *homeostazy*, działający na zasadzie ***ujemnego sprzężenia zwrotnego***, a więc przeciwdziałający zakłócającym czynnikom otoczenia.

Wydaje się, że Ziemia w pewnym zakresie zachowuje się jak homeostat: utrzymuje w miarę stałą temperaturę powierzchni. Wzrost nasłonecznienia powoduje zwiększone parowanie wód, a narastająca ilość chmur zwiększa albedo atmosfery i redukuje ilość docierającego do Ziemi promieniowania. Ziemia więc skutecznie przeciwdziała wytrąceniu jej z termicznej równowagi, przynajmniej w obserwowanych przez nas granicach czasowych. Że nie każda planeta jest zdolna tak czynić, dowodzi przykład Wenus, gdzie w wyniku efektu cieplarnianego temperatura planety stale wzrasta i osiąga w chwili obecnej ok. 800 kelwinów.

Jednak również i w przypadku naszej planety nie można do końca być pewnym jej przyszłości. Czy w istocie istnieje ujemne sprzężenie zwrotne pomiędzy zanieczyszczeniami atmosfery a klimatem? W takiej sytuacji Gaja musiałaby usuwać nadmiar dwutlenku węgla, stabilizując klimat, nie dopuszczając do jego ocieplenia.

Jednak ostatnie badania dowodzą czegoś zupełnie przeciwnego. Analiza powietrza zamkniętego w lodach Grenlandii i Antarktydy pokazuje, że w ostatniej epoce lodowej stężenie dwutlenku węgla było o 30-40% mniejsze; było także dwa razy mniej metanu. Jeżeli życie miało jakiś wpływ na zmiany klimatyczne epoki lodowcowej, było to przyspieszenie zmian klimatycznych, a nie opóźnienie. Podczas zlodowaceń w atmosferze było mniej gazów cieplarnianych, co sprzyjało oziębieniu. Wraz z ogrzewaniem się Ziemi wzrastała zawartość metanu i dwutlenku węgla w atmosferze, a to jeszcze bardziej przyspieszało ogrzewanie. Z tego wynika, że może istnieć *dodatnie sprzężenie zwrotne* pomiędzy wydzielanymi do atmosfery gazami a zmianami klimatu [15].

Na ustalenie się dodatniego sprzężenia zwrotnego może wpłynąć także inne zjawisko, które powiększa ilość dwutlenku węgla w atmosferze: podwyższona temperatura sprzyja uwalnianiu się CO_2 ze skał, przez co jego zawartość w atmosferze szybko narasta (zjawisko takie ma miejsce na Wenus). W skrajnym przypadku, gdyby temperatura przy powierzchni Ziemi przekroczyła 100° C, nastąpi wyparowanie wszystkich zbiorników wody i zagłada życia [3].

Niektórzy wszak autorzy uważają, że mogły istnieć stosunkowo krótkie okresy, rzędu dziesiątków milionów lat, kiedy ustalało się stabilizujące ujemne sprzężenie zwrotne chroniące Gaję przed przegrzaniem. Mogło tak być w okresie szybkiego rozprzestrzeniania się dna morskiego, gdy ciepły i wilgotny klimat oraz duża ilość CO_2 sprzyjały gwałtownemu rozwojowi fitoplanktonu, który później, po śmierci organizmów, wiązał węgiel w osadach dennych [15].

Ustalenie charakteru sprzężenia zwrotnego pozwoliłoby przewidzieć kierunek zmian klimatycznych na Ziemi. Co więcej, mogłoby zaważyć na samej hipotezie Gai, jej przyjęciu bądź odrzuceniu.

Modelowanie takich procesów jest jednak trudne. Tym bardziej, że na efekt cieplarniany nakłada się proces przeciwdziałający – zapylenie atmosfery. Atmosfera działa w sprzężeniu z powierzchnią Ziemi. Promieniowanie słoneczne dochodzące do atmosfery jest częściowo odbijane, a częściowo przenika do powierzchni planety. Sama atmosfera także promieniuje – zarówno wysyła promieniowanie w kierunku powierzchni Ziemi, jak i w Kosmos. Powierzchnia Ziemi wysyła promieniowanie, które jest głównie pochłaniane przez atmosferę, ale częściowo przenika ją i ulatuje w Kosmos. Do tego dochodzi wymiana ciepła pod wpływem parowania i konwekcji.

Pyły zarówno przemysłowe, jak i naturalnego pochodzenia – wulkaniczne, piaski pustyni, pyły meteorowe zwiększają pochłanianie promieniowania słonecznego w atmosferze, powodując ochłodzenie, a tworząc jądra kondensacji dla pary wodnej – zwiększają ilość opadów. Pyły, osiadając na powierzchni planety – zwiększają jej *albedo* – zdolność do odbijania promieniowania. Największe albedo ma świeży śnieg – ok. 85%, piaski – ok. 30%, a lasy i łąki – 10-15%. Te czynniki – mniej słonecznego ciepła, wzrost opadów, zwiększone albedo – przy dodatnim sprzężeniu zwrotnym mogą doprowadzić do całkowitego zlodowacenia Ziemi i wyginięcia wszelkiego życia [3].

Żaden z tych „czarnych scenariuszy" nie musi się wydarzyć, jednak nieznajomość charakteru sprzężeń zwrotnych w obu przypadkach pozwala obawiać się obydwu kataklizmów. Teoretycznie – możliwe są obydwa kierunki przebiegu obu procesów, jak również ich wypadkowa. Z tego też względu Schneider postuluje mówienie nie o sprzężeniu zwrotnym, a raczej o *koewolucji* życia i klimatu [15].

Obliczając ewolucję przykładowego klimatu („*zerowymiarowy model cieplarniany*"), otrzymujemy w przypadku „białej Ziemi" (dużej ilości śniegu, dużego albedo) temperatury przy powierzchni poniżej -3° C, a w przypadku „zimy nuklearnej" (zanieczyszczenia przez dymy i pyły) – spadek temperatury średnio o ok. $4{,}4^{o}$ C. Natomiast pokrycie 1/3 powierzchni Ziemi bateriami słonecznymi zmniejszyłoby albedo zaledwie o 0,01%, co spowodowałoby wzrost temperatury przy powierzchni Ziemi o 0,2 stopnia [2].

Jak z powyższych rozważań wynika, biotyczne i abiotyczne cykle obiegu pierwiastków (szczególnie węglowy), zakłócane przez działania antropogeniczne, mogą w sposób istotny odbić się na klimacie i o ile ustali się dodatnie sprzężenie zwrotne, mogą stanowić przyczynę kataklizmu biologicznego, a nawet zagrozić istnieniu życia na Ziemi. Jest to zagrożenie globalne, postępujące niejako od góry w dół (od stanu atmosfery – do poszczególnych ekosystemów).

2. Związki międzygatunkowe

Innym zagrożeniem, nie mniej groźnym, choć wciąż niedocenianym, są działania degradacyjne postępujące w przeciwnym kierunku, polegające na unicestwianiu przez człowieka poszczególnych osobników i w konsekwencji gatunków roślin i zwierząt, co ma wpływ globalny i odbija się na funkcjonowaniu całej biosfery.

Jakkolwiek optymistycznie nie spojrzelibyśmy na ostateczne możliwości adaptacyjne Gai, to nie ulega kwestii, że dotychczasowa i obecna działalność człowieka jest przyczyną istotnych zmian w biosferze – zmniejszania się populacji i wymierania ostatnich osobników gatunków roślin i zwierząt.

Na Ziemi żyje coraz mniej gatunków, zmniejsza się ich *różnorodność*.

