



## EDITORIAL BOARD

### EDITOR-IN-CHIEF

**Professor Irena Roterman-Konieczna**

Jagiellonian University, Medical College, Kraków, Poland

### HONORARY ADVISORS

**Professor Ryszard Tadeusiewicz**

AGH – University of Science and Technology, Kraków, Poland

**Professor Jan Trąbka**

Jagiellonian University, Medical College, Kraków, Poland

### EDITORIAL BOARD (in alphabetical order)

**Professor Piotr Augustyniak**

AGH – University of Science and Technology, Kraków, Poland

**Dr Marian Bubak**

AGH – University of Science and Technology, Kraków, Poland

Amsterdam University, Amsterdam, Netherlands

**Dr Jacques Chomilier**

Universite Pierre et Marie Curie, Paris, France

**Dr Dragan Gamberger**

Rudjer Boskovic Institute, Zagreb, Croatia

**Dr Inga Hege**

University of München, München, Deutschland

**Dr Sören Huwendiek**

University Children's Hospital Heidelberg, Heidelberg, Deutschland

**Professor Kalina Kawecka-Jaszcz**

Jagiellonian University – Medical College, Kraków, Poland

**Professor Leszek Konieczny**

Jagiellonian University, Medical College, Kraków, Poland

**Dr Valeria Krzhizhanovskaya**

University of Amsterdam, Amsterdam, Netherlands

**Professor Fabio Polticelli**

University RomaTre, Roma, Italy

**Dr Paweł Spólnik**

Jagiellonian University, Medical College, Kraków, Poland

**Professor Katarzyna Stąpor**

Silesian University of Technology, Katowice, Poland

**Professor Nabil Zary**

Karolinska Institutet, Stockholm, Sweden

**Polish Ministry of Science and Higher Education journal rating: 6.000**

Sustaining institution: Ministry of Science and Higher Education

Edition: 300 copies

Copyright by individual authors and Jagiellonian University, Medical College

ISSN 1895-9091 (hard copy)

ISSN 1896-530X (electronic version)

<http://www.bams.cm-uj.krakow.pl>

The publisher: Księgarnia Akademicka, Kraków, St. Anny 6, Poland

<http://www.akademicka.pl>

# Contents

- 5 Role of Image Processing in the Cancer Diagnosis  
Jerzy Detyna, Łukasz Jeleń, Michał Jeleń

## LARGE SCALE COMPUTING

- 9 Snapshot backup system in distributed computing environments – problems, solution, results  
M. Magryś, M. Pogoda, G. Sułkowski, M. Twardy, P. Winiarczyk

## OPTIMIZATION PROCEDURE

- 13 Further improvement of the Glowworm Swarm Optimization algorithm by adding a conservation of agent move direction  
Piotr Oramus

## IMAGE PROCESSING

- 21 Grand Challenges Less Challenging: New Possibilities Provided by Graphics Processing Units  
Dawid Kuna, Marcin Makowski, Grzegorz Mazura, Paweł Russek, Marcin Janiszewski<sup>a</sup>, Kazimierz Wiatr<sup>a</sup>,

## MEDICAL SIGNAL INTERPRETATION

- 27 Application of standard algorithms of automatic signal separation in medicine  
Agnieszka Nowak  
39 Implementation and Simulation of the Scalar Timing Model  
Maciej Komosinski, Adam Kups  
51 A continuous approach to the ECG noiseprint estimation  
Piotr Augustyniak

## BIOINFORMATICS

brak tekstu 5

- 57 SSTMProt, a point mutation sensitive tool to combine results and to predict the consensus sequence, and secondary structure of transmembrane proteins  
Jan Paweł Jastrzębski  
69 Analysis of data generated by insulin pump using LabVIEW environment  
Bartłomiej Matejko  
73 The motion capture system to conduct an electrical muscle stimulation experiment on upper limb movement  
Andrzej Skalski, Mirosław Socha, Łukasz Malicki, Marek Iwaniec

## MEDICAL SCIENCE

- 79 Python based simulator of ICU patients hyperglycemia  
Jerzy Baranowski, Waldemar Bauer, and Paweł Płuciennik  
85 An application software for anatomical structures segmentation from tomographic data of abdominal cavity  
Grzegorz Toporek, Mirosław Socha, Andrzej Skalski

## NEURAL NETWORK

- 91 Neural Networks for Medical Image Processing  
Tomasz Pieciak, Joanna Jaworek, Marek Gorgon



## ROLE OF IMAGE PROCESSING IN THE CANCER DIAGNOSIS

JERZY DETYNA<sup>1</sup>, ŁUKASZ JELEŃ<sup>2,3</sup>, MICHAŁ JELEŃ<sup>4</sup>

<sup>1</sup>*Institute of Materials Science and Applied Mechanics, Wrocław University of Technology  
Smoluchowskiego 25, 50-370 Wrocław, Poland,* <sup>2</sup>*Wrocław School of Applied Informatics  
Wejherowska 28, 54-239 Wrocław, Poland,* <sup>3</sup>*Institute of Agricultural Engineering  
Wrocław University of Environmental and Life Sciences, Chelmońskiego 37-41, 51-630 Wrocław, Poland*  
<sup>4</sup>*Department of Pathology and Clinical Cytology, Medical University of Wrocław,  
Borowska 213, 50-556 Wrocław, Poland*

**Abstract:** Cancer is still one of the most deadly diseases. It is a well known fact that the early diagnosis is crucial and allows for the successful treatment while cancers diagnosed in their late stage are almost impossible to treat. For precise and objective diagnosis there is a need for a computerized method for cytological image processing, which is an integral part of a diagnosis process. In this work we present a classification system for grading cancer malignancy. In particular, issues of image processing in the aspect of medical diagnosis presented by prof. R. Tadeusiewicz and Dr. J. Śmietański in [1].

**Keywords:** cancer, image processing, malignancy classification, cancer classification

### INTRODUCTION

Information science (bioinformatics, bio-cybernetics) is a discipline that has found its important place in the modern medicine. One of the main reasons for that is huge amount of data used by physicians in the diagnostic and clinical processes. That considerable volume of information has to be properly collected and processed, which is very difficult, or just impossible, without the application of modern computer systems. Most of the contemporary medical equipment is able to transmit results of their tasks directly into a computer (the carriers), which essentially facilitates collection of the necessary information.

In view of the above, with great joy and gratefulness for the contribution, I accept the publication of prof. Ryszard Tadeusiewicz and dr Jacek Śmietański devoted to the issue of applying computer methods in medicine. In the modern way the book captures the problems of acquiring medical images and their processing. The information concerning medical apparatus presented in the individual chapters is given in the reliable and up to date way (and even in prospect), which, considering the fast technological progress, indicates huge experience and commitment of the Authors in editing the book (Fig. 1).

Due to the wide expression of the medical informatics topics the discussed publication may find wide circle of readers both, among the experienced medical specialists (physicians, radiologists) and the medicine or similar discipline students, e.g. of biomedical engineering. In case of that second group of

recipients the control questions placed at the end of each chapter would prove to be extremely useful.

The first chapter of the book introduces a reader into arcana of the medical imaging issues, with particular emphasis on the aspects of digital acquisition and processing of images. The second chapter is fully devoted to medical imaging techniques. In the easily adaptable way the Authors bring closer the commonly used imaging techniques such as: X-ray radiography, computer tomography, magnetic imaging, radioisotope methods and PET. Besides the description of modern equipment used in the diagnostic imaging, a reference is also made to the history and physical base of a research. The third and fourth chapters shortly present methods of computer image processing and their automatic interpretation. In a very accessible way the methodology of computer aided medical diagnostic has been presented, the exemplification of which is found in chapter five – Exemplary application of automatic medical images interpretation in prostate cancer diagnostics. Besides presentation of the research methods currently used in diagnosing a prostate cancer the Authors introduce the study of a selected case, in which the reader finds himself as being a participant of the diagnostic procedure. Such embrace of the topic enables better understanding of the computer information processing mechanisms and points to the general significance of informatics and exact science methods in the modern medicine (example in breast cancer diagnosis - see section 2 and 3) [1, 2].

The book has been written in the literary-and-technical style, to which prof. Ryszard Tadeusiewicz has already accustomed us through his numerous publications. The enormous publication experience manifests itself in presenting the engineering issues in a perfect-language way.

The presented physics or mathematical methods issues do not require from a reader the extensive and exact preparation. Frequent drawings, diagrams and photographs are very helpful in understanding of the provided knowledge. Bibliography presented at the end of the book and sources specified at graphical objects will surely enable a reader (and especially those studying) to extend the presented subject matter.

## BREAST CANCER DIAGNOSIS

Breast cancer is the most often diagnosed cancer among middle-aged women. According to the World Health Organization there are 7.6 million deaths worldwide due to the cancer each year out of which 502,000 are caused by breast cancer alone. With such a high rate, breast cancer also is one of the most deadly cancers. Successful treatment is a key to reduce the high death rate. Most of the diagnosed cases can be fully recovered when diagnosed at an early stage. Cancers in their early stages are vulnerable to treatment while cancers in their most advanced stages are usually almost impossible to treat [3].

During the diagnosis process, the cancer is assigned a malignancy grade that is used to determine the appropriate treatment. Malignancy grading allows doctors to precisely estimate cancer behavior with or without undertaking treatment and therefore is called a prognostic factor. It plays an important role in breast cancer diagnosis and the appropriate treatment is chosen ac-

ordingly to this factor. This is a complicated process that involves assessing numerous nuclear features that allow for malignancy estimation. Cancer malignancy is graded based on a numeric scale that was introduced by Bloom and Richardson in 1957 [4]. The grading scheme proposed by the authors was derived to assess malignancy from histological slides and is now widely used among pathologists to grade not only histological but also cytological tissue.

The grading system originally proposed by Bloom and Richardson [4], later modified by Scarff [21] and known as modified Scarff-Bloom-Richardson system, for grading breast cancer malignancy is one of the best known prognostic factors for this type of cancer [5]. These systems are based on grading of cells' polymorphy, ability to reform histoformative structures, and mitotic index. All of these features are described by the Bloom-Richardson scheme as three factors that use a point based scale for assessing each feature according to the following description:

1. **Degree of structural differentiation (SD)** – In histopathological slides this is also described as tubule formation, which reflects cell tendency to form tubules. Since in cytological smears tubules are not preserved, the scoring given below for this factor is based on the classification of cell groupings within a smear, see Fig. 2a for example. On the right image in the Fig. 2a only one group is visible, which indicates lower malignancy than the case in the left image where dispersed cells are visible.
  - One point – cells in the image are grouped and spread regularly.
  - Two points – both grouped and single cells found within the image.
  - Three points – cells are spread irregularly.

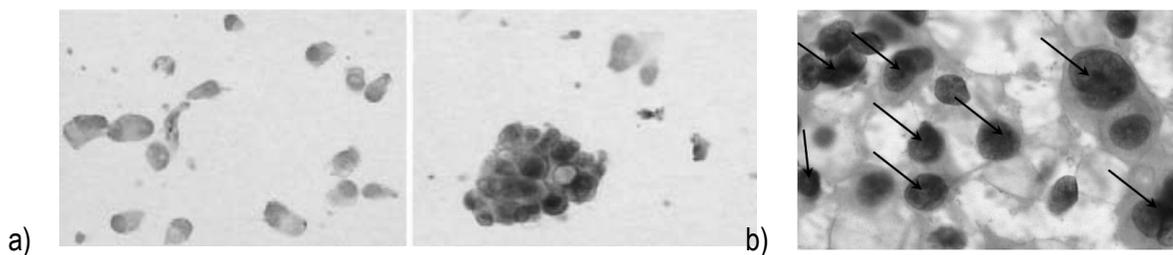


Fig. 2. FNA smears: a) Structural differentiation; b) Pleomorphism [3]

2. **Pleomorphism (P)** – This factor takes into consideration differences in size, shape and staining of the nuclei. This scoring is fairly straightforward because with the growth of irregularity of the nuclei the prognosis becomes worse. Fig. 2b shows an example of these variations. Arrows in the image indicate cells with visible variations in shape and color. Arrows seem to be very small in Fig. 2b.
  - One point – nuclei with uniform size, shape and staining.
  - Two points – moderate variations are found.
  - Three points – very significant variations (see Fig. 2b).
3. **Frequency of hyperchromatic and mitotic figures (HMF)** – Mitosis is a process in the cell life cycle in which

a mother cell divides into two identical cells. Main objective of this factor is to assess the number of mitosis in the field of view. Several fields of view on the same slide are taken into account because this step is done in a large magnification. The more cases of mitosis found, the worse the prognosis is. During the staining process, mitotic cells stain the most intensively providing the darkest areas in the nucleus.

- One point – occasional figures per field are found.
- Two points – smears with two or three figures in most fields.
- Three points – more than three figures per fields are found.

According to the BR scheme, the malignancy of the tumor is assigned a grade that depends on the quantitative values of the above factors and is determined by the following equation:

$$G = SD + P + HMF. \tag{1}$$

The final grade is obtained by the summation of all the awarded points for each factor described earlier. Depending on the value of G, the tumor is assigned one of three grades according to the chart shown in Fig. 3.

Points									
3	4	5	6	7	8	9			
Grade I			Grade II		Grade III				
Low			Intermediate		High				
malignancy			malignancy		malignancy				

**Fig. 3.** Grade determination for the Bloom – Richardson scheme, taken from [5]

Based on the evaluation of the malignancy of the tumor an appropriate treatment is suggested.