Na problem niezwykłej różnorodności gatunków bytujących na Ziemi pierwszy zwrócił uwagę Karol Darwin w dziele: *O powstawaniu gatunków drogą doboru naturalnego. „Dlaczego życie na Ziemi odznacza się tak wielką różnorodnością? Dlaczego jest tak wiele (a właściwie: ile?) gatunków? [...] Czy musi być akurat tyle gatunków? Czy mogłoby być mniej albo więcej? I czy w ogóle potrafimy odpowiedzieć na te pytania?*" [18].

Określenie *bioróżnorodności* narodziło się na konferencji w Waszyngtonie we wrześniu 1986 roku (*Forum on BioDiversity).* Odtąd zagrożenie bioróżnorodności stało się synonimem zagrożenia całej żywej Przyrody.

Jako najważniejsze szkody wyrządzane biosferze wymienia się: 1) bezpośrednie tępienie zwierząt i roślin, 2) wprowadzanie obcych gatunków, 3) niszczenie środowiska – dewastacja siedlisk, 4) fragmentacja siedlisk (wciąż niedoceniane), 5) zanieczyszczanie środowiska. Najgroźniejsza jest *synergia*, czyli nałożenie się tych wszystkich czynników [15].

Głównymi „winowajcami" zanieczyszczania środowiska są przemysł i urbanizacja, a także rolnictwo. Przemysł znalazł się na cenzurowanym stosunkowo wcześnie. Pyły oraz gazy emitowane przez fabryki powodują zmianę stanu atmosfery, zwiększają pochłanianie promieniowania słonecznego oraz efekt cieplarniany.

Dłużej niedocenianym elementem degradacji środowiska był rozwój miast. Okazało się jednak, że degradacja środowiska przez urbanizację jest nie tylko lokalna (co było oczywiste), ale także globalna: zdrenowanie gruntów wokół miast wiedzie do obniżenia poziomu wód gruntowych i stepowienia dużych obszarów, a cieki wodne przenoszą trujące substancje do mórz. Wskutek tego giną gatunki roślin i zwierząt nie tylko w rejonach zurbanizowanych, ale na oddalonych od nich obszarach (np. w deltach rzek). Istotne jest także ogrzewanie się atmosfery; nad miastami temperatura powietrza podnosi się nawet o kilka stopni.

Jednym z najbardziej degradujących środowisko czynników było i jest rolnictwo. Pasterstwo przed wiekami, a współcześnie intensywna hodowla (na świecie jest ok. 30 miliardów zwierząt hodowlanych) i związane z nią działania i procesy (wypalanie lasów pod pastwiska, uwalnianie się do atmosfery wielkich ilości dwutlenku węgla i metanu produkowanego przez żołądki przeżuwaczy) w sposób globalny zmieniają stan atmosfery i niszczą cienką warstwę ozonu (O_3) w stratosferze, która przepuszcza coraz więcej słonecznego ultrafioletu. Pewne wahania stężenia ozonu mają swe źródło również w zjawiskach naturalnych [6], jednak działania antropogeniczne znacznie nasilają ten proces, co jest przyczyną tego, że odbudowa warstwy ozonowej staje się znacznie utrudniona.

W wyniku wzrostu natężenia promieniowania ultrafioletowego ginie morski fitoplankton, wytwarzający 80% tlenu na świecie. Ocenia się, że zmniejszenie ilości ozonu o 10% stanie się przyczyną ok. 45% wzrostu uszkodzeń wywołanych przez średniofalowe promieniowanie ultrafioletowe UV-B [2]. Wtedy promieniowanie może uszkadzać nawet kiełki roślin i mniejsze gatunki ryb (ultrafiolet jest wysokoenergetyczny, ale ma niewielką przenikliwość).

Autorzy nie są zgodni co do ilości obecnie żyjących na Ziemi gatunków: w zależności od metody rachunków i poczynionych założeń oceny wykazują rozpiętość od 5 do 50, a nawet 80 milionów. Tę ogromną różnorodność osiągnęła biosfera w okresie od 3,9 – 3,4 miliarda lat temu do chwili obecnej i rozwój ten przebiegał według krzywej wykładniczej. Ocenia się, że od czasów Karola Linneusza zarejestrowano w muzeach od 1 do 1,8 miliona gatunków organizmów wyższych. Rokrocznie opisywane są nowe gatunki, nawet pośród najbardziej znanych grup zwierząt – przeciętnie 15 gatunków ssaków i 3 gatunki ptaków (!). Mało zbadane są także rośliny puszczy równikowej i zwierzęta głębin – ocenia się, że i tam wiele jeszcze gatunków pozostaje nieznanych [18].

W 1982 roku T. L. Erwin dla zgrubnego szacunku ilości gatunków dokonał wątpliwego etycznie czynu: opylił środkiem owadobójczym korony 19-tu drzew i policzył uśmiercone owady: było wśród nich 1200 gatunków chrząszczy. Na podstawie tego ocenił liczbę wszystkich stawonogów w lasach równikowych na ok. 30 milionów [18], [1].

Z danych paleontologicznych wynika, że w minionych epokach następowały po sobie zarówno spadki, jak i eksplozje bioróżnorodności. Epoki geologiczne wyróżnione zostały na podstawie występujących w nich gatunków, więc należy przyjąć, że każde przejście do kolejnej epoki to gwałtowne zmiany świata organicznego: wymieranie jednych gatunków, a potem rozpowszechnienie się innych.

Od początku paleozoiku (540 miliony lat temu) zarejestrowano 15 okresów wielkiego wymierania. Uważa się, że największym wymieraniem był przełom permu i triasu (przejście z paleozoiku do mezozoiku, ok. 245 milionów lat temu), kiedy wyginęło 96% wszystkich gatunków i kolejno przełom kredy i trzeciorzędu (przejście z mezozoiku do kenozoiku, przed ok. 65 milionami lat), kiedy znikło ich 70%, między innymi osławione dinozaury.

Fakty te tłumaczy się przyczynami kosmicznymi, głównie wielkimi zderzeniami (świadczą o tym znajdowane w osadach mikrotektyty i iryd występujący w meteorytach), wulkanicznymi albo tektonicznymi.

Obecnie jesteśmy świadkami kolejnego wielkiego wymierania, jednak spowodowanego nie przez czynniki kosmiczne czy geologiczne, ale przez jeden gatunek – *Homo sapiens.* Wymieranie plejstoceńsko-holoceńskie, zapoczątkowane kilkadziesiąt tysięcy lat temu (!) wskazuje na człowieka jako głównego winowajcę. Człowiek, zgodnie z ostatnimi badaniami, narodził się w Afryce i ok. 50 tysięcy lat temu zasiedlił Australię i Nową Gwineę, 20 tysięcy lat temu – Syberię, 11 tysięcy lat temu – obie Ameryki, 4 tysiące – wyspy Pacyfiku. Każdej ekspansji towarzyszyło wytępienie miejscowych gatunków (w Australii człowiek wytępił 86 % gatunków ssaków, w Ameryce Południowej – 80%, w Ameryce Północnej – 73%) [18].

Tępienie gatunków w czasach prehistorycznych następowało jedynie z użyciem prymitywnej broni, a już wtedy osiągnęło rozmiary porównywalne do wielkich katastrof geologicznych. Eskalacja nastąpiła, gdy ok. 10 000 lat temu człowiek zaczął uprawiać rolę, a osiągnęła jeszcze większy poziom w wyniku rozwoju przemysłu kilka wieków temu. Giną przede wszystkim gatunki dużych i stadnych roślinożerców oraz będące z nimi w łańcuchu pokarmowym drapieżniki i padlinożercy. Ich różnorodność zmniejsza się, co ma głębokie znaczenie: zrywają się delikatne łańcuchy współegzystencji pomiędzy poszczególnymi gatunkami i ekosystemami.