Assigning a malignancy to a case is a very difficult task and is dependent on the experience of the pathologist. More experienced pathologists that have seen more cases are more reliable in their diagnosis. On the other hand, due to overwork and fatigue, seeing more similar cases may lead to misclassification of the malignancy. To address this problem we present an automated grading approach that is able to evaluate and assign a grade to Fine Needle aspiration biopsy (FNA) tissue. To achieve this we convert the Bloom – Richardson [4] grading scheme into a classification problem.

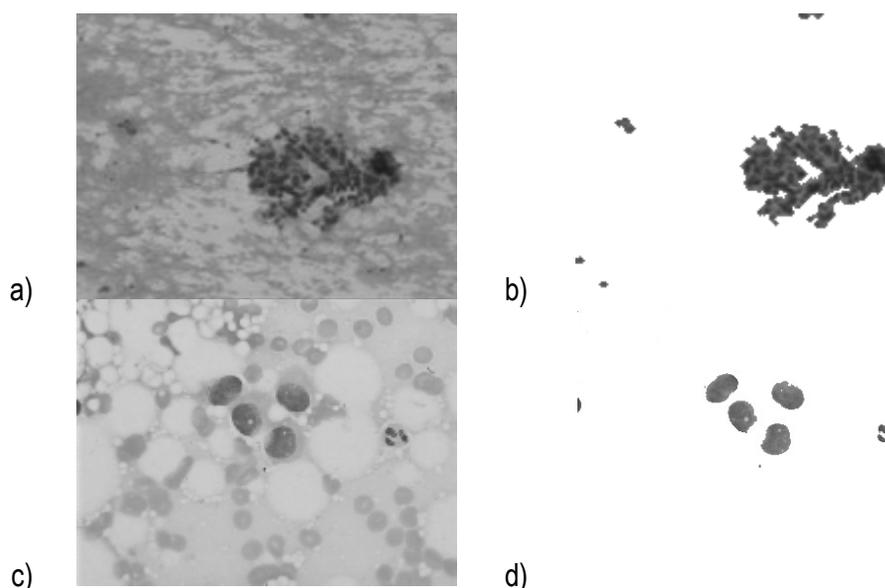
In the literature one can find approaches for breast cancer classification [6-12]. Most of these approaches involve classification of breast cancer as benign or malignant. The proposed system implements the idea of multilayer perceptron to achieve the classification of malignancy into intermediate and high malignancy classes. The results obtained with the perceptron are then compared with those obtained by a pathologist as described in section 3.

### FEATURE EXTRACTION AND CLASSIFICATION

Feature extraction is an important part of each classification task. Poor definition of features can lead to a high error rate of a classification system. Each classification system takes a feature vector as an input and responds with a category to which the object belongs. A feature vector is a set of features extracted from the input data. Before we can extract features used for classification our input data needs to be preprocessed and segmented. Preprocessing is a task of removing not important information from the data. Segmentation is an operation during which we isolate the boundaries of the important parts of the data that are then used for feature extraction and classification. In this study we make use of automatic thresholding as proposed by Riddler and Calvard [13] and fuzzy C-means (FCM) segmentation [14-16] to retrieve the nuclear information from fine needle aspiration biopsy (FNA) slides.

During the cytological examination two kinds of images are taken. The first type are the images that are recorded at 100x magnification. The second set of images is recorded at 400x magnification. All of the images used in this study had a resolution of 96 dots per inch (dpi) and a size of 764x572 pixels.

Due to the fact that different structures are visible at different magnification, the choice of the segmentation technique depends



**Fig. 4.** Segmentation results: a) FNA slide at low magnification; b) Segmented image; c) FNA slide at high magnification; d) Segmented image [3]

on the magnification of the image. For images recorded with low magnification the automated thresholding method shall be used (see Fig. 4a), while for the high magnification images the FCM segmentation will be applied (see Fig. 4b).

Based on the segmentation we are able to extract a set of 16 features that allow for training and testing the perceptron. For breast cancer malignancy classification two kinds of images are used. One subset consist of images recorded in low magnification, which allows for extracting structural features, the second subset of images is built from high magnification images which are used for polymorphic feature extraction [17-20].

## CONCLUSIONS

The article presents the role of image processing in relation to cancer diagnosis. Modern methods of image acquisition and image processing in medicine are presented in the book [1] by prof. Tadeusiewicz R. and J. Smietański. In the book, the authors have shown many examples of the application of imaging techniques in specific clinical cases.

The article quotes the examples of cancer image processing to show the importance of the problem. Mathematical methods can help to identify cancer cells. More advanced methods allow for the assessment of the tumor grade. Such a procedure helps in making a good diagnosis, which in turn determines the effectiveness of recovery. Automatic pattern recognition algorithms are successfully implemented to modern diagnostic equipment in medicine.

Over all we can conclude that from the pathologist point of view, having a system that is able to assist them during the decision making process, is very useful and helpful especially in the situations where it is difficult to decide to which malignancy class the case should belong. Such a system would help to make their diagnosis more objective and precise.

## REFERENCES

1. Tadeusiewicz R., Śmietański J.: *Pozyskiwanie obrazów medycznych oraz ich przetwarzanie, analiza, automatyczne rozpoznawanie i diagnostyczna interpretacja*. Kraków: Wydawnictwo STN, 2011.
2. Tadeusiewicz R.: *Informatyka medyczna*. Lublin: Wydawnictwo UMCS, 2011.
3. Jeleń Ł., Lipiński A., Detyna J., Jeleń M.: Grading breast cancer malignancy with neural networks. *Bio-Algorithms and Med-Systems* 2011, 7, 2: 47-53.
4. Bloom H.J.G., Richardson W.W.: Histological grading and prognosis in breast cancer. *British Journal of Cancer* 1957, 11: 359-377.
5. Le Doussal V., Tubiana-Hulin M., Friedman S., et al.: Prognostic value of histologic grade nuclear components of Scarff-Bloom-Richardson (sbr). An improved score modification based on a multivariate analysis of 1262 invasive ductal breast carcinomas. *Cancer* 1989, 64(9): 1914-1921.
6. Street N.W.: Xcyt: a system for remote cytological diagnosis and prognosis of breast cancer. In: Jain L.C. (ed.), *Soft Computing Techniques in Breast Cancer Prognosis and Diagnosis*, Singapore: World Scientific Publishing, 2000, pp. 297-322.
7. Street W.N., Wolberg W.H., Mangasarian O.L.: Nuclear feature extraction for breast tumor diagnosis. In: *Imaging Science and Technology/Society of Photographic Instrumentation Engineers 1993 International Symposium on Electronic Imaging: Science and Technology*, IS&T/SPIE, San Jose, California, 1993, Vol. 1905: 861-870.
8. Nezafat R., Tabesh A., Akhavan S., Lucas C., Zia M.A.: Feature selection and classification for diagnosing breast cancer. *Proceedings of International Association of Science and Technology for Development International Conference, IASTED*, Cancun, Mexico, 1998, pp. 310-313.
9. Walker H.L. Jr., Albertelli L.E.: Breast cancer screening using evolved neural networks. *IEEE International Conference on Systems, Man, and Cybernetics*, San Diego, USA, 1998, 2: 1619-1624.
10. Cheng H.D., Li X.Q., Riodan D., Scrimger J.N.: A Parallel Approach to Tubule Grading in Breast Cancer Lesions and its VLSI Implementation. *Computer-Based Medical Systems: Fourth Annual IEEE Symposium*, 1991, pp. 322-329.
11. Schnorrenberg F., Tsapatsoulis N., Pattichis C.S., Schizas C.N., Kollias S., Vassiliou M., Adamou A., Kyriacou K.: A modular neural network system for the analysis of nuclei in histopathological sections. *IEEE Engineering in Medicine and Biology Magazine* 2000, 19: 48-63.
12. Jeleń Ł., Fevens T., Krzyżak A.: Classification of Breast Cancer Malignancy using Cytological Images of Fine Needle Aspiration Biopsies. *Int. J. Appl. Math. Comput. Sci.* 2008, 18, 1: 75-83.
13. Ridler T., Calvard S.: Picture thresholding using an iterative selection. *IEEE Trans. System, Man and Cybernetics* 1978, 8: 630-632.
14. Klir G.J., Yuan B.: *Fuzzy sets and fuzzy logic: Theory and applications*. New Jersey: Prentice Hall, 1995.
15. Theera-Umpon N.: Patch-Based white blood cell nucleus segmentation using fuzzy clustering. *ECTI Transactions on electrical eng., electronics and communications* 2005, 3(1): 15-19.
16. Bezdek J.C.: *Pattern recognition with fuzzy objective function algorithms*. New York: Plenum Press, 1981.
17. Jeleń Ł.: *Computerized cancer malignancy grading of fine needle aspirates*. PhD thesis, Concordia University, 2009.
18. Rangayyan R.M.: *Biomedical Image Analysis (Biomedical Engineering)*. Boca Raton, FL: CRC Press, 2004.
19. Lapedes A., Farber R.: *Nonlinear signal processing using neural networks: prediction, and system modelling*. Technical Report, LA-UR-87-2662, 1987.
20. Wodzislowski W., Detyna J., Jeleń Ł., Kaczyński R.: Ocena wybranych metod klasyfikacji pacjentów (naiwna metoda Bayesa, metoda wektorów nośnych) w aspekcie reakcji kości miednicznej na wszczepienie panewki stawu biodrowego. *Sci. Bull. of Chem. Sect. of Math. and Comp. Science* 2010, 1: 179-212.
21. Scarff R.W., Torloni H.: *Histological typing of breast tumors. International histological classification of tumours*. *World Health Organization. Geneva* 1968, 2, 2: 13-20.

## SNAPSHOT BACKUP SYSTEM IN DISTRIBUTED COMPUTING ENVIRONMENTS – PROBLEMS, SOLUTION, RESULTS

M. MAGRYŚ<sup>1</sup>, M. POGODA<sup>1</sup>, G. SUŁKOWSKI<sup>1</sup>, M. TWARDY<sup>1</sup>, P. WINIARCZYK<sup>1</sup>

*AGH University of Science and Technology Academic Computer Center  
CYFRONET AGH, ul. Nawojki 11, 30-950 Krakow, Poland;  
email: [m.magrys,m.pogoda,g.sulkowski,m.twardy,p.winiarczyk]@cyfronet.pl;  
phone: (+48 12) 632 33 55, fax: (+48 12) 634 10 84*

**Abstract:** One of the fundamental problems in distributed computing environments is to ensure proper security of user data, which due to the nature of the system, are mostly stored in the dedicated distributed shared disk resources. These resources are available for computing machines (WN) via a special server – Storage Element (SE) with dedicated access protocols such as NFS or LUSTRE. Currently, the numbers of concurrent computing machines exceed a few thousand which, combined with shared disk resources, whose capacity often is counted in Peta Bytes (PB), prevents the efficient archiving (backing up) user data. This can cause the additional load on the SE, while maintaining efficiency in data access. The solution to user data in a distributed shared disk resources is to use the snapshot mechanism (snapshot) which is mostly available in a disk array. The authors present the results of using a dedicated backup system with snapshot mechanism on the disk array with the intermediate backup server and backup software HP DataProtector6.0 for protection of user data in PL-Grid environment.

**Keywords:**

### 1. Introduction

Dynamic development of grid computing infrastructure raises the challenge for dispersed mass storage systems. The most important issues include satisfying the constantly growing demand for capacity, high efficiency and accessibility of storage resources, as well as the very important and widely discussed problem of safety of stored data. Therefore, proper configuration of mass storage systems, together with providing necessary computing power, becomes one of the key parameters decisive for the efficiency and safety of the entire Computing Centre. Administrators of storage infrastructure constantly search for an optimum solution that satisfies the needs of the most demanding users, who often consider high security of stored data as the most important feature of mass storage.

Taking current large volumes of individual hard disks and methods of sharing their resources into consideration, ensuring data security is a huge challenge for administrators, who are responsible for backups. In most cases, key user information is

stored on external arrays equipped with specialized mechanisms for data protection, such as RAID groups or hot-spare disks. Distribution of data to users and so-called Worker Nodes (WN) is provided by special Storage Element (SE) servers, utilizing various protocols for remote sharing of files (i.e. NFS or LUSTRE). In such a dispersed and shared mass storage infrastructure, backups based on Disk-to-Tape (D2T) mechanisms heavily degrade the work efficiency of both disk systems and client workstations. Such a solution is unacceptable from the perspective of the aforementioned Computing Centre's priorities. This fact has become a starting point for the search of an alternative solution for backup of stored data.

The paper presents results of work focusing on a solution for creating backups of a dispersed disk system with implementation of the Disk-To-Disk-To-Tape (D2D2T) method connected to snapshot copies of a disk array and utilizing of a proxy backup server with HP DataProtector 6.0 software. The resulting solution has been tested in an actual computing grid cluster environment utilized in the Cyfronet Academic Computing Centre.

## 2. Infrastructure Characterization; Definition of the Problem

Figure 1 presents the dispersed disk system at AGH's CYFRONET ACC from the perspective of creating backup copies of data processed in grid computing clusters.