Tempo tego wymierania nie jest dokładnie znane, ale ocenia się, że jest to ok. 50 gatunków dziennie. Według niektórych autorów bioróżnorodność jest obecnie taka, jakbyśmy się cofnęli w historii ewolucji ok. 60 milionów lat! [5].

Rodzi się niebagatelne pytanie, jaką wartość stanowi bioróżnorodność? Czy biosfera będzie funkcjonowała inaczej (gorzej), jeśli różnorodność się zmniejszy? Dla kogo gorzej – dla człowieka, dla innych gatunków czy dla samej siebie jako całości? Intuicyjnie wydaje się to oczywiste, jeśli wziąć pod uwagę łańcuchy związków i zależności pomiędzy poszczególnymi gatunkami, łańcuchy zależności nie tylko pokarmowych, ale wszystkie inne, dotąd tym bardziej nieznane i niezbadane, że nieznana jest ani liczba wszystkich gatunków, ani zasady ich bytowania.

Dla oceny konsekwencji zmniejszania się bioróżnorodności próbuje się stosować jej najróżnorodniejsze modele, zarówno matematyczne, jak i opisowe. Jak dotąd żadne z otrzymanych wyników opartych na którymkolwiek modelu nie wydają się wiarygodne.

J. H. Lawton w swej *hipotezie idiosynkrazji* twierdzi, że wpływ bioróżnorodności na funkcjonowanie biosfery jest nieprzewidywalny. Z tego też względu niektórzy autorzy decydują, że nie ma *„żadnych przesłanek, aby rozważać znaczenie różnorodności gatunkowej dla biosfery"* [18]. Jej znaczenie rozpatrywane w kategorii wartości dla człowieka to przede wszystkim dostarczanie pokarmu, leków i paliw. Bardziej wysublimowane wartości dotyczą przyjemności obcowania z Przyrodą, badania ścieżek ewolucji, co jest niemożliwe w zubożonym środowisku. Czy jednak można tak beztrosko ograniczać się tylko do stanowiska antropocentrycznego, rozpatrując wartość bioróżnorodności jedynie dla człowieka?

Cybernetyka poucza, że stabilne funkcjonowanie układu jest bardziej prawdopodobne wtedy, kiedy układ cechuje *różnorodność* [16]. Różnorodność okazała się także cechą konieczną dla ewolucji Wszechświata i osiągania przez niego coraz wyższych szczebli złożoności [7]. Niektórzy filozofowie, jak Leibniz w swojej *Monadologii,* intuicyjnie dostrzegali konieczność istnienia różnorodności w Świecie, ażeby mógł się rozwijać, był piękny (doskonały) i można go było opisać [10].

Jeśli *stabilność* to dysponowanie procesami i siłami, które są w stanie przywracać stan pierwotny, to z większym prawdopodobieństwem możemy jej oczekiwać w przypadku *różnorodności* układu, czyli *posiadania rozmaitości cech i form oraz bycia złożonych z wielkiej ilości elementów* [14]. Oczekiwać więc można, że po zaburzeniach biosfera powraca łatwiej do pierwotnego stanu, gdy jest bardziej różnorodna, a zubożona – funkcjonuje gorzej. Gorzej „dla siebie", ale przecież także dla gatunków na nią się składających, w tym i dla człowieka.

Podaje się przykłady zubożonych środowisk, które łatwo ulegają degradacji. Należą do nich monokultury (np. leśne), ale takim przykładem może być także eksperymentalna *Biosfera 2,* która była na tyle uboga, że w stosunkowo krótkim czasie utraciła stabilność [18].

Inni autorzy uważają, że są gatunki mniej i bardziej znaczące – *tzw. gatunki zwornikowe,* jak nazywa je R. T. Paine, których brak wpływa w istotny sposób na funkcjonowanie biosfery. Według słynnej już teorii P. R. i A. Ehrlichów *wypadających nitów*, dopiero po utraceniu określonej ich liczby – jak w przypadku nitów wypadających z konstrukcji samolotu – następuje katastrofa [4].

Jednak, jak to zauważa C. Maser, również odwołując się do lotniczej analogii, nigdy nie można przewidzieć, który z drobnych elementów („nitów") okaże się nie do zastąpienia. Najważniejsza w śmigłowcu wydaje się być sprawność silnika, a uległ on katastrofie z powodu wypadnięcia maleńkiej drucianej zawleczki, zabezpieczającej zsunięcie się nakrętki śruby mocującej boczny statecznik [14].

Są wreszcie teorie, jak modele matematyczne Roberta M. Maya, pokazujące, że w pewnych sytuacjach różnorodność może nawet pogorszyć funkcjonowanie układu i jego stabilność [18].

Sądzę, że metafora nitów, jak również wszelkie „mechaniczne" modele nie są tutaj odpowiednie i przywodzą na myśl Kartezjuszowy błąd mechanicyzmu. Ani biosfera, ani poszczególne żywe organizmy nie funkcjonują tak, jak skręcone ze sobą śrubami sztywne części mechanicznego urządzenia. Podstawową cechą struktur żywych jest *elastyczność,* pozwalająca na modyfikację zachowań, zmienność, przystosowanie się. Dlatego bardziej adekwatnym byłby *model elastycznej sieci,* w której węzłami są poszczególne gatunki, a łączącymi je nićmi – zależności pomiędzy nimi. Zależności tych może być wiele i jeśli istnieje różnorodność gatunków, jest większa szansa, że przy „wypadnięciu któregoś węzła" nici naciągną się, a pozostała sieć węzłów nie dopuści do zachwiania się funkcjonowania całego systemu. Gdy węzłów – gatunków będzie mniej, wypadnięcie któregoś łatwiej doprowadzi do zerwania naciągniętych nici sieci, pojawienia się „dziury" i do katastrofy.

Zadając pytanie o wagę bioróżnorodności dla ogromnego organizmu Gai, należałoby się zastanowić, czy nie jest właściwym wpisać wydarzenia katastroficzne jako konieczne i naturalne w historię ziemskiej biosfery, nie martwiąc się szczególnie rolą człowieka jako sprawcy? Hipoteza Gai, która była często postrzegana jako wsparcie postawy ekologicznej i służyć miała idei ochrony środowiska, w istocie uzasadnia *paradygmat ewolucyjny* [8]. To prawda, że ingerując w najdrobniejszy ekosystem biosfery, jednocześnie zaburzamy związki całego systemu, jednak skoro Gaja ma możliwość autoregulacji i samonaprawy, to czyż człowiek powinien się martwić o jej ochronę? Być może Gaja jako homeostatyczną całość zdoła się obronić. Następna epoka lodowa pokryje ziemię na kilka tysięcy lat grubą warstwą lodu, gnijące szczątki padłych roślin i zwierząt nawiozą glebę w naturalny sposób, a jałowe ziemie po ustąpieniu lodowców znów porosną bujną roślinnością.

Odwoływanie się do możliwości autoregulacyjnych Gai stawia pod znakiem zapytania sensowność większości działań pro-ekologicznych. Może zrodzić wątpliwości, czy taki stosunek człowieka do Gai nie jest nazbyt „opiekuńczy" (jak to widziała Margulis) i nie należałoby potraktować wszelkich zniszczeń cywilizacyjnych jako przejawy wpływu na środowisko jednego z wielu gatunków, a więc element naturalny?