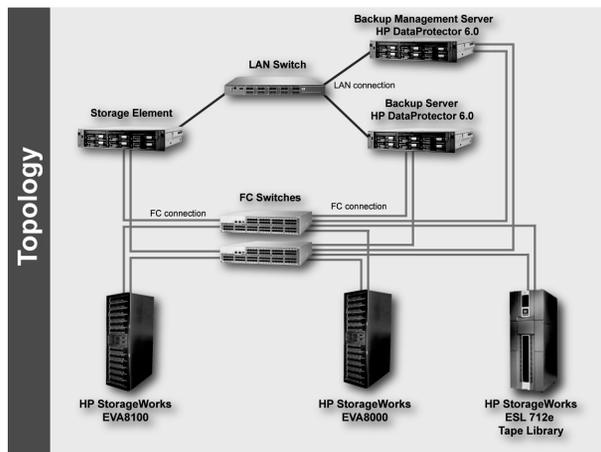


Fig. 1. Topology of the dispersed disk system at AGH's CYFRONET ACC

The system's principal components presented in figure 1 include: a Backup Server (BS), a client backup server (Storage Element – SE), a managing server for the backup copy system (Backup Management Server – CM), HP EVA8100 disk array (EVA), HP ESL 712e tape library, LAN (Local Area Network) and SAN (Storage Area Network), together with the necessary infrastructure of physical connections.

A typical backup scenario of a dispersed disk system assumes direct copying of data from the SE backup client to the BS backup server via LAN. The whole copying process can be divided into several intermediate stages. During the first stage, the SE client reads data from disk resources within SAN and sends them via LAN to the BS backup server. In the next, the Backup Server (BS) transfers received data to magnetic media located in the tape library that connects to the BS via SAN. During the last stage, an update of the backup session database (located on the managing server for the backup system) is performed. Figure 2 presents a diagram of this process.

Such a method for securing user resources requires the use of three independent servers: a SE client, a BS backup server and a CM database server. In this case, the stored information is being sent via two independent network infrastructures: LAN and SAN. The main disadvantage of this solution is the additional load for the SE client, resulting from the necessity of reading and transmitting data during the backup creation process. Thus, a significant reduction of the SE's performance occurs, necessary for performing current tasks for local clients of the computing node. Other negative features of the presented solution, apart from such performance degradation, are as follows:

- Degradation of bandwidth available to the SE (Storage Element) client.

- Lack of flexibility in defining a backup window, due to the way computing nodes are utilized (placement of a window depends on currently running computing tasks).
- Performance degradation of the SE's (Storage Element) disk resource access during the data backup procedure.
- Low efficiency of data copying processes.
- Conventional backup methods for dispersed disc resources also cause numerous problems on the backup system side, such as:
  - Low utilization of network bandwidth.
  - Low utilization of tapes (due to the necessity of providing a continuous data stream to the drives, especially in the case of new, more efficient LTO-4 and LTO-5 cartridges).
  - Shortening of the lifespan of a tape.
  - Inefficient copying of small files.

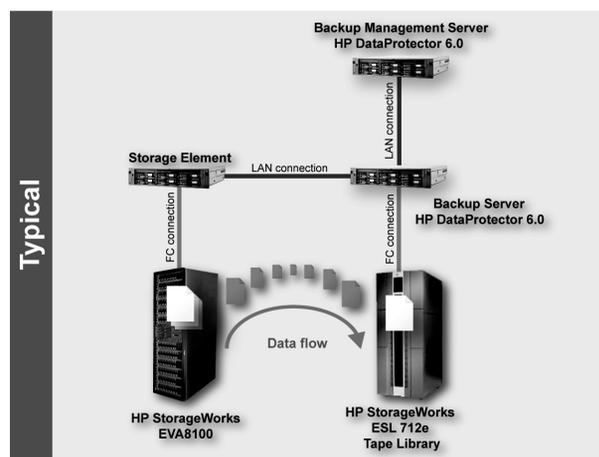


Fig. 2. Diagram of a typical backup scenario

## 3. Snapshot Copy Solution

Utilization of the Disk-To-Disk-To-Tape (D2D2T) data storage method with the snapshot mechanism for disk array copies is a way of eliminating the problems described in point 2. The proposed solution divides the main backup scenario into two stages: Disk-To-Disk (D2D) and Disk-To-Tape (D2T).

During the first stage (D2D), snapshot copies are made on the EVA array of source disk volumes utilized by the SE (Storage Element) client, which contains the data to be copied. Generated volume images are exact copies of the source disks at the moment they are generated. Snapshots are then made available via SAN to the backup server, which initiates a process of copying local data onto an intermediate disk medium without any participation from the client (Storage Element – SE). At the same time, the information about backed-up files is saved in the CM (Backup Management Server) server's database. During the second stage of this scenario (D2T), after the copying of all snapshots from the D2D stage is completed, a consolidation of intermediate disk media onto tape media occurs. Use of a single, continuous copying process causes a significant increase in the

efficiency of utilization and lifespan of magnetic tapes. The D2T process, similarly to D2D, is performed locally in the BS backup server without SE client participation. An overview of the proposed solution, with specification of D2D and D2T processes, is presented in figure 3.

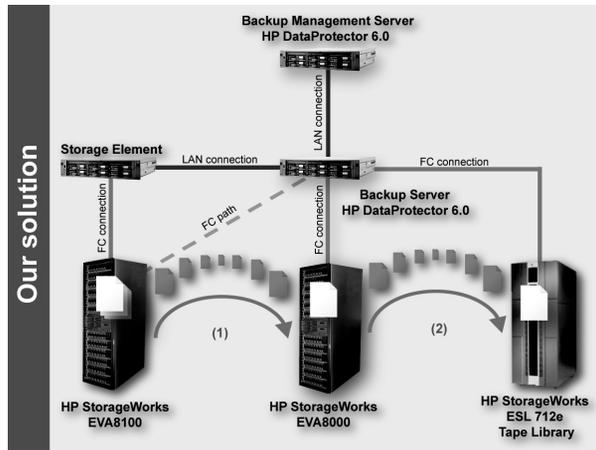


Fig. 3. Diagram of a proposed snapshot backup scenario

### 4. Results

Efficiency tests were performed in order to compare both previously discussed methods. Both scenarios utilize the actual grid computing cluster infrastructure, described in detail in point 1. HP DataProtector 6.0 was the software package used for automatic creation of backup copies. Figures 4 and 5 show the SE server load graphs and utilization of EVA array for a conventional data backup scenario, described in step 2.

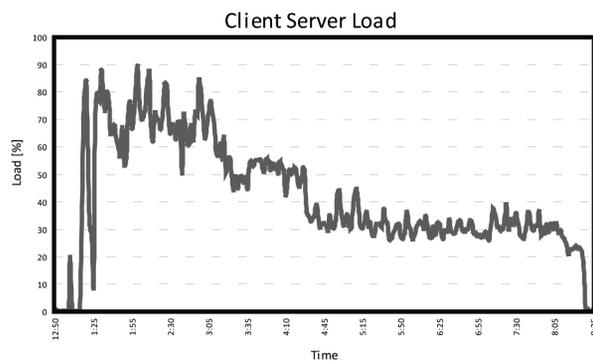


Fig. 4. Client server load

As the graphs show, the SE client's load (expressed as a percentage of the server's CPU occupation) is inadequate for the resulting read speed of source disks. At the same time, the transfer during reading of data from the SE client (Fig. 4) corresponds to data transfer during the process of saving media

on a BS backup server onto tape, which results in inefficient utilization of magnetic tapes.

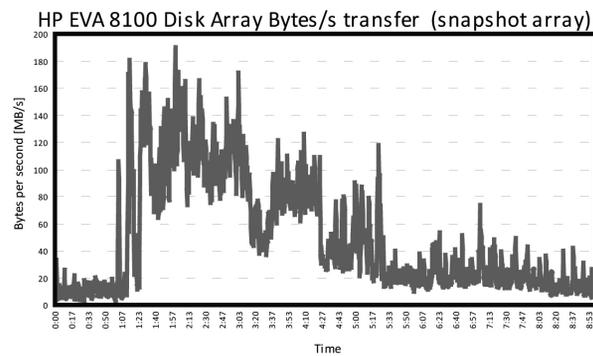


Fig. 5. HP EVA 8100 Disk Array Bytes/s transfer (snapshot array)

The resultant low transfer values of the copying process (Fig. 5) are not acceptable from the perspective of a backup copy system. Figures 6 and 7 include graphs of BS server load and EVA array utilization within the D2D2T scenario.

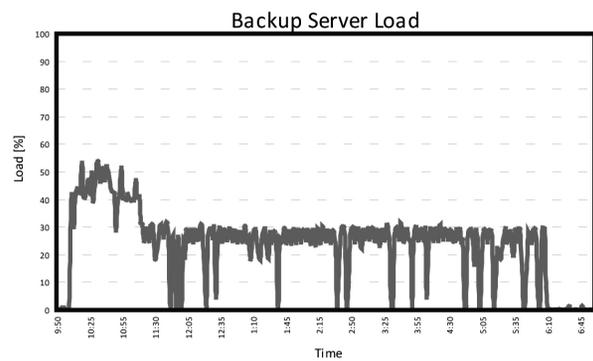


Fig. 6. Backup server load

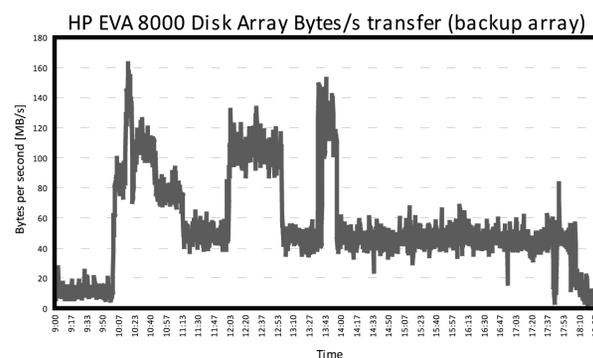


Fig. 7. HP EVA 8000 Disk Array Bytes/s transfer (backup array)

Implementation of the proposed solution resulted in constant, efficient utilization of disk resources during the D2D stage (Fig. 7), as well as of tape drives during the D2T stage (Fig. 8). At the same time, with the use of the two-stage copying process, the BS server's utilization dropped to an acceptable level (Fig. 6), while the read and write ratios for the backup media (Fig. 7 and Fig. 8) remained satisfactory.

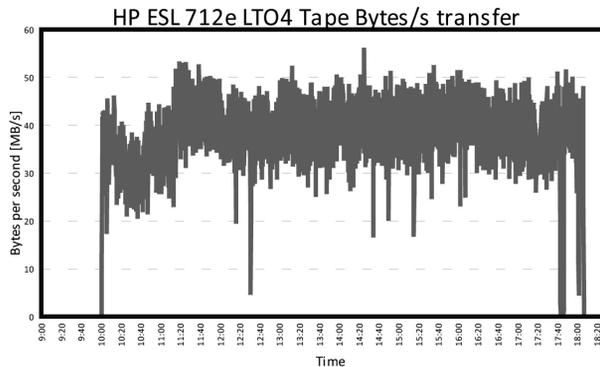


Fig. 8. HP ESL 712e LTO4 Tape Bytes/s transfer

## 5. Conclusions and Summary

The performed tests confirm that implementation of the solution presented in point 3 allows one to achieve satisfactory ratios of data transfer onto the backup media, as well as server load reduction. Also, with the proposed Disk-To-Disk-To-Tape data protection method, which utilizes a mechanism of creating snapshot copies of the arrays, it was possible to eliminate

problems present in a conventional backup scenario. Specifically, it was possible to:

- Significantly reduce SE server load due to transferring of the copying process onto the BS server (D2D stage).
- Increase the flexibility of backup windows' definition and minimize access performance degradation of the SE server's disk resources due to the use of snapshot copies of disk arrays.
- Increase bandwidth available for computing tasks by eliminating the use of LAN for copying purposes.
- Increase utilization and lifespan of magnetic tapes due to the use of consolidation during the D2T stage.
- Remove the problem of small file copying due to transferring of backup processes onto the BS server and the use of consolidation.

## References

1. Pogoda M., Sułkowski G., Twardy M.: Preparing Storage Infrastructure to Meet the Requirements of the Grid – Environment. Proceedings of the CGW'06, Krakow, 2006.
2. Dwużnik M., Pogoda M., Sułkowski G., Twardy M.: Performance optimization of storage resources accessibility in grid system environment after migration to Scientific Linux 4. Proceedings of the CGW'07, Krakow, 2007.
3. Flis Ł., Lasoń P., Pogoda, M., Sułkowski G., Twardy M.: Comparing two Lustre implementation scenarios – based on storage servers and enterprise SAN disk arrays. Proceedings of the CGW'08, Krakow, 2008.
4. Pratt S., Heger D.: Workload Dependent Performance Evaluation of the Linux 2.6 I/O Schedulers. Proceedings of the Linux Symposium, Ottawa, 2004.
5. Grider G.: Scalable i/o, file systems, and storage networks: R&D at Los Alamos, 2005.

## FURTHER IMPROVEMENT OF THE GLOWWORM SWARM OPTIMIZATION ALGORITHM BY ADDING A CONSERVATION OF AGENT MOVE DIRECTION

PIOTR ORAMUS

*Department for Information Technology, Faculty of Physics, Astronomy and Applied Computer Science, Jagiellonian University in Krakow, Reymonta 4, 30-059 Krakow, Poland; piotr.oramus@uj.edu.pl*

**Abstract:** Glowworm Swarm Optimization is an algorithm, which can localize multiple optima of the multi-modal function during a single run. Unfortunately, original algorithm manifests several disadvantages: an agent can change its position only in the presence of other agents and exclusively in direction of randomly chosen neighboring agent. This paper shows how GSO algorithm can be significantly improved by simple modifications: agents receive an alternative method of changing the position by random jumps and they prefer directions consistent with the direction of the previous movement.

**Keywords:** Swarm intelligence, Glowworm Swarm Optimization, Multi-modal function optimization

### Introduction

Usually, we think of optimization in the sense of choosing the best element from many available alternatives. However, there are situations where a problem has multiple similar solutions, which are global and local optimums. In order to find all good solutions of a problem, one has to use multi-modal optimization and special algorithms, because multiple use of a classical optimization algorithm may result in obtaining the same single solution.

Multi-modal optimization is especially important from a practical point of view: e.g. similar results of the production process can be obtained under different conditions, but some of them may be easier to achieve on an industrial scale. Moreover, the multi-modal optimization algorithm can be used for an efficient control of robots, which must cooperate in order to fulfill the task of locating multiple sources of e.g. thermal radiation. The creation of fast and efficient algorithm is therefore an important task.

In general, evolutionary algorithms are able to provide multiple answers to an optimization problem, because they are simultaneously exploring a search space at many points. For example, Niching Particle Swarm Optimization (NichePSO) algorithm [1], which is a variant of Particle Swarm Optimization algorithm [2] uses Niching [3] technique to divide simulated particles into subswarms in order to find many solutions. Glowworms Swarm Optimization [4-6] is another evolutionary multi-modal optimization algorithm, in which the work is done by a swarm of agents, which behave like glowworms. The GSO agent encodes the fitness of its own position by a level of *Luciferin* (a luminance quantity)

and moves step-by-step toward randomly chosen agent, which broadcasts higher *Luciferin* level (i.e. “shines brighter”). It must be emphasized, that a simulated swarm of glowworms is protected against convergence at a point of the global optimum by limiting the distance of an information exchange: the agent can look for “brighter” glowworms only in its own vicinity. The GSO algorithm is given in [6] in a form of a pseudo-code.

### Goal of this work

Despite many positive attributes, the GSO algorithm has its downsides as well:

- in the case of low concentration of agents, many of them are not able to find the neighbors and in consequence they are unable to move at all,
- agents can move only in direction of other agents. It means, that many agents have to collaborate in order to effectively look for a single solution (in a simple, two-dimensional case, at least 3 agents are required). The general rule is: the more dimensions of the search space, the more agents have to cooperate.
- GSO has not got a built-in stopping condition, so the optimization process is terminated after an arbitrary chosen number of iterations. This in turn means, that the final location of agents may suggest successful finding of an optimum only when a certain number of agents are grouped together. In

particular, the location of a single agent has no meaning, due to lack of an information, why a simulated glowworm was stopped at that particular place.

The first two drawbacks can be bypassed by a simple improvement, which was proposed by the author last year [7]. In modified GSO algorithm (GSOv1), always, when an agent has no neighbors (i.e. glowworms, which broadcast higher value of the *Luciferin* and are in its close vicinity), it performs random step in the search of a better position (this part of the algorithm will be called a *backup algorithm*). In the *backup algorithm* only steps improving current solution are accepted, all the others are discarded. This modification significantly improves an efficiency of the algorithm, measured here by the number of correctly detected optimums.

This paper is a continuation of earlier work [7] and is aimed at improvement of GSO algorithm by adding a conservation of agent move direction.

### Description of a problem solution

The GSOv1 algorithm (the standard Glowworm Swarm Optimization plus the *backup algorithm*) was further improved by adding a capability of conservation of agent move direction. In the new algorithm, always, when an agent has a choice between multiple directions of the next move (either there are several neighboring glowworms, which “shine brighter” or a new position is generated by random by the *backup algorithm*), the greatest probability of the choice have these directions, which do not deviate too much from the direction of a previous motion of this agent. *New modification changes only the probability of the next move selection; the algorithm core (GSOv1 presented in [7]) is not altered.*

In the new, described here, version of the algorithm (GSOv1C), the last change of agent location is stored in a memory in a form of a unit vector. This vector represents orientation of the agent ( $O$ ) and is always overwritten, when the agent changes location. Directions consistent with the direction of the vector  $O$  are promoted by scaling the probability of their selection by the following weighting factor:

$$W_f = 0.495 \left[ \left( \frac{N \times O}{\|N\|} - 1 \right) + 2 \right] + 0.01$$

where:

- $N$  is a vector from a current position of the agent to a new, evaluated location,
- $\| \cdot \|$  denotes the length of the vector.

A constant 0.01 makes the weighting factor always larger than zero, what prevents algorithm from being blocked, when return to the last position (U-turn) is the only available option (such a situation can happen, when several agents are placed on the edge of the search space). A constant 0.495 makes the weighting factor not larger than 1.0.

Additionally, if a current move of a certain agent is governed by the random *backup algorithm*, the length of the vector  $O$  is reduced by the factor 0.9 always, when new location is not accepted (i.e. when it leads to worse solutions). By means of this feature, the new algorithm has an ability to decrease influence of the old orientation, when it becomes out of date.

In Figure 1 the weighting factor is plotted versus angle between  $O$  and  $N$  vectors. The shape of the curve depends on the number of previous positions discarded by the *backup algorithm*.

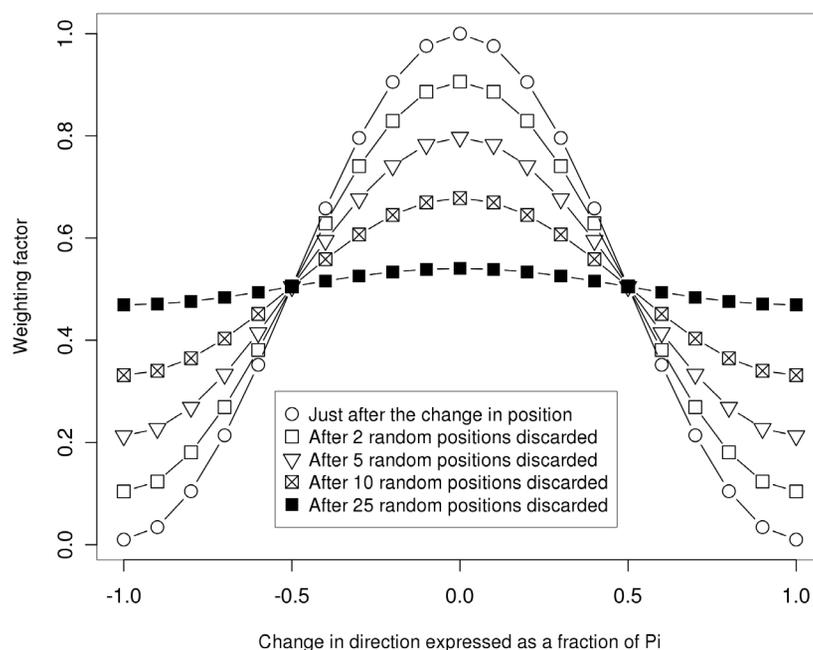


Fig. 1. Weighting factor as a function of an angle

It must be emphasized, that a concept of a reduction of the  $\mathbf{O}$  vector length has very important side-effect. When, at the end of calculations, a single, solitary agent owns very short  $\mathbf{O}$  vector, it means, that this agent successfully found the optimum! Previously, this particular agent repeatedly tried to move by the random jumps generated by the *backup algorithm* (only this part of the GSOv1C is allowed to decrease the length of the  $\mathbf{O}$  vector), but could not, because all nearby locations were worse, so the optimum was located.

## Software and hardware used

The software used for computer experiments was written in GNU R language [8]. Parameters and results of the calculations were stored in the relational database SQLite [9] by means of RSQLite library [10]. Due to the rich plotting and charting capabilities, GNU R language was also used for figures preparation.

The SQLite database was divided into three files. All tables required for planning calculations were placed in the first file. The second file contained information about the progress of calculations and crude results. The processed, averaged results were inserted in tables saved in the last file. The information, stored in these files, was always added and never modified. Therefore, the new version of the database files contained all previous data, so it was possible to simultaneously plan for new calculations, to do them and to analyze the result of previous computer experiments.

Main calculations were performed on zeus.cyfronet.pl and mars.cyfronet.pl clusters at ACK Cyfronet AGH in Cracow. Results were elaborated by a standard PC computer.

## Results

Three versions of GSO algorithm (original GSO, GSOv1 and GSOv1C) were tested by means of many well-known multimodal test functions. Presented results were received for a small swarm composed of 25 agents (glowworms). All calculations were terminated after 2000 iterations. GSO model constants were taken as in paper [7].

The following two-dimensional test functions were arbitrary chosen to show results in this paper: Peaks function [11], Rastrigin's function [12], Himmelblau's function [1], Equal-peaks-B function [13], Ackley's Path function [14], Branin's rocs function [14], Easom's function [14] and Fifth function of De Jong [15]. All optimums of these functions are known. Some of the test functions were re-scaled in order to limit the search space, which was usually a square  $10 \times 10$ .

The algorithms were compared due to the number of localized optimums. Because the test functions differ in number of optimums and calculations were repeated 100 times, the normalized average number of localized optimums (denoted as:  $\langle N_{loc} \rangle / N_{opt}$ ) was shown in Figures 2-7 as a function of iteration number.

The normalized average number of the localized optimums was evaluated three times, because of different success criteria:

- according to criteria A) optimum was treated as found when at least three agents were localized in its vicinity. This requirement was used in papers [6] and [7].
- according to criteria B) even one agent near optimum was enough to consider that optimum as found.
- according to criteria C) optimum was regarded as discovered when either condition A) was fulfilled or at least one agent with the orientation vector  $\mathbf{O}$  shorter than 0.001 was

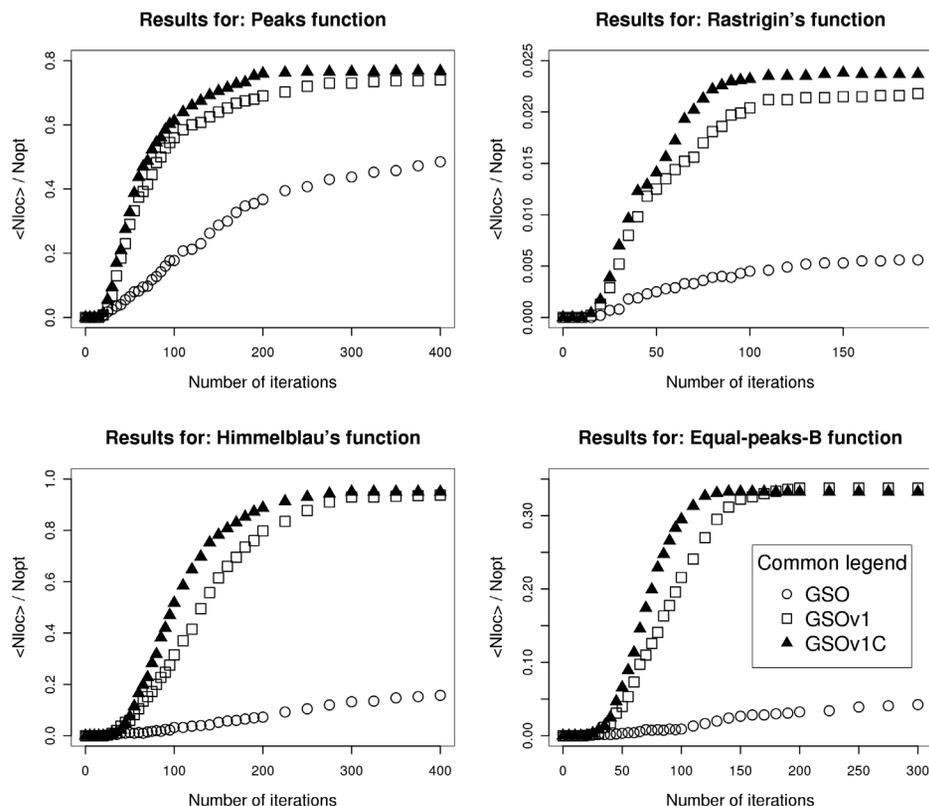


Fig. 2. The performance of the algorithms evaluated according to criteria A)

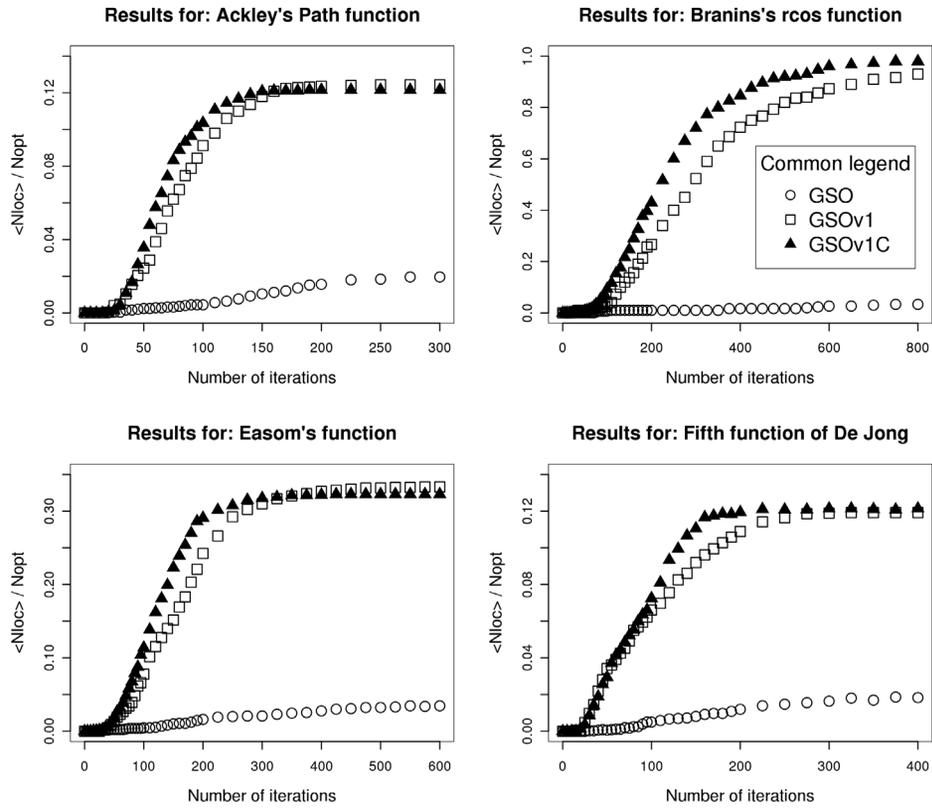


Fig. 3. The performance of the algorithms evaluated according to criteria A)