Stanowisko takie, zrodzone – paradoksalnie – z przekonania o homeostatycznych własnościach Gai, traktuje człowieka nie jako potencjalnego obrońcę, ale jako szkodnika, z którym Gaja musi sobie radzić, rozkładając produkty jego metabolizmu, a być może niszcząc w końcu i jego samego.

Inne jest stanowisko *ekologii głębokiej* i *etyki środowiskowej*, traktujących bioróżnorodność jako podstawową cechę Natury, stanowiącą wartość samą w sobie, niezależnie od jej użyteczności dla człowieka.

Wymieranie gatunków spowodowane działalnością człowieka to zachwianie naturalnego procesu ewolucji życia. Dostrzeżenie tego problemu zrodziło przekonanie o konieczności podjęcia *odpowiedzialności człowieka za ewolucję*

i traktowania tego problemu jako podstawowego dla etyki środowiskowej [9].

Te fakty pozwalają traktować bioróżnorodność nie jako niczemu nie służącą obfitość Przyrody, ale jako cechę immanentną Natury i jej *wartość witalną* [17]. Bioróżnorodność jest najważniejszym i najwyższym etapem rozwoju środowiska kosmicznego, tworzonym przez Wszechświat przez całe miliardy lat rozwoju. Różnorodności i złożoności takiej, jaka występuje w ziemskiej biosferze, nie znajdziemy w żadnym ze znanych zakątków Kosmosu. Dla jej powstania było konieczne różnicowanie się materii i wspinanie się jej na wyższe stopnie złożoności przez całe miliardy lat ewolucji Wszechświata.

Obecnie zarysowują się dwa zasadnicze kierunki ratunku dla biosfery, próbujące powstrzymać degradację środowiska. Jeden z nich – nazwę go *drogą umiarkowania* – nawołuje do powstrzymania ekspansji ludzkości i zmniejszenia wybujałej konsumpcji krajów rozwiniętych, co znalazło wyraz m.in. w *idei zrównoważonego (trwałego) rozwoju, ochronie Przyrody, wegetarianizmie*, promocji *energii odnawialnej*.

Rozwinięte technologie przyczyniają się do wzrostu konsumpcji dóbr. Produkowanie nowoczesnego sprzętu, zużywającego nieco mniej energii niż sprzęt dawniejszy, nie równoważy jego masowej dostępności i wzrostu produkcji. Zużywanie energii w społeczeństwach przemysłowych, czyli *metabolizm społeczny*, ciągle wzrasta. Ocenia się, że w społeczeństwach rolniczych jest on 5 razy większy niż u łowców i zbieraczy, a w przemysłowych – aż 20 razy większy. Oblicza się, że masa produktów zużywanych przez jednego mieszkańca w ciągu roku wynosi w Japonii ok. 17 ton, w Niemczech i Austrii – 29 ton, a Stanach Zjednoczonych – aż ok. 70 ton!

Na wykorzystywane przez człowieka produkty składa się przede wszystkim energia produkowana przez rośliny w procesie fotosyntezy (współcześnie – tlen, pokarm roślinny, oraz w przeszłości – paliwa kopalne). Ocenia się, że pod koniec XX wieku człowiek zawłaszczał ponad 25% energii wyprodukowanej przez rośliny. Nasza niezwykła zależność od świata roślinnego, z której najczęściej zupełnie nie zdajemy sobie sprawy, jest zbyt duża, tym bardziej, że człowiek w szybkim tempie niszczy roślinne zasoby planety (czy to wycinając puszczę amazońską, czy wyniszczając słonecznym ultrafioletem fitoplankton morski poprzez zaburzanie warstwy ozonowej).

W związku z nadmierną eksploatacją biosfery pojawiła się idea *trwałego (zrównoważonego) rozwoju*, spopularyzowana na „Szczycie Ziemi" w Rio de Janeiro w 1992 roku. Jej celem jest powstrzymanie wzrostu metabolizmu społeczeństw przemysłowych.

Stosowanie energii odnawialnych – wiatru, spadającej wody, promieniowania słonecznego, pływów jest najlepszą drogą do bardziej efektywnego – z punktu widzenia termodynamiki – wykorzystania energii Słońca, bez nadmiernego zawłaszczania jej roślinom. Drogą – co niezwykle ważne – zarazem praktycznie nieszkodliwą dla środowiska. Jak obliczono bowiem, korzystając z *zerowego modelu cieplarnianego,* pokrycie 1/3 powierzchni kuli ziemskiej bateriami słonecznymi spowodowałoby wzrost średniej temperatury powierzchni Ziemi zaledwie o ok. 0,2^{o} [2].

Drugi kierunek pokłada nadzieję w nowoczesnych technologiach, które niczym *deus ex machina* mają naprawić błędy cywilizacyjne wyrządzone przez wieki. Przekonania takie uważam w dużej mierze za dowód naiwności i braku świadomości istnienia nieznanych jeszcze, delikatnych związków pomiędzy gatunkami i ekosystemami. Czy bowiem manipulacja środowiskiem, owe zdolności wynoszące nas ponad inne gatunki, są w stanie zniweczyć i odwrócić szkody wyrządzane biosferze przez tysiąclecia i przywrócić właściwe funkcjonowanie Gai? Wiedza o jej procesach jest wciąż powierzchowna i wyrywkowa. Jest to przyczyną, dla której nie udaje się przewidzieć wielu skutków wpływów na środowisko. Znamy tego liczne i tragiczne przykłady. Czy może się powieść klonowanie i introdukowanie ginących gatunków do zubożonego środowiska? Na ile taka introdukcja mogłaby zaburzyć jego relacje z innymi gatunkami? Czy nie jest naiwnością sądzić, że działanie takie się powiedzie, kiedy zostały zerwane zależności pomiędzy nimi a innymi, nieznanymi może nawet, wytępionymi gatunkami? Próba odpowiedzi na te pytania powinna się oprzeć na *modelu elastycznej sieci* związków pomiędzy gatunkami.

Świadomość, że Gaja przeżywała niejednokrotnie wielkie kataklizmy, pozwala mieć nadzieję, że być może znów powróci do stabilności i przetrwa. Na ile zubożone będzie jednak przyszłe ziemskie środowisko? Czy przetrwa człowiek? Oto pytania, na które nie umiemy odpowiedzieć.