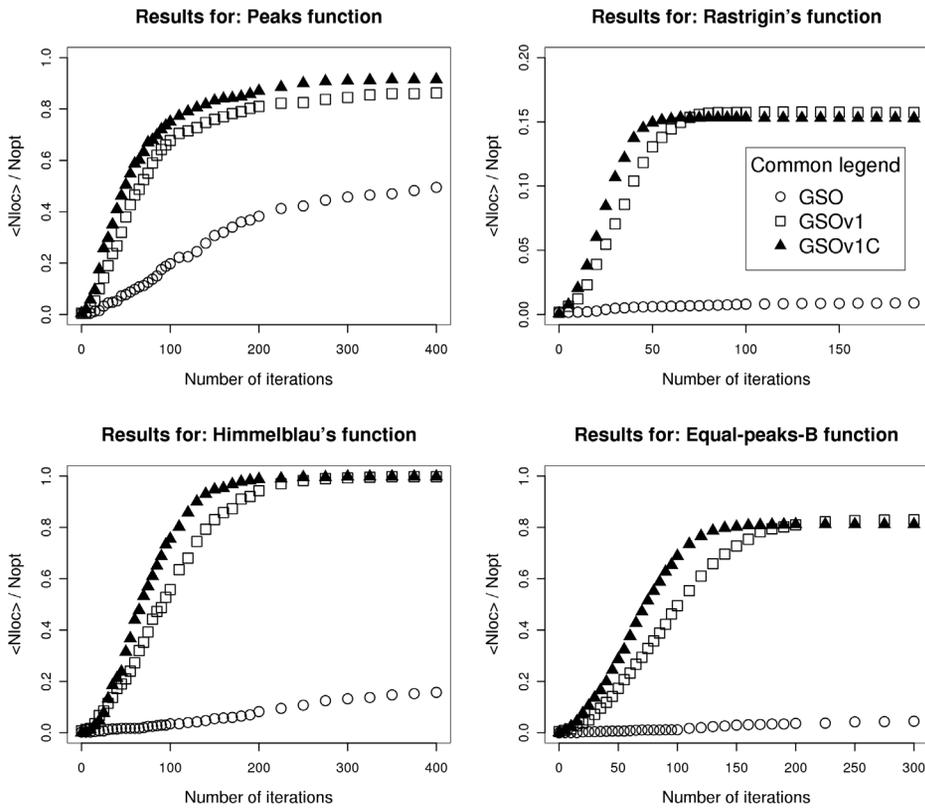


Fig. 4. The performance of the algorithms evaluated according to criteria B)

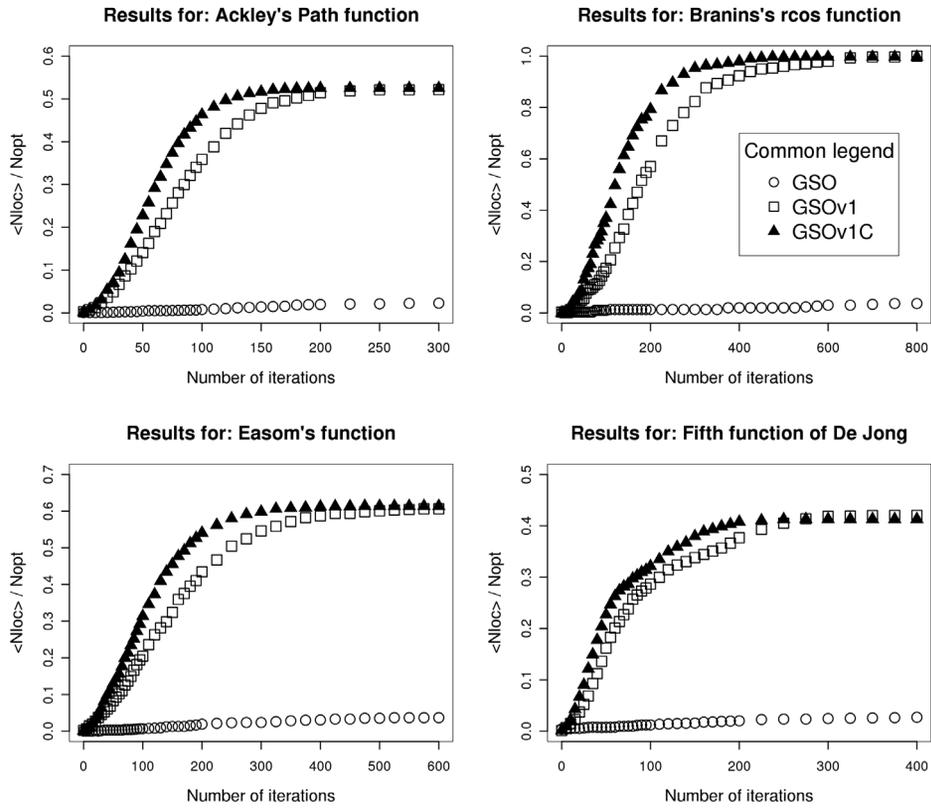


Fig. 5. The performance of the algorithms evaluated according to criteria B)

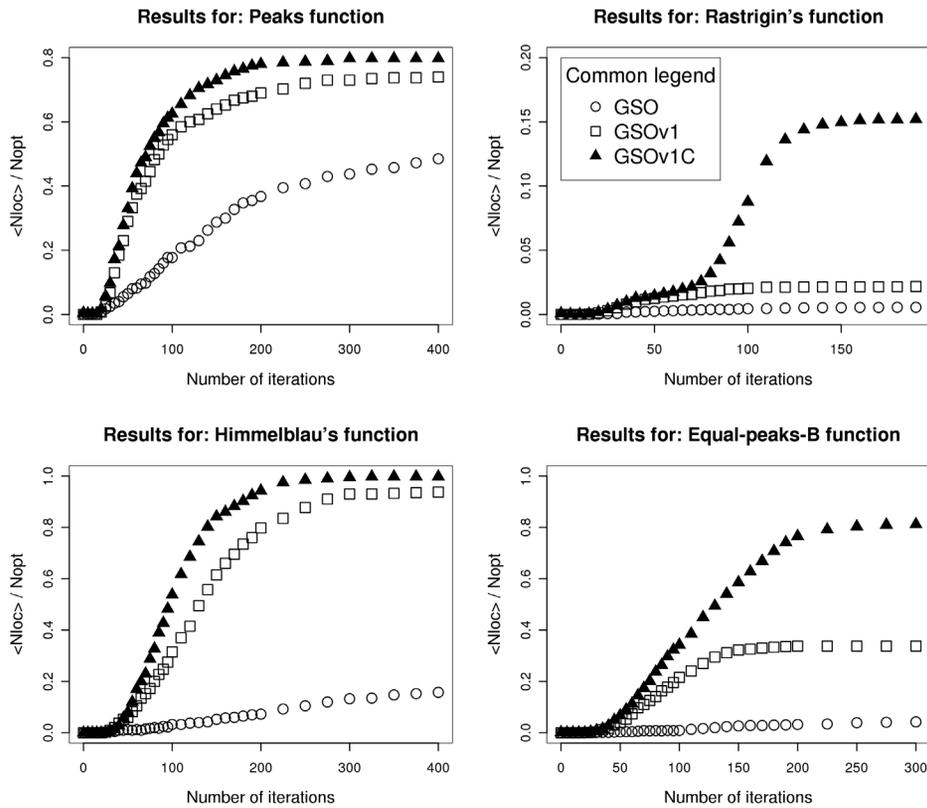


Fig. 6. The performance of the algorithms evaluated according to criteria C)

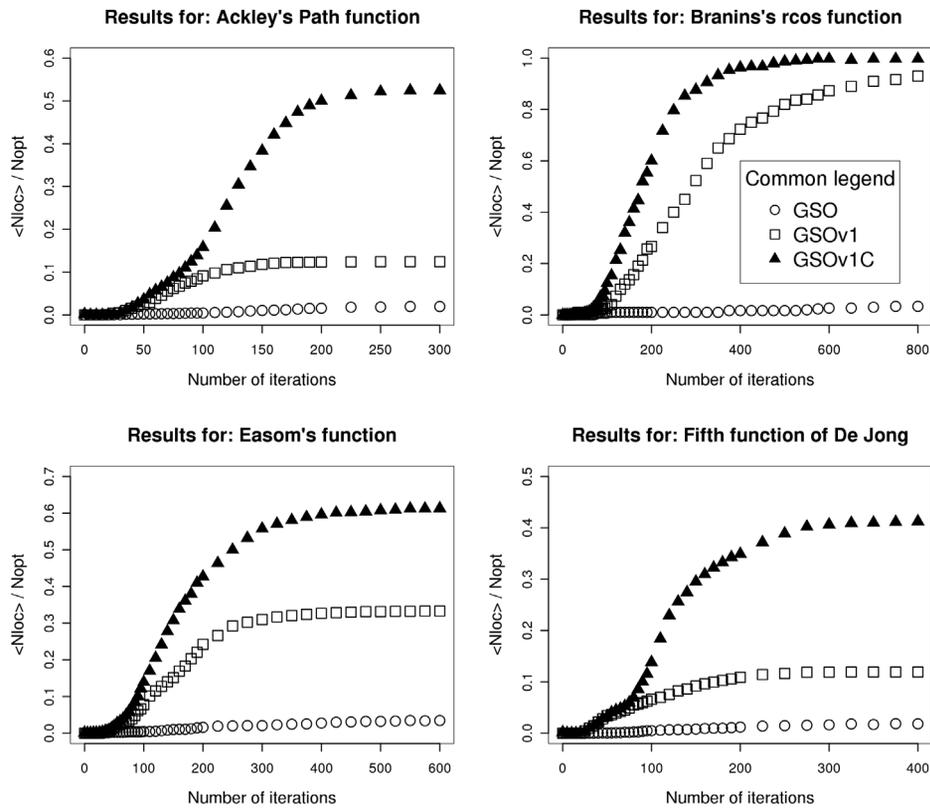


Fig. 7. The performance of the algorithms evaluated according to criteria C)

found near optimum. The second part of this condition can be accomplished only in the case of the new version of GSO algorithm (GSOv1C). However, the introduction of this measure is justified, because only this one algorithm provides the direct information needed to correlate the positions of the single agent and the optimum.

In all cases only agents, which distance to the nearest optimum was shorter than 0.05 (a step size was equal to 0.03), were taken into account.

In all Figures below the original GSO algorithm is marked by means of open circles, GSOv1 by open squares and for the new algorithm (GSOv1C) filled triangles are used.

In Figures 2 and 3 the performance of the algorithms evaluated according to criteria A) is shown. It can be easily seen, that algorithms GSOv1 and GSOv1C are comparable in terms of the number of the localized optimums, but GSOv1C is faster and requires less iterations to finish the work. Of course, it is a direct result of more efficient choice of the next agent position.

Results received with the usage of the criteria B), which is a impaired version of the criteria A), are given in Figures 4 and 5. Both GSOv1 and GSOv1C have the *backup algorithm* implemented and they are able to explore the search space even by means of a single agent, so they gain in comparison with the original GSO algorithm and now yield significantly more localized optimums.

As it was previously described, in the case of GSOv1C algorithm, it was enough to check the length of the orientation vector  $\mathbf{O}$  for unambiguous assessment of whether a single agent was

able to find the optimum. This feature positively distinguishes GSOv1C from other versions of GSO, what can be observed in Figures 6 and 7, where criteria C) was used. The same level of the normalized average number of the localized optimums was achieved by GSOv1C quicker in comparison to GSO and GSOv1. Moreover, for GSOv1C the final values of  $\langle N_{loc} \rangle / N_{opt}$  parameter, were the largest.

## 5. Conclusions and future work

It has been shown, that the standard GSO algorithm was significantly improved by simple modifications.

The new algorithm GSOv1C is able to locate more optimums than standard GSO. Moreover, GSOv1C requires less computer resources (uses smaller number of the iterations) than previous version (GSOv1) to obtain the same result. Excellent performance is achieved because GSOv1C efficiently uses all agents and explore the search space even by means of a single agent. The new, proposed in this paper, version of GSO uses memorized direction of the last agent movement in order to reduce the chance of a return to the old solution. Only GSOv1C can clearly indicate the optimum by means of a single agent.

Recent results received for 5, 10 and 15-dimensional test functions also indicate the superiority of the algorithm in the version GSOv1C.

A sensor range is a very important GSO model constant. It limits the resolution of calculations - two different optimums can be

detected, if the distance between them is larger than the sensor range. The conversion of this constant to a variable dynamically adjusted during calculations is a goal for the future work.

## Acknowledgments

The research reported here was partially supported by the grant no. MniSW/SGI4700/UJ/126/2007.