Literatura cytowana

1. Attenborough D.: Szanujmy Ziemie. Przesłanie Dawid Attenborough, Editor. Jo Payne, Produced R. Barringtyon, BBC Discovery, 2000.
2. Boeker E., Rienk van Grondelle: Fizyka Środowiska, tłum. K. Bajer, A. Kopystyńska, T. Szoplik, Wydawnictwo Naukowe PWN, Warszawa 2002.
3. Dworak Z. T.: Porównanie atmosfer ciał typu ziemskiego w Układzie Słonecznym dla celów inżynierii środowiska planetarnego, [w:] Inżynieria Środowiska, t.7, z.3, 523 (202), 2002.
4. Ehrlich P. R., Ehrlich A. H.: Extinction, Random House, New York 1981.
5. Kendall H. W.: Globalne perspektywy [w:] Czy nauka jest dobra?, red. M. Moskovits, tłum. J. i M. Jannaszowie, Wydawnictwo Cis, Warszawa 1997.
6. Korpikiewicz H.: Kosmoekologia z elementami etyki holistycznej. Hipoteza Gai-Uranosa, Prodruk, Poznań 2002.
7. Korpikiewicz, w druku. Kosmoekologia. Obraz zjawisk
8. Kośmicki E.: Podstawowe problemy etyki ekologicznej [w:] red. B. Andrzejewski, Humanistyka i ekologia, Fundacja Warta, Poznań 1992.
9. Kośmicki E.: Odpowiedzialność człowieka za ewolucję. Podstawowy problem etyki środowiskowej, [w:] Człowiek – Zwierzę – Cywilizacja. Aspekt humanistyczny, red. H. Korpikiewicz, Prodruk, Poznań 2001.
10. Leibniz G. W.: Monadologia, tłum. H. Elzenberg, Uniwersytet Mikołaja Kopernika, Toruń 1991.
11. Lovelock J. E.: GAIA. A New Look at Life on Earth, Oxford University Press, Oxford 1982.
12. Lovelock: The Ages of Gaia: A Biography of our Living Earth, Oxford University Press, Oxford 1988.
13. Margulis L.: Symbiotyczna planeta, tłum. M. Ryszkiewicz, Wydawnictwo CiS, Warszawa 2000.
14. Maser C.: Nowa wizja lasu, tłum. J. P. Listwan, J. Majewski, Stowarzyszenie Pracownia na Rzecz Wszystkich Istot, Bystra k/Bielska-Białej 2003.
15. Schneider S.H.: Laboratory Earth. The Planetary Gamble, We Can`t Afford To Loose, Orion Publishing Group Ltd., 1997.
16. Trąbka J.: W uniwersalnym kręgu cybernetyki, Oficyna Wydawnicza Abrys, Kraków 2001.
17. Wawrzyniak J.: Teoretyczne podstawy neonaturalistycznej bioetyki środowiskowej, Pisma Filozoficzne, t.73, Wydawnictwo Naukowe Instytutu Filozofii UAM, Poznań 2000.
18. Weiner J.: Życie i ewolucja biosfery, wyd. II poprawione i unowocześnione, Wydawnictwo Naukowe PWN, Warszawa 2003.
19. Wiernadski W.I.: Biosfera. Izbrannyje trudy po biogeochemii, Moskwa 1967.

BIO-ALGORITHMS AND MED-SYSTEMS
JOURNAL EDITED BY MEDICAL COLLEGE – JAGIELLONIAN UNIVERSITY
Vol. 1, No. 1/2, 2005, pp. 381-383.

O RELACJI LICZBY DO RZECZY

THE *NUMBER-THING* RELATION

TERESA GRABIŃSKA, MIROSŁAW ZABIEROWSKI

Instytut Filozofii, Uniwersytet Zielonogórski, Zielona Góra

Streszczenie. We współczesnej nauce występuje graniczne uabstrakcyjnienie opisu liczbowego i jego relacji do rzeczywistości będącej przedmiotem danej dyscypliny naukowej. Dla zrozumienia tego problemu konieczne jest odtworzenie tej rzeczywistości, do której odnosi się granicznie uabstrakcyjniony opis. Pokazujemy, że w ramach realizmu krytycznego możliwe jest określenie statusu realności owej rzeczywistości jak również ustalenie, na ile przeniesienie granicznie kwantytywnego opisu rzeczywistości przyrody nieożywionej jest limitujące w obszarze przedmiotów biologicznych, co podnosił także J. Trąbka.

Słowa kluczowe: ontologiczny status liczby, ilość a jakość, przedmiot matematyczny a przedmiot zjawiskowy

Abstract. Present in modern science is the extreme abstracting of quantitative description and the extreme abstracting of its relation to reality which is ultimately the subject of the given science. The first step in resolving the problem of quantitative abstraction is the reconstruction of the reality which is in fact the subject of such a description. In the frame of critical realismus we show the possibility to define the status of this reality as well as to settle the restricting character of extreme quantitative description in the field of biological objects.

Key words: onthological status of number, quantity and quality, mathematical object and phenomenal object

„Po mózgu zbyt wiele błąkało się idei pustych,
Które nie miały żadnej konkretnej podkładki.
To przeidealizowanie neurologii było
rezultatem
wprowadzenia szerokim nurtem formalizmu
matematycznego do mózgu".
J. Trąbka

„a całe niebo jest harmonią i liczbą"
Arystoteles, *Metafizyka*, A

Pitagorejska ontologia przedmiotów (rzeczy) matematycznych, które u pitagorejczyków sprowadzały się wyłącznie do liczb, ich właściwości i relacji między nimi, upatruje bytowości liczb w oderwaniu od przedmiotów zmysłowych (zjawiskowych) i ich niezmienności – „z wyjątkiem przedmiotów astronomicznych"[1], co zawsze podreślaliśmy za Arystotelesem. W pitagorejskiej ontologii liczby (przedmioty arytmetyczne) są źródłem zasad wszelkiej rzeczywistości: poszczególne właściwości liczb są przejawami rzeczywistości, którą inni nazywają idealną lub transcendentalną, duchową, przyrodniczą. Prymat liczby z jednej strony usposabiał pitagorejczyków do *czystych*[2] badań teoretycznych przedmiotów liczbowych (teorii liczb i stosunków liczbowych), z drugiej zaś strony podporządkowywał ogląd każdej innej rzeczywistości wzorowi rzeczywistości arytmetycznej. I nie tylko faworyzował ilościowy (matematyczny) ogląd (tzn. kształt poznania), ale przede wszystkim ignorował rzeczywistość zjawiskową, ignorował teorię przyrody. Wszelkie obserwowane przypadłości czy regularności nie miały bytowego znaczenia, a więc nie mogły stać się punktem wyjścia poznania. Liczba jest początkiem i źródłem wszystkiego. Co jest więc przyczyną liczby – pyta Arystoteles[3]. Jeśli to pytanie zlekceważyć, to ontolog ma obowiązek zapytać o podstawę samoistności liczby.

Platon kwestię samoistności liczby próbował rozwiązać, dzieląc liczby na rozumowe (które są przyczynami) i inne, które są przedmiotami zjawiskowymi. Te pierwsze miałyby bytowość idealną. W tym jednak rozwiązaniu – jak już zauważył Arystoteles[4] – nie da się np. wyeksplikować wyraźnej różnicy między rodzajami liczb, jak i między rodzajem liczby a stosunkami liczb; nie można ustalić

[1] Por. Arystoteles, *Metafizyka*, tłum. K. Leśniak, A. 8, PWN, Warszawa 1984.

[2] Termin „czysty" jest przez nas wyróżniony kursywą, która ma podkreślić specyficzność owej czystości w porównaniu ze współczesnym programem czystej matematyki (np. francuska grupa Bourbaki). U pitagorejczyków poszukiwanie odpowiedniości między liczbami i np. przedmiotami geometrycznymi (przedmioty geometryczne w innych klasyfikacjach są zaliczane do przedmiotów matematycznych, u pitagorejczyków były wtórne wobec liczb) i zjawiskowymi odbywało się na sposób specyficznie empiryczny. Por. G. Sarton, *A History of Science. Ancient Science through the Golden Age of Greece*, Cambridge 1959 (za Kazimierzem Leśniakiem: przypisy do *Metafizyki* Arystotelesa, *op. cit*., s. 396).

[3] Arystoteles, *Metafizyka*, op. cit., A. 8, s. 29; dalsza część krytyki pitagorejskiej koncepcji liczby (a także platońskiej,również w Księdze N) występuje w Księdze M.

[4] *Ibidem*, A. 9.

tożsamości jednostek składających się na liczbę; nie można z nieruchomości liczby (czy idei) wyprowadzić ruchu.