## References

1. Brits R., Engelbrecht A.P., van den Bergh F.: A niching particle swarm optimizer. Proceedings of the 4th Asia-Pacific conference on simulated evolution and learning, 2002, pp. 692-696.
2. Kennedy J., Eberhart R.: Particle Swarm Optimization. Proc. IEEE International Conference on Neural Networks, 1995, pp. IV: 1942-1948.
3. Mahfoud S.W.: Niching methods for genetic algorithms. Doctoral dissertation, University of Illinois at Urbana-Champaign. *Dissertation Abstracts International* 1995, 56(9), 4987B (University Microfilms No. 9543663).
4. Krishnanand K.N., Ghose D.: Detection of multiple source locations using a glowworm metaphor with applications to collective robotics. IEEE Swarm Intelligence Symposium, Pasadena, California, USA, 2005, pp. 84-91.
5. Krishnanand K.N., Ghose D.: Glowworm swarm based optimization algorithm for multimodal functions with collective robotics applications. *Multi-agent and Grid Systems* 2006, 2(3): 209-222.
6. Krishnanand K.N., Ghose D.: Glowworm swarm optimization for simultaneous capture of multiple local optima of multimodal functions. *Swarm Intelligence* 2009, 3, 2: 87-124.
7. Oramus P.: Improvements to glowworm swarm optimization algorithm. *Computer Science* 2010, 11: 7-20.
8. <http://www.r-project.org/> (accessed 2011-03-30).
9. <http://www.sqlite.org/> (accessed 2011-03-30).
10. <http://cran.r-project.org/web/packages/RSQLite/index.html> (accessed 2011-03-30).
11. Reutskiy S.Y., Chen C.S.: Approximation of multivariate functions and evaluation of particular solutions using Chebyshev polynomial and trigonometric basis functions. *International Journal for Numerical Methods in Engineering* 2006, 67(13): 1811-1829.
12. Törn A., Zilinskas A.: Global optimization. New York: Springer, 1989.
13. Parsopoulos K., Vrahatis M.N.: On the computation of all global minimizers through particle swarm optimization. *IEEE Transactions on Evolutionary Computation* 2004, 8(3): 211-224.
14. <http://www.geatbx.com/docu/fcnindex-01.html> (accessed 2011-03-30).
15. <http://www.zsd.ict.pwr.wroc.pl/files/docs/functions.pdf> (accessed 2011-03-30).



## GRAND CHALLENGES LESS CHALLENGING: NEW POSSIBILITIES PROVIDED BY GRAPHICS PROCESSING UNITS

DAWID KUNA<sup>A</sup>, MARCIN MAKOWSKI<sup>A,B</sup>, GRZEGORZ MAZURA<sup>A,C</sup>,  
 PAWEŁ RUSSEK<sup>A</sup>, MARCIN JANISZEWSKI<sup>A</sup>, KAZIMIERZ WIATR<sup>A</sup>,

<sup>a</sup> ACC CYFRONET AGH, Nawojki 11, 30-950 Kraków, Poland; <sup>b</sup> Department of Theoretical Chemistry, Jagiellonian University, Ingardena 3, Kraków, 30-060, Poland; <sup>c</sup> Department of Computational Methods in Chemistry, Jagiellonian University, Ingardena 3, Kraków, 30-060, Poland

**Abstract:** The specific features of general-purpose computing on graphics processing units (GPGPU) in the area of scientific computing are highlighted. The hardware and software resources provided by CYFRONET are presented and selected benchmark calculations are briefly introduced.

**Keywords:** GPGPU, molecular modelling, molecular mechanics, quantum chemistry

### 1. Short history

There was a time when computer graphics were generated solely on the Central Processing Units (CPUs) in a computer's memory and then, if necessary, displayed on screens by video cards or printed on paper by printers. However, as larger and larger image resolutions were required, and processing became more complex as more sophisticated models were used, the CPUs reached a performance limit and couldn't generate the graphics in real-time anymore with a reasonable number of frames per second. The response came from hardware manufacturers who designed dedicated hardware to off-load the CPUs from common graphics operations. Some of those devices targeted the entertainment market, which is known for its persistent need of higher and higher graphics quality. This brought a constant growing acceptance of hardware acceleration, enabling manufacturers to gradually design more advanced devices. This trend led to the Graphics Processing Units (GPUs). In their present form they are far from their first ancestors'.

Some basic 3D operations are easy to understand and implement, but nevertheless put a lot of strain on the CPU, because they require a lot of random memory accesses and lots of arithmetic calculations. These 3D operations include for example filling a triangle's interior with an image (while keeping a correct angle and having it adequately stretched), filling a triangle with a colour gradient coming from interpolation of its three vertex colours, etc. It is not enough to interpolate either colours or texture coordinates linearly, as it doesn't realistically reflect the

perspective of the introduced artefacts. Nonlinear interpolation is necessary, however, it requires the use of more time-expensive operations that are especially slow if performed by the CPU.

Later, the graphics processing hardware was augmented with a functionality to perform 3D coordinate transformations. So that the hardware was able to accelerate a larger part of graphics generation, not only polygon filling. Some stages in the GPU's multi-stage graphics generation process (i.e. graphics processing pipeline) contained certain dedicated circuits that were absent at every other stage. This meant the customisation of stages of the processing units that depended on performance functionality. The units implementing different pipeline stages varied. For example: the rotating points in 3D space differ from putting a texture onto a polygon in terms of the required resources. Recently, a new trend has been observed: that is to replace the dedicated units with general-purpose resources whenever it is feasible and reasonable. This approach simplified the GPU programming for tasks not related to graphics. Meanwhile, the GPU companies provided the software developers community with programming models. This allowed the GPU programming to perform various, typically computationally intensive tasks. Having gradually more general units, the GPUs became in some aspects similar to the CPUs, more approachable and easier to use as a simple co-processor. At the same time, the GPUs still possess their inherent property of being able to process a lot of independent data using common operation sequences, i.e. they have a large number of computational units, far exceeding that which is soon to be seen on the CPUs, they can also efficiently

execute simple computational tasks (kernels) over large sets of data, element by element.

The GPUs are not meant as a replacement for the CPUs, due to their different characteristics. The GPU processing units are simple in-order execution units. There is no sophisticated memory model implemented in the GPU, contrary to the modern multi-CPU architecture. The GPU can offer much higher memory bandwidth as memory access is simplified. The computational throughput is provided by the GPUs in a plain, inflexible form.

As their predecessors have done this for graphics, the modern GPUs can significantly reduce the CPUs load in various computational applications. To get noticeably shorter total computation times, the operation sequence has to be expressed in a way suitable for the GPUs and map well to its resources. The GPUs are usually installed in the system as add-on cards, so there are two major usage scenarios: the GPU as a replacement or an extension to the CPU. The first one is easier to implement: at the beginning of a computationally intensive application part, all the data is written to the GPU memory, then the GPU starts data processing, and the CPU is waiting idly for the GPU to finish, finally the results are copied back to the system memory. The latter one assumes some sort of cooperation between the CPU and the GPU, either by splitting the problem into pieces and distributing them among the devices in the system, or by working simultaneously on different tasks. The higher the complexity of the software pays off by providing better application performance overall.

Typically, a single modern GPU is connected to the rest of the computer system through a PCI-Express bus. All the data, results and commands have to be exchanged using that bus. In a sense it resembles a separate node in a computer cluster where the explicit message and data exchange is implemented. Some communication fabric connects devices in a cluster and this structure is a performance bottleneck because it introduces delays in communication.

Recently, the GPU-based systems have made their way to the computing centres establishing the General-Purpose computing on Graphics Processing Units (GPGPU) paradigm. This triggered the development and deployment of hardware-accelerated scientific software. In the following sections a short introduction to a GPGPU scientific computation is given, a brief overview of the GPU systems available at CYFRONET is presented and finally the results of the benchmark molecular modelling calculations are shown. Finally, the resources for development of the GPU-accelerated scientific software are briefly discussed.

## 2. The GPGPU scientific computations

The software vendors and the open source projects' contributors noticed the potential to improve the execution time in several areas when the GPUs are present in a system. As for now, support varies for the GPUs. Some software packages use them for computations which are both mostly used and the most time-consuming, others provide just a selected few computation methods on the GPUs, having the rest done by the CPUs. The latter may not address the usual users' needs, but it's a good point to begin with. Next the selected scientific tools are presented,

grouped by their domain. In [2] one can find a similar survey, focused more on molecular modelling.

### 2.1 Quantum chemistry

In quantum chemistry, to explore the chemical properties of molecules, the laws of quantum physics are used. The model representing a real system is detailed enough to account for the presence of electrons. Usually, the Schrodinger equation is being used to obtain the probable distribution of the electrons' locations. Having the molecule's electronic structure is the entry point to discover the molecule's properties which less accurate models cannot provide. Several approaches have been developed to approximately solve the system, with the accuracy good enough. However, the detailed model's nature inherent to quantum chemistry obviously limits the molecule size that can be computed, with the limiting factor being computing time. To solve the Schrodinger equation, the most well-known methods are the Hartree-Fock with its variants and density functional theory.

The problem maps well to the GPUs ([5, 6]), which led to the design of the TeraChem software. Unlike a lot of software written for the CPUs and later adapted to the GPUs, from the beginning TeraChem was targeted toward the modern nVidia GPUs. As a consequence, significant speed-ups have been observed when compared to software performing similar computations solely on the CPUs. Especially for larger molecules speed-ups are easily observable. TeraChem supports, among others, the Hartree-Fock (Restricted and Unrestricted), the DFT method, the TIP3P force field for modeling how water in fluencies explored molecules.

### 2.2 Molecular mechanics

When molecules are too large to be solved with quantum chemistry or QC specific properties are not required, less accurate models can suffice. This is often the case in biology, where for many applications molecules can be modelled as balls connected with springs. Balls represent molecule's atoms, while springs are there to indicate bonded interactions between atoms. When the physics follows Newtonian mechanics, the model is called molecular mechanics. Interactions are expressed through so called 'fields' that, given two atoms, return the combined mutual force acting upon these atoms. So, the resultant force applied to an atom is equal to the sum of all the partial forces coming from interactions with other atoms. It takes  $O(n^2)$  field evaluations, where  $n$  denotes the number of atoms in the molecule. As there are few kinds of atom-atom interactions, the combined force can be split into components:  $F_{total} = F_{bonded} + F_{nonbonded}$ , where  $F_{bonded}$  relates to bonds linking adjacent atoms, and  $F_{nonbonded}$  relates to interactions between all atoms, such as electrostatic forces. Saving on insignificant computations is possible by skipping nonbonded forces for distant atoms, since nonbonded interactions decline with the distance growth.

One of the possible optimisations is to use the Particle-Mesh Ewald method ([7]), giving high accuracy at significantly reduced computational cost. The approximation works by dividing the potential into the direct and the reciprocal sum. The direct sum, related to atoms close enough, is computed directly. The recip-

cal sum, on the other hand, is handled by introducing a grid, onto which charge density coming from further atoms is interpolated. Then, the Fourier Transform on the grid is performed, followed by convolution and the Inverse Fourier Transform, returning potential on the grid. The reciprocal sum relates to long-range interactions. The PME assumes periodic boundary conditions. Molecular mechanics is well-suited to the GPU implementations. It is a building block for the molecular dynamics software packages, where the MM is employed for a large number of simulation frames. The computational complexity combined with a need to perform computations many times per simulation put high performance demands on the MM codes.

## 2.3 Molecular dynamics

For several applications, the computing of a system at an isolated point in time is of no use by itself, but it is necessary as a building block for larger computations. There is a need to observe how the systems behave, and evolve over time. When interactions between the system elements follow molecular mechanics and the system is simulated in consecutive time slots, the computations are called molecular dynamics (MD).

MD is a major tool to explore how biomolecules interact. E.g. knowing and understanding protein-protein docking (i.e. forming complexes from two or more biological molecules) allows us to understand the processes occurring in living organisms, it is also of huge importance for drug design. Besides its significance in biological sciences, MD is also used heavily in material sciences, to investigate material's properties, such as its melting temperature, analyse crystal behavior in the presence of defects, determine the structure of X-ray crystallography as well as of Nuclear Magnetic Resonance (NMR) experiments, to name just a few of its applications.

For MD to return realistic results, the timestep has to be set appropriately to be in the order of femtoseconds. Larger timesteps lead to skipping important intermediate system stages, leading to states diverging from reality. A small and not varying timestep means the simulation length is severely limited by the available computing capacity.

There are many scientific applications for MD. Of the most well-known and able to use the GPUs, there are the Gromacs, the NAMD and the Amber. The Gromacs 4.5 support the GPUs through the OpenMM library, currently only the subset of functionality benefit from the GPUs in the system. Nevertheless, authors point to a large performance improvement for the implicit solvent when compared to the CPU implementation. Support for more features is under development. The NAMD is an MD package, geared towards scalability on parallel architectures. A 2.7 version supports the GPU. The Amber package is also capable of performing MD simulations. With version 11, the Particle-Mesh Ewald method is supported on the GPUs, with some sub-features not yet implemented. This is still a work in progress.

## 2.4 Aminoacid sequence alignment

In the area of searching for amino acid sequence similarity, it is not enough to check whether two sequences are exactly the same

or not. Rather than that, one needs to know to what extent one sequence resembles another one and how many subsequences they have in common.

Probably the most widely used tools are the BLAST, the FASTA, and the Smith-Waterman algorithm. All of them need to be implemented as quickly as possible, due to the amount of data to process. For the former ([3]) and the latter one (e.g. [4]), the GPU versions have been implemented with noticeable speed-ups. The three tools are not completely interchangeable, as they differ in provided accuracy. The Smith-Waterman algorithm returns optimal matches, thus it is the slowest one and requires the most memory. The two others use the heuristic method to increase the processing speed at the cost of lower accuracy – the returned results might not be optimal. The HMMER, a similar tool, has also been ported to the GPU ([1]).