W swojej realistycznej ontologii substancji Arystoteles określił relację liczby do definicji. Definicja w części podmiotowej „musi odgrywać rolę materii", a w części orzecznikowej – „formy"[5]. Pokazał też, w jakim sensie należy porównywać definicje z liczbą: 1) Ze względu na naturę podzielności – formuła definicyjna, jak i liczba są podzielne na skończone części; 2) Ze względu na rolę braku lub nadmiaru części – brak lub nadmiar części zmienia zarówno liczbę, jak i definicję; 3) Ze względu na niemożliwość ustalenia absolutnej jedności tak liczby, jak i definicji – podstawą jedności jest jedność substancji; 4) Ze względu na rozumienie relacji „więcej i mniej" odnośnie do liczby i definicji – tak jak konkretna liczba nie dopuszcza owej relacji, tak nie dopuszcza jej „substancja w znaczeniu formy", lecz relacja ta jest naturalna dla substancji w jej składowej materialnej. Składowa materialna substancji dopuszcza więc różne liczby, co sprawia, że liczbowy opis substancji (bytu) dotyczy wyłącznie składowej materialnej[6].

Realistyczna ontologia Arystotelesa wymagała jasnego określenia, czy przyjmowana w innych ontologiach za realną substancja idealna bądź matematyczna jest samoistna i w jakiej pozostaje relacji do hylemorficznej substancji („substancji zmysłowej", rzeczywistości zjawiskowej). Stagiryta podjął szeroko ten temat w Księdze M. *Metafizyki* i wykazał, że przedmioty matematyczne nie mogą istnieć w sposób absolutny, bowiem nie istnieją „jako oddzielne substancje ani w rzeczach zmysłowych, ani niezależnie od rzeczy zmysłowych". Przedmioty matematyczne wszakże istnieją jako „abstrakcje", jako przedmioty umownie „odłączone" od bytu realnego („entelechii"), ale faktycznie z nim związane poprzez odniesienie do potencjalności materialnej. Wszelkie zaś twierdzenia matematyczne, jeśli wyrażają jakąś prawidłowość dla przedmiotów matematycznych, to ta prawidłowość równocześnie obowiązuje w rzeczywistości zjawiskowej. Nie są jednakże w żadnym razie przedmioty matematyczne przyczyną w każdym przez Arystotelesa wyróżnionym sensie „w odniesieniu do pierwszych zasad", nie są przyczyną formalną[7].

Liczba byłaby zatem odniesiona: 1. do racjonalności i poznania i 2. do przyrody[8]. Także w tym fundamentalnym ujęciu ontologii liczby zbliżamy się do problemu zbadania logicznej konieczności ziarnistości od strony zasady kosmologicznej (traktowanej jako teza ontologiczna[9]), skoro realnie dana rzeczywistość galaktyczna i pozagalaktyczna jest ziarnista. Wszelka kosmogonia i kosmologia nieziarnista byłaby wadliwa. Substrat kosmologiczny musi być, i to ziarnisty, musi zawierać *mirę*, a nie astrometrycznie niepomiarowe obiektowe *nice*. Astrometryczna mira pozwala dokonywać pomiaru – nie chcemy powiedzieć, że uprawiamy opomiarowanie bez substancji, substratu i bez miry, czyli ziarnistości.

Przedmiot zjawiskowy przedstawia się również w *stosunkach liczbowych*. Podstawą (korzeniem) liczby jest przyroda, zatem obiekty, zjawiska *odtąd-dotąd*. Inaczej jest u przedarystotelików, których krytykował Arystoteles, za to, że liczbę umieszczają między niebieskiem a rzeczami: „Mam na myśli to, że ci myśliciele [tzn. platończycy] umieszczają przedmioty matematyczne między Ideami a rzeczami zmysłowymi jako trzeci rodzaj rzeczy, niezależnie zarówno od Idei, jak i od rzeczy w tym świecie; ale przecież nie istnieje według nich ani trzeci człowiek, ani trzeci koń"[10].

Heterogeniczność świata to m.in. konsekwencja liczby. Zasada względności (ZW) Kopernika (mimo złamania przez heliocentryzm w kosmologii Arystotelesa zasady substancjalności przestrzeni przez uczynienie Ziemi, substancjalnie różnej od planet, im równą) „kontroluje" (oddaje, ale jak? – przez nadzór) liczby – fakt, obserwację, sens zliczenia. Zliczenie to stwierdzenie – stwierdzenie czegoś, obrotu nieba, a więc obiektu, skoro nie ma obrotu bez miry, minimum kosmologicznej (pomiaru całości) ziarnistości. To minimum kosmologiczne implikuje problem kosmogenezy, czyli koncepcję Kanta z 1755 r., który chciał rozwiązać chaosem genezę[11].

Nie o to chodzi, że coś mniemamy o samych liczbach - wszak nie możemy ani konkretnej liczby, ani sensu liczby, oderwać od rzeczywistości zjawiskowej (przyrodniczej). A przyroda jawi się poprzez jakości (mierzone intensywnością jakościowego różnicowania) i przez ilości (wyznaczane w pomiarze). ZW (odnosząca poszczególne rzecz i ich opisy do siebie wzajemnie) nadaje pomiarowi, czyli faktowi, odniesienie do zaktualizowanego istnienia. Liczba jest zakotwiczona w bycie a dlatego pozwala indywiduom być mierzonymi, bo wymaga realnej (nieabstrakcyjnej) miry, substancjalności dla mierzenia. To pociąga za sobą nie tyle abstrakcyjne podobieństwo obiektów zliczanych, ile ostatecznie pociąga za sobą problem kosmogenezy, czyli powstania rzeczy. Gdy się odrywa liczby od rzeczy, jak w mechanicyzmie czy w bigbangu, to należałoby przyjąć, że kosmogenezy nie było. Nie chodzi o to, że liczba zdaje relację z liczności obiektów, lecz że zadaje faktualność, epizodyczność od A do B, a nie trwa-

[5] *Ibidem*, H. 3.

[6] *Ibidem*.

[7] *Ibidem*, N.

[8] Por. J. Trąbka, *Gnoza a nauki neurologiczne*, Wyd. Atykwa, Kraków 1997. Autor w rozdz. *Rzeczywistość wirtualna mózgu* pragnie nie tyle zbudować, ile zaznaczyć możliwości nie tyle zbudowania ile rozwijania refleksji nad inteligencją. Podkreśla, że Babilończycy „nigdy nie odważyli się oderwać liczby od konkretnej sytuacji" i gwałtu na rozumieniu istoty liczby „dokonali dopiero Grecy przez to, że operowali samą liczbą, abstrahując od konkretu." „ (...)Niewątpliwie za błogosławione następstwo owego gwałtu uznać należy rozwój współczesnej komputeryzacji." „(...)Sprawia to wrażenie, jakby same liczby wyczarowywały rzeczywistość." Wypada nam sprecyzować, że liczbę oderwali od rzeczy przede wszystkim Grecy przed Arystotelesem. Koncepcje – platońska i pitagorejska – autonomizujące (także bytowo) liczbę, zostały skrytykowane na gruncie realizmu ontologicznego Arystotelesa, który – jak wspomnieliśmy w tekście głównym – rozprawił się *Metafizyce* z owym oderwaniem liczby od rzeczy. Współczesna nam epoka poznania kwantytywnego ma już inne niż greckie źródła. Jest np. konsekwencją postulatu analityczności Kartezjusza i mechanicyzmu, a z drugiej strony – w praktyce życia społecznego i gospodarczego – konsekwencją nowej moralności handlowej Adama Smitha. Objawem rozumienia liczby natomiast przypomina greckie przedarystotelesowskie stawianie liczby przed jednostkowym bytem realnym. (W związku z tym warto jeszcze przypomnieć, że sam J. Trąbka – ale czy na przekór naszej diagnozie? – czuje duże powinowctwo z Kartezjuszem, por.: J. Trąbka, *Dusza mózgu*, Wyd. WAM, Kraków 2000, s. 29-32).

[9] Por. T. Grabińska, Od nauki do metafizyki, Wyd. Nauk. PWN, Warszawa-Wrocław, 1998, rozdz. III.

[10] Arystoteles, *Metafizyka*, op. cit., Księga K.

[11] Por. I. Kant, *Allgemeine Naturgeschichte und Theorie des Himmels oder Versuch von der Vefassung und den mechnischen Ursprunge des ganzen Weltgebaeudes nach Newtonischen Grundsaetzen abgehandelt*, ed. H. Ebert, Verlag von Wilhelm Engelmann, 1890. Inna, niekantowska kosmogeneza, uwzględniająca ową ziarnistość w postaci elementarnych embrionów, została przedstawiona w: M. Zabierowski, „*O pewnym programie badawczym w kosmologii i kosmogonii. O niekantowskim schemacie rozwoju materii*", Z Zagadnień Filozofii Przyrodoznawstwa i Filozofii Przyrody VII (1985) 69-155.

łość, bezepizodyczność, wieczność. Trwałość daje dopiero ZW, bo abstrahuje zarówno liczbę, jak i fakt, obserwację, zliczenie. Liczba nie jest pierwiastkiem jak ogień, ziemia, powietrze, woda, O, C, N, H i inne ogniwa w łańcuchu przemian na poziomie trwałego. Nie chodzi o stabilizację przez liczby tego, co wyjątkowe, niepowtarzalne, ciągle się zjawiające – to ZW jest stabilizatorem.

Mimo uznawania – podobnie jak Arystoteles – liczby jako abstrakcji odnoszącej się do materialności twierdzimy, że jednak liczba jest wyróżniającą się abstrakcją nie dlatego, że charakteryzuje się trwałością i niezmiennością, ale dlatego, że: 1. legitymuje fakt, pomiar, zliczenie; 2. doskonali realne obiekty przez rozwój pomiarowej miry. Ciąg powiązań jest następujący: pomiar łamie jedność, zestawia jedność z epizodycznością i również implikuje ziarnistość, wielość, a ciągłość. Jest więc tym bardziej błędem oddzielanie liczby (trwałej) od ciał (zmiennych, zdarzeniowych), ale w porównaniu z wykładnią arystotelesowskiej metafizyki związek ten powinien zostać pogłębiony[12]. W liczbowym (matematycznym) opisie spotyka się świat wieczny i świat zmienny. W kosmologii liczba łamie trwałość i nieskończoność substratu (materialnego kosmosu) i prowadzi do ziarna oraz wskazuje na to, że błędna jest aziarnista kosmogonia kantowska czy spencerowska, czy koncepcja ciągłego substratu ogólnej teorii względności. Liczba nie jest zasadą świata, ale jest czymś więcej niż tylko abstrakcją materialności arystotelesowkiej rzeczy indywidualnej; liczba tę indywidualność rzeczy określa i w tym sensie generuje. Nie jest ani samodzielna, ani nie jest ideą, ale ma swoją niezastępowalną rolę w ustaleniu tożsamości izolowanego, odgraniczonego indywiduum. Nie jest to zatem rozumienie liczby ani czysto arystotelesowskie, ani tym bardziej pitagorejskie lub platońskie.

Wróćmy do konceptualizacji substratu we współczesnej kosmologii ogólnoteoriowzględnościowej. Jest ciągły nie z powodu danych obserwacyjnych, lecz z powodu dopasowania rzeczywistości zjawiskowej do rzeczywistości przedmiotów matematycznych geometrii różniczkowej, które muszą być ciągłe. Czy można mówić tu o idealizacji[13] rozkładu materii we wszechświecie? Z punktu widzenia budowania modelu ogólno teorio względnościowego wszechświata (przedmiotu globalnego) można tak twierdzić, ale sprawa w tym przypadku jest bardziej subtelna. Czy z punktu widzenia jakichkolwiek lokalnych właściwości obiektów we Wszechświecie jest to ich idealizacja, czy raczej jest to ich deindywidualizacja, unicestwienie? Tak naprawdę w problemach kosmologicznych spotyka się to, co globalne z tym, co lokalne[14]; czyni to rozważania o liczbie i substancji bogatszymi niż u Arystotelesa. Gdybyśmy pierwszą wymienioną arystotelesowską funkcję liczby polegającą na dzieleniu, wyróżnianiu części z całości chcieli zastosować do wydzielenia ziarnistych indywidualnych rzeczy (przedmiotów lokalnych) z ciągłej jednolitej całości wszechświata (przedmiotu globalnego), to na poziomie modelu kosmologicznego musielibyśmy się uciec do matematycznego zabiegu wymodelowania fluktuacji w ciągłym substracie, co by o tyle się jeszcze mieściło we wspomnianej pierwszej funkcji liczby, gdyby matematyczne byty arytmetyczne, czyli liczby jakoś uogólnić na skomplikowane przedmioty matematyczne reprezentujące (modelujące) fluktuacje, a to wcale nie jest kwestią rozwiązaną. Gdyby nawet tę kwestię pozytywnie rozwiązać, to i tak pozostaje problem ontologiczny, dotyczący pierwszeństwa owego przedmiotu globalnego, owej całości w stosunku do realnego, bytowości poszczególnych rzeczy.

Z punktu widzenia arystotelesowskiej metafizyki istnieją poszczególne rzeczy, ziarna większe i mniejsze we Wszechświecie. Gdyby więc przedmiot globalny, czyli Wszechświat był modelowany jako ciągły rozmazany substrat, to nie mógłby być w tej konceptualizacji traktowany jako całość jednocząca w sobie poszczególne ziarniste przedmioty lokalne (rzeczy o statusie bytowości), nawet gdyby dopuścić taki rodzaj jedności, w której poszczególne wyodrębnione części zjawiają się ewolucyjnie, bo ta możliwość (potencjalność) wyodrębnienia musiałaby być w nim wiecznie i trwale zakodowana, a tego w klasycznych modelach teoriowzględnościowych nie ma. Gdyby ów kod rozwoju struktury ziarnistej (embrionalność) udało się wprowadzić do matematycznej ogólnoteoriowzględnościowej aparatury pojęciowej, to z kolei można byłoby mniemać o prymacie przedmiotu matematycznego nad zjawiskowym. Ale na pewno rozsądnie trzeba by stwierdzić, że uogólniona liczba wyrażałaby istotowość rzeczywistości zjawiskowej, nie tylko jej materialność.

Ta podwójna funkcja liczby (przedmiotu matematycznego), w porównaniu ze statusem nadanym jej przez Arystotelesa, podwójna ze względu na istotę i materialność bytu, jaka wyłania się z analizy ontologicznej przedmiotu współczesnej kosmologii, wymaga rewizji arystolesowskiej koncepcji relacji liczby i rzeczy. W tej sytuacji związek liczby i rzeczy byłby silniejszy. I choć (uogólniona) liczba – inaczej niż u Arystotelesa – miałaby swój wymiar istototowy, to – inaczej niż u przedarystotelików – nie byłaby od rzeczy oderwana, ale tym bardziej byłaby związana z każdą jakością (jako liczba uogólniona) i z konkretną rzeczą (jako przedmiot arytmetyczny).