## 2.5 Other applications

Apart from ready-made software, like the molecular modelling packages described above, the GPU cluster serves as a development environment for high-performance scientific applications. To facilitate implementation of the GPU-aware code, the CUDA C and the OpenCL compilers are provided with a set of advanced libraries as well, allowing for easy delegation of computationally-demanding kernels to the graphic subsystem. Among others, the Niedoida [11] computational chemistry package is being extended to make use of the GPU-accelerated kernels implementing core ab-initio and the Density Functional Theory (DFT) algorithms.

## 3 CYFRONET GP GPU resources

In April 2011 ACC Cyfronet introduced users to a new computer system equipped with NVIDIA® Tesla™ GPU cards. This was the newest addition to the most powerful Polish supercomputer cluster "Zeus". GPU based system, which is intended to solve biochemical, geological, pharmacological and materials science problems. Thanks to its outstanding performance much larger scale problems can also be solved.

The system is composed of 24 computing nodes each equipped with two NVIDIA® Tesla™ M2050 graphic cards. The combined performance of a single computing node is 1.1 Tflop which includes the processing power of 12 Intel Xeon X5670 CPU cores and 896 NVIDIA CUDA cores. Overall the computing power of the entire GPU cluster is comparable with roughly 176 computing nodes with exact configuration excluding graphic cards. The nodes are connected using low-latency high bandwidth Infiniband interconnect to facilitate efficient message passing. The specific configuration was selected to optimally support the variety of scientific software, with each specific application having specific preferences with respect to the number of CPU cores and GPU subsystems. In particular, the cluster allows for efficient modelling of large molecular systems at both molecular and quantum mechanics levels.

Theoretical computing power can be calculated for every processor model. The actual computing power, however, is much harder to describe because it depends on algorithm properties. Moreover, achieving theoretical values is impossible in most

cases. In the case of graphic cards this task is even harder due to the architectural and the programming complexities. The problems in this field are the main focus of a newly created Center of Competence in the field of GPU programming. Research is carried out with the support of the „HPC Infrastructure for Grand Challenges of Science and Engineering” Project.

Current research is focused on chemistry-related problems including:

- The development of GPU-accelerated quantum-chemistry software
- Molecular modelling of macrocyclic inclusion complexes, particularly the examination of the influence of ions on the formation of guest/host systems
- Modelling the influence of the ketocyanine dye-solvent interaction on the HOMO-LUMO gap

The Zeus GPU cluster provides many popular software packages for scientific calculations. These include the NAMD, the Gromacs, the Gamess, the TeraChem and others. The full list of available packages is provided at [12].

#### 4. Results of experiments on Cyfronet's GPGPU

The new ACC Cyfronet installation provides an excellent environment for performing large-scale scientific calculations. The installed GPU-aware software includes molecular mechanics (NAMD [8]) and quantum-chemical (TeraChem [9]) packages. To illustrate the efficiency gained stemming from the hardware-assisted acceleration, selected benchmark calculations are presented.

Molecular dynamics calculations were performed for the Satellite Tobacco Mosaic Virus (STMV, Fig. 1) in an explicit water environment. The system comprises of over one million atoms. The Langevin dynamics with a timestep of 1 fs was performed with periodic boundary conditions. The Particle Mesh Ewald (PME, [7]) method was used for calculating the Coulomb interactions. A series of calculations was performed for the varying number of nodes, with and without GPU acceleration. The results (see Fig. 2) show consistent almost three-fold speed-up stemming from the hardware-assisted acceleration.

To assess the performance of typical quantum-chemical calculations, the Born-Oppenheimer dynamics of olestra (see Fig. 3) was performed using the 6-31G basis set at the B3LYP level of theory. A comparison was made with the same calculations performed using the GAMESS [10]. The result shows that the TeraChem running on two GPUs is over 170 times faster than the GAMESS using 12 CPU cores.

#### 5. Summary

The modern graphics processing units can be harnessed to provide tremendous acceleration for the various numerically intensive scientific tasks. Specifically, the molecular modelling applications are well suited to the GPUs, due to their extensive computational requirements and immanent data-parallelism. The benchmark calculations performed using the GPU resources offered by CYFRONET show that very substantial gains in efficiency

can be expected in simulations based both on the molecular and the quantum mechanics paradigms.

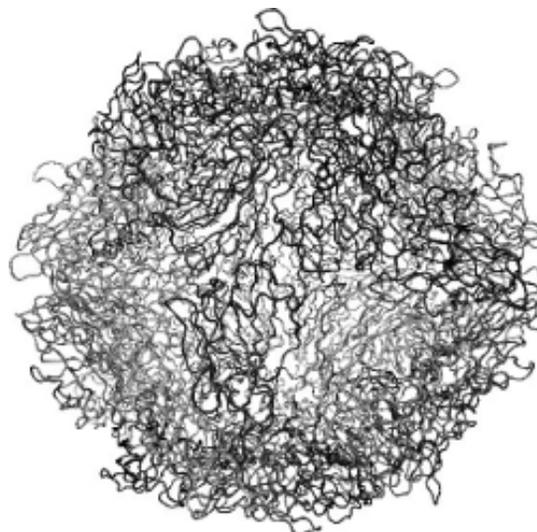


Fig. 1. Satellite Tobacco Mosaic Virus

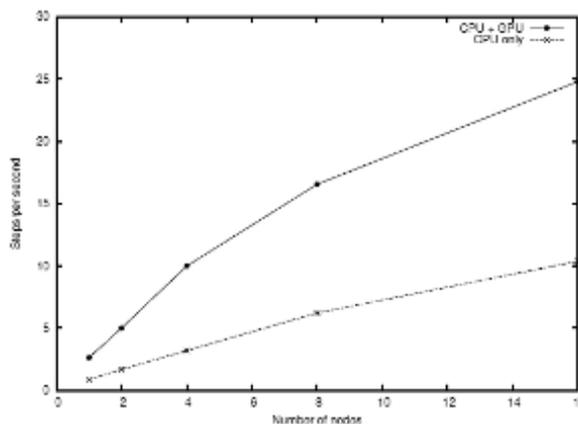


Fig. 2. NAMD performance with respect to the number of nodes (see text for details)

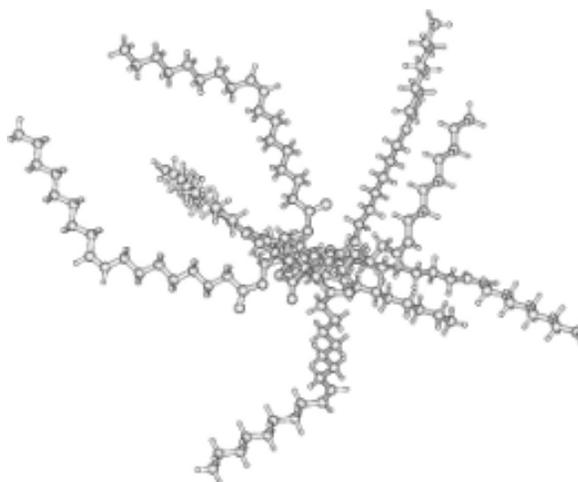


Fig. 3. Olestra molecule

## Acknowledgements

This work was carried out thanks to the "POWIEW" POIG.02.03.00-00-018/08-00 project. This project is co-funded by the European Regional Development Fund (ERDF) as a part of the Innovative Economy program.

## References

1. Walters J.P., Balu V., Kompalli S., Chaudhary V.: Evaluating the use of GPUs in Liver Image Segmentation and HMMER Database Searches. Proceedings of the 2009 IEEE International Symposium on Parallel & Distributed Processing, 2009, pp. 1-12.
2. Stone J.E., Hardy D.J., Ufimtsev I.S., Schulten K.: GPU-accelerated molecular modeling coming of age. *Journal of Molecular Graphics and Modelling* 2010, 29: 116-125.
3. Vouzis P.D., Sahinidis N.V.: GPU-BLAST: using graphics processors to accelerate protein sequence alignment. *Bioinformatics* 2001, 27, 2: 182-188.
4. Yongchao L., Maskell D.L., Schmidt B.: CUDASW++: optimizing Smith-Waterman sequence database searches for CUDA-enabled graphics processing units. *BMC Research Notes* 2009, 2(73). Retrieved 16 August 2011 from <http://www.biomedcentral.com/1756-0500/2/73>.
5. Ufimtsev I.S., Martinez T.J.: Graphical Processing Units for Quantum Chemistry. *Computing in Science & Engineering* 2008, 10, 6: 26-34.
6. Guochun S., Kindratenko V., Ufimtsev I., Martinez T.: Direct self-consistent field computations on GPU clusters, Parallel & Distributed Processing (IPDPS). 2010 IEEE International Symposium, 19-23 April 2010, pp. 1-8.
7. Toukmaji A.Y., Board J.A. Jr.: Ewald Summation Techniques in Perspective: A Survey. *Computer Physics Communications* 1996, 95: 73-92.
8. Phillips J.C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R.D., Kale L., Schulten K.: Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* 2005, 26: 1781-1802.
9. Ufimtsev I.S., Martinez T.J.: Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients and First Principles Molecular Dynamics. *Journal Chemical Theory and Computations* 2009, 5: 2619.
10. Schmidt M.W., Baldrige K.K., Boatz J.A., Elbert S.T., Gordon M.S., Jensen J.H., Koseki S., Matsunaga N., Nguyen K.A., Su S., Windus T.L., Dupuis M., Montgomery J.A.: General Atomic and Molecular Electronic Structure System. *Journal of Computational Chemistry* 1993, 14: 1347-1363.
11. Mazur G., Makowski M.: Development and Optimization of Computational Chemistry Algorithms. *Computing and Informatics* 2009, 28: 115-125.
12. <http://www.wielkiewyzwania.pl>.



## APPLICATION OF STANDARD ALGORITHMS OF AUTOMATIC SIGNAL SEPARATION IN MEDICINE

AGNIESZKA NOWAK

*Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie agnowak@agh.edu.pl*

**Abstract:** In the present article various methods of automatic separation of acoustic signals have been described. The biggest focus was placed on two methods, Blind Signal Separation (BSS) and Independent Component Analysis (ICA). In order to verify the efficacy of these methods, selected separation algorithms have been used for deconvolution of a specially prepared sinusoidal and saw-tooth sound signals as well as natural signals such as recordings of human voice. The obtained results have been compared and presented. More accurate results have been acquired from the analysis of artificially prepared signals that is the sinusoidal and saw-tooth signals which were mixed together using numerical transformations. Due to the potential practical usage of speech signal separation in medicine, more stress has been put on the analysis of life taken signals, which were created by mixing voices of few persons speaking simultaneously. The assessment of the usability of different algorithms, which effected from the research, may have practical application due to the fact that in the available literature the authors usually limit themselves only to presenting (and praising) algorithms created on their own, scarcely mentioning algorithms of different authors predominantly without doing necessary comparative researches. These missing researches constitute the essential part of the work presented in this article.

**Keywords:** Algorithms of automatic signal separation, Blind Signal Separation (BSS), Independent Component Analysis (ICA), SOBI, AMUSE, SONS

### Introduction

There are many spheres of life and science in which separation of various signals is necessary. Very often these signals occur in the mixed form and have to be analyzed or processed separately. For example denoising of signals in order to enhance the quality of reception in cell phones, in automatic speech recognition or even espionage in which speech signal of a particular speaker has to be separated from other signals which are not the point of interest. Automatic speech signal separation might also be useful in medicine in automatic protocols creation. These protocols can be derived from various form of medical treatment while the doctor performing medical activities records his statements which are the basis of the required report. Because of the fact that microphone picks up not only doctor's voice but also surrounding noise (for example answers of the operating room nurses which are irrelevant for the protocol), the doctor's voice separation is indispensable before using the speech recognition systems.

Due to the reasons mentioned above many algorithms which allow for separation of various signals have been created. A few examples of such algorithms can be found in the literature appended in the present paper. These algorithms make use of

information obtained from the second or higher order statistics and from temporary-spatial decorrelation. Due to their popularity and effectiveness in separation of mixed signals, two methods have been employed in the research performed in this paper; these methods are Blind Signal Separation (BSS) [2, 6] and Independent Component Analysis (ICA) [8, 7]. Additionally different methods such as Principal Component Analysis (PCA) and Factor Analysis (FA) are described in the present paper but their significance is only auxiliary.

The choice of analyzed methods is caused by the fact that BSS is based on estimation of source signals derived only from the information about mixed signals, which have been observed. The ICA method is applied to these actions which are the conceptual fundamentals of all methods of signal separation. Both methods reflect real situations in which a person encounters only one mixture of signals and the need for their splitting occurred. As mentioned before these situations take place relatively often. For example when the need for tracking the speech of one person occurred while other people are speaking in the background. Automatic signal separation might also be used to prepare recorded signal for the speech-to-text application. Both of these

examples are further analyzed in the present article. Additionally, artificially generated signals have been used in the research in order to acquire more accurate assessment of the process of separation and its results. To be more specific different frequencies of sinusoidal and saw-tooth signals have been analyzed.

## 1. Algorithms of signal deconvolution

### 1.1. Blind Signal Separation (BSS)

Blind Signal Separation is based on recovering of independent source signals while given mixtures of these signals. The word "blind" suggests that there is no information about the source signals and the mixing process or the information is insufficient. This problem has become particularly important in the discipline of speech signal processing, for example in applications used for automatic recognition of the speaker or telecommunications as well as in processing of image or medical signals [3].