Ta skomplikowana analiza relacji przedmiotu globalnego do przedmiotu lokalnego wskazuje na to, że matematyzacja współczesnego przyrodoznawstwa wcale nie musi skłaniać do reaktywacji ontologii idealistycznych. Umiarkowana, nakierowana na zjawiskowość realistyczna ontologia Arystotelesa powinna jednak zostać rozwinięta w zakresie relacji między rozmaitego typu przedmiotami. Platońska koncepcja podziału na trzy grupy przedmiotów zróżnicowanych pod względem stopnia abstrakcji, która przeniknęła do arystotelesowskiej epistemologicznej koncepcji trójpodziału wiedzy, wydaje się upominać o uwzględnienie jej także na poziomie ontologicznym. Problem polega na tym, że w ontologii arystolesowskiej porządek platoński tych przedmiotów jest wykluczony. Należałoby więc stworzyć taki porządek przedmiotów, w którym abstrakcje mają swoje miejsce, ale nie są bytowo fundamentalne. Powyższe rozważania o liczbie i rzeczy odkrywają taki kierunek.

[12] Ten problem w innej perspektywie badawczej wystąpił np. w: T. Grabińska, *Kanony estetyczne modelowania przedmiotu zjawiskowego*, w: *Homo experimentator*, red. D. Sobczyńska i P. Zeidler, Wyd. Nauk. IF UAM, Poznań 2003, 237-254.; *idem*, *Intelektualne zabiegi modelowania a piękno obiektu zjawiskowego*, Estetyka i krytyka (2005), w druku

[13] Por. T. Grabińska, *Człowiek we Wszechświecie*, Wiedza i Życie 10 (1986) 11-15; *idem*, „Teoria, model, rzeczywistość", Ofic. Wyd. Polit. Wroc., Wrocław 1993.

[14] Więcej o fundamentalnym problemie lokalność *versus* globalność znajduje się w: Zabierowski, *O pewnym....*, op. cit.

INDEX

A

Arodź Tomasz 107, 287
Augustyniak Piotr 307, 349

B

Bieńkowski Paweł 281
Bilo Grzegorz 365
Bochenek Grażyna 355
Bogdał Piotr 233
Borkowski Paweł 143
Boryczko Krzysztof 287

C

Całka Michał 139
Crone Nathan 339

D

Darowski Marek 65
Dębowska Małgorzata 205
Dindorf Ryszard 147
Długosz Tomasz 281
Dobrowolska Anna 111
Drath Maria 129
Drozd Anna 113, 133
Duplaga Mariusz 191, 355
Durka Piotr 339
Dyduch Grzegorz 333
Dzwinel Witold 287

F

Filipowicz Bogusław 3
Filipowicz Elżbieta 3
Fogtman Anna 117
Franaszczuk Piotr 339

G

Gajewska Elżbieta 121
Gałach Magda 209
Gonciarz Marcin 291
Gorzelańczyk Edward Jacek 139, 217
Grabińska Teresa 381
Grabska-Chrząstowska Joanna 9
Groborz Anna 291
Grzanka Antoni 175
Grzesik Stanisław 361
Gólczewski Tomasz 65, 73, 81
Guzik Anna 247
Guzik Przemysław 139

H

Helbin Jadwiga 21
Herman Roman M. 99
Horzyk Adrian 13
Huflejt Michał 217

I

Izworski Andrzej 213, 297

J

Janik Paweł 51
Jasińska Anna 129
Juliszewski Tadeusz 291
Jung Aleksandra 87

K

Kaczmarek Piotr 157
Kaganek Krzysztof 221
Kalitzin Stiliyan 259
Kamiński Wiesław A. 27, 321
Kasiński Andrzej 157
Kasper Łukasz 355
Kawecka-Jaszcz Kalina 365
Klimkowicz Aleksandra 23
Kłys Małgorzata 3
Kniat Jan 139, 217
Kolarzyk Emilia 21
Kokot Teresa 227
Konarski Łukasz 129
Konieczka Grzegorz 139
Konieczny Leszek 113, 133
Kononowicz Andrzej 165, 365
Konopka Wiesław 175
Korpikiewicz Honorata 374
Kozowicz Andrzej 227
Król Marcin 113, 133
Królczyk Grzegorz 87, 99
Krzysiek Józef 165
Kwiatkowski Jacek 21
Kwiecień Joanna 3
Kurdziel Marcin 287
Kuśka Jacek 125

L

Laskowska Ewa 217
Lasoń Wojciech 57, 233
Legutko Janusz 361
Leluk Jacek 117, 121, 125
Lesyng Bogdan 117, 125
Liebert Adam 277
Libuszowski Wiesław 9
Lopes da Silva Fernando 259
Luboch Radosław 139

M

Maciejewski Henryk 129
Maniewski Roman 277
Matuszak Zenon 237
Mazurkiewicz Joanna 339
Mazurkiewicz Przemysław 171
Micek Stanisław 301
Micek Grzegorz 301
Michalska Marta 175
Michałek Marcin 213, 297
Mikrut Zbigniew 241, 307
Migacz Andrzej 213
Muc-Wierzgoń Małgorzata 227
Muszyński Siemowit 27
Murakowski Janusz 217

N

Niżankowska-Mogilnicka Ewa 355
Nowak Leszek 31
Nowakowska-Zajdel Ewa 227

O

Obtułowicz Krystyna 57
Ogiela Marek R. 311

P

Piela Lucjan 1
Pietruch Rafał 175
Piekarska Barbara 113, 133
Pietrzycka Agata 21
Piwowar Piotr 183
Podsiadły-Marczykowska Teresa 247
Pokora Maciej 89
Porębski Grzegorz 57
Pyrczak Wiesław 233, 365

R

Reroń Elżbieta 103
Roterman-Konieczna Irena 113, 133, 165, 365
Rogoż Monika 23
Rudzińska Monika 297
Rybarska Janina 113, 133

S

Sarapata Krzysztof 365
Sawczyk Karol 139
Sawow Aneta 237
Sobocki Jacek 99
Socha Mirosław 191
Sokalski Krzysztof 253
Spólnik Paweł 113, 133
Stopa Barbara 113, 133
Stanuch Helena 221
Suffczyński Piotr 259
Szczeklik Andrzej 355
Szura Mirosław 361
Szyszko-Bohusz Andrzej 373

Ś

Święcicki Mariusz 31, 37, 51

T

Tchórzewski Jerzy Rudolf 61
Thor Piotr 87, 99
Tomalak Waldemar 9
Trąbka Jan 103, 267, 369
Trąbka-Zawicki Paweł 103, 369
Trzupek Mirosław 311
Turek Michał 45
Twardowska Magdalenia 355

W

Wais Piotr 31, 37, 51
Wajs Wiesław 37, 51
Walecki Piotr 57, 233, 267
Waniewski Jacek 205
Wasilewska-Radwańska Marta 237
Weryński Andrzej 209
Wierzgoń Janusz 227
Włodarski Jerzy 197, 317
Wojas-Pelc Anna 333
Wojtowicz Hubert 51
Woźniak Piotr 217
Woźniakiewicz Ewa 361
Wójcik Grzegorz M. 321
Wójcik Mariusz 273

Z

Zabierowski Mirosław 381
Zajączkowski Michał 241
Zemanek Grzegorz 113, 133
Zyss Tomasz 325

Ż

Żabińska-Płazak Ewa 333
Żabińska-Popiela Marta 165
Żołek Norbert S. 277
Żygierewicz Jarosław 339