Many different algorithms devoted to the problem of blind signal separation exist, however, the main idea of this problem is usually similar for all of them. First of all it is taken under consideration that the sources (which are modeled as random despite their deterministic way of production) are statistically independent. The principal criterion of signal separation is the independence of the sources which is estimated by calculating some cost function  $J$ . Obtaining the extremum of this function means generating more or less independent component signals which are treated as solution. Obviously the parameters of unmixing process have to satisfy the assumptions in order to give correct results [6].

This general idea of the considered algorithm is shown in Fig. 1.

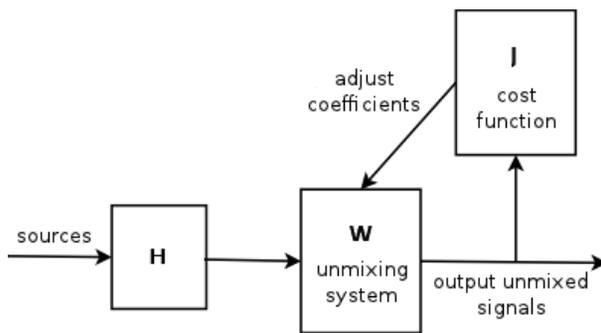


Fig. 1. Block diagram of separating system

From the mathematical point of view, the blind signal separation problem is based on observation of the recorded signals received in the output of the unknown MIMO (Multiple Input, Multiple Output) system which has mixed and filtered source signals. The formula for these collected data can be written as in 1, where  $k$  denotes discrete time.

$$x(k)=[x_1(k), \dots, x_m(k)]^T \tag{1}$$

The problem to be solved can be formulated as followed: To find (if exists and is stable) an inverse system to identify all the original signal sources or at least parts of them with previously defined properties. These signals are defined by 2.

$$s(k)=[s_1(k), \dots, s_n(k)]^T \tag{2}$$

The overall calculations are carried out only on the basis of the output signals obtained from the algorithm written as in equation 3 and recorded signals. The output vector  $y$  contains components as independent as possible and is the best approximation of the unknown source vector  $s$ .

$$y(k)=[y_1(k), \dots, y_n(k)]^T \tag{3}$$

Although the mixing signal process uses different physical or mathematical models depending on particular application, the simplest form of majority of blind signals separation's linear models may be written as a specific problem of the matrix decomposition. Given the observation matrix,  $X=[x(1), \dots, x(N)]$ , where  $X \in R^{(m \times N)}$ , factorization of the matrix is defined by the formula 4 [5].

$$X=HS+Z \tag{4}$$

where:

- $N$  – the number of available samples,
- $m$  – the number of observations,
- $n$  – the number of sources,
- $H \in R^{(m \times n)}$  – an unknown mixing matrix,
- $Z \in R^{(m \times N)}$  – an unknown matrix of noise and errors,
- $S=[s(1), \dots, s(N)]$ , where  $S \in R^{(n \times N)}$ , where – an unknown matrix of source signals with specific features or temporal structures.

Usually the arrays in the above equation define physical features of the system, hence they have to meet certain requirements.

Frequently in various applications the observation matrix  $X$  is represented by the vectors  $x(k)$ , where denotes current sample's number. Matrix equation 4 can be written as the system of linear equations. This equivalent form is given in 5.

$$x(k)=Hs(k)+z(k), \quad k=1,2,\dots,N \tag{5}$$

where  $x(k)$  and  $s(k)$  are defined respectively as in formula 1 and 2. Equation 5 can be written as  $x=Hs+z$  for any value of  $k$  as it will be used for simplicity later. However, one should take under consideration that the data is not always represented in the time domain but often in the time-frequency domain. Therefore the index  $k$  may have different meanings.

Algorithms of blind source separation may use various methods, for example Independent Component Analysis (ICA), Sparse Component Analysis (SCA) or Non-negative Matrix Factorization (NMF). However, one should keep in mind that these methods are only mathematical, mechanical formulas. Effective use of them depends on the researcher's a priori knowledge, common sense and appropriate preprocessing and postprocessing tools. During these transformations a relevant knowledge is indispensable for obtaining reliable and valid from the physical point of view

components. The sequence of three basic procedures used to perform blind signal separation is shown in Fig. 2 [5].

In the following section selected algorithms used in these procedures are discussed.

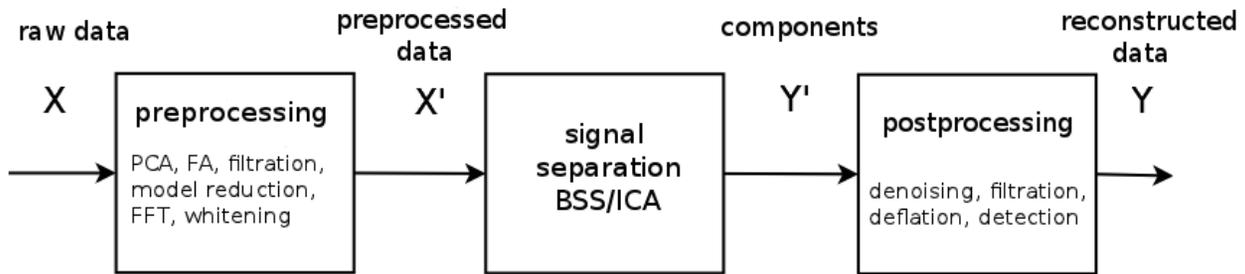


Fig. 2. Sequence of the basic procedures exploit in BSS algorithms for effective decomposition (PCA – Principal Component Analysis, FA – Factor Analysis, FFT – Fast Fourier Transform)

## 1.2. Preprocessing algorithms

### 1.2.1. Centering

Centering is the most basic and very often necessary form of preprocessing. In this method the value of the mean vector (6) is subtracted from the input data in order to make it a zero-mean variable. This suggests that  $\mathbf{s}$  is zero-mean as well.

$$\mathbf{m} = E\{\mathbf{x}\} \tag{6}$$

This process is made only to simplify the ICA algorithm. After establishing of the mixing matrix  $\mathbf{H}$ , with centered data, the estimation can be completed by adding the mean vector  $\mathbf{m}$  to the vector  $\mathbf{s}$ . It has to be noticed that the process of centering is performed only when the model can be described with the equation 7 [8], a simplified version of equation 5.

$$\mathbf{x} = \mathbf{H}\mathbf{s} \tag{7}$$

### 1.2.2. Whitening

Whitening is a linear transformation of the vector, described in 7, in order to obtain a new vector which components are uncorrelated and their variances equal unity. After this process the covariance matrix is equal to the identity matrix as shown in 8.

$$E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\} = \mathbf{I} \tag{8}$$

This kind of transformation is always possible. One of the widely known methods of whitening is the use of decomposition of the covariance matrix as in 9:

$$(\mathbf{R}_{xx}) = E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\} = \mathbf{V}\mathbf{D}\mathbf{V}^T \tag{9}$$

where:

- $\mathbf{V}$  – the orthogonal matrix of eigenvectors of the covariance matrix,
- $\mathbf{D} = \text{diag}(d_1, \dots, d_n)$  – the diagonal matrix of eigenvalues of the matrix  $\mathbf{R}(\mathbf{R}_{xx})$ .

Whitening can be performed as shown in the equation 10:

$$\tilde{\mathbf{x}} = \mathbf{V}\mathbf{D}^{(-1/2)}\mathbf{V}^T\mathbf{x} \tag{10}$$

where: matrix  $\mathbf{D}^{(-1/2)} = \text{diag}(d_1^{(-1/2)}, \dots, d_n^{(-1/2)})$ .

This transformation is useful because the new mixing matrix is orthogonal which reduces the number of parameters to be estimated [8].

### 1.2.3. Principal Component Analysis (PCA)

The PCA and Factor Analysis (FA) belong to the second-order methods, which means they use only the information contained in the covariance matrix of the data vector. When the data is not centered, the mean value is used in order to center it. Additionally, if the variable has a normal or Gaussian distribution, its distribution is absolutely determined by the second-order information. That is why it is not necessary to include any other information. What is more, these methods are not mathematically complex and very often require only classical matrix calculations [7]. PCA can be changed to a problem of estimating eigenvalues of covariance matrix from vector, which makes it an equivalent of Karhunen-Loeve transform. Computing eigenvalues and eigenvectors from covariance matrix is shown in 9. The Karhunen-Loeve transform describes linear transformation of data vector, as in 11 [2].

$$\mathbf{y}_s = \mathbf{V}_s^T\mathbf{x} \tag{11}$$

where:

- $\mathbf{y}_s = [y_1(k), \dots, y_n(k)]^T$  output vector also called the vector of principal components,
- $\mathbf{V}_s = [\mathbf{v}_1, \dots, \mathbf{v}_n]^T \in \mathbf{R}^{(m \times n)}$  the set of signal subspace eigenvectors consisting of orthonormal vectors.
- Vectors  $\mathbf{v}_i$  ( $i=1, 2, \dots, n$ ) are the eigenvectors of the covariance matrix.

According to the above considerations, one can state that the issue that needs to be solved is a standard problem of eigenvalues and can be formulated like in the equation 12.

$$\mathbf{R}_{xx}\mathbf{v}_i = \lambda_i\mathbf{v}_i \tag{12}$$

where:

- $v_i$  eigenvectors,
- $\lambda_i$  eigenvalues corresponding to the eigenvectors,
- $\mathbf{R}_{xx} = E\{\mathbf{x}\mathbf{x}^T\}$  the covariance matrix.

Principal Component Analysis allows for a division of observed data:  $x(k) = x_s(k) + z(k)$  into two subspaces: *signal subspace* corresponding with the main factors, which are connected with the highest eigenvalues:  $\lambda_1, \lambda_2, \dots, \lambda_n$  along with corresponding eigenvectors and noise subspace along with eigenvalues:  $\lambda_{(n+1)}, \dots, \lambda_m$ . Subspace spanned by the first  $n$  eigenvectors can be considered as an approximation of subspace without interference. One question can arise from this approach, that is how to establish a borderline dividing the two subspaces especially when the interferences are high. Fortunately, algorithms which solve this problem have been created and can be found in [2].

### 1.2.4. Factor Analysis (FA)

The FA method stays in close relationship with the Principal Component Analysis. An auto-generating model is assumed as shown in 13, hence the existence of interference is allowed.

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{z} \tag{13}$$

All variables in and are assumed to be Gaussian. Additionally, assumption has been made that dimension of  $\mathbf{s}$  is smaller than dimension of  $\mathbf{x}$ . In general the factor analysis is a method of reducing of the dimension of the data.

Two principal methods of computing the factor analysis exist [7]:

1. The principal factors method which is basically a modification of the Principal Component Analysis. The difference between those two methods is that the principal factors method takes into consideration the influence of interference. It can be assumed that the covariance matrix of the noise is known and can be determined in the following manner (eq. 14).

$$\mathbf{N} = E\{\mathbf{z}\mathbf{z}^T\} \tag{14}$$

Subsequently the PCA algorithm is performed on the modified covariance matrix  $\mathbf{L}$ , so equation 15 is employed.

$$\mathbf{L} = \mathbf{R}_{xx} - \mathbf{N} \tag{15}$$

Therefore the vector  $\mathbf{s}$  is the vector of the principal components of  $\mathbf{x}$  with noise removed.

2. The method based on maximum likelihood estimation. Also this method can be reduced in order to determine the principal components of a modified covariance matrix.

## 1.3. Independent Component Analysis (ICA)

### 1.3.1. Basic concepts and definitions

Firstly, it is necessary to explain elementary notion of this method – the independence. Supposing given random variables  $y_1, y_2, \dots, y_m$  total probability density function is equal to  $f(y_1, y_2, \dots, y_m)$ . The variables  $y_i$  are mutually independent, if the total probability density function can be decomposed into factors according to the formula 16. This independence is called statistical, unlike some others, for instance linear independence.

$$f(y_1, \dots, y_m) = f_1(y_1) f_2(y_2) \dots f_m(y_m) \tag{16}$$

Two concepts should be distinguished from each other: variables' independence and uncorrelation. If the variables  $y_i$  and  $y_j$  are uncorrelated, they satisfied the equation 17.

$$E\{y_i y_j\} - E\{y_i\} E\{y_j\} = 0, \quad \text{for } i \neq j \tag{17}$$

While independence is generally a stronger requirement. If the variables  $y_i$  and  $y_j$  are independent, they fulfill the equation 18 for any function  $g_1$  and  $g_2$

$$E\{g_1(y_i) g_2(y_j)\} - E\{g_1(y_i)\} E\{g_2(y_j)\} = 0 \tag{18}$$

There are several definitions of linear Independent Component Analysis [7], however, many ICA algorithms are based on simplified definition which does not take into account the noise vector:

Independent Component Analysis of random  $m$ -dimensional data vector consists of determining the data model shown in 19:

$$\mathbf{x} = \mathbf{H}\mathbf{s} \tag{19}$$

where the unknown components  $s_i$  of the vector  $\mathbf{s} = (s_1, \dots, s_n)^T$  are accepted as independent and the matrix  $\mathbf{H}$  is a constant mixing matrix of dimension  $m \times n$ .

To ensure traceability of the ICA method without interference (defined as above), following requirements are imposed:

1. All the independent components  $s_i$ , besides one acceptable exception, have to be non-Gaussian.
2. The number of observed linear mixtures  $m$  have to be equal or greater than the number of independent components  $n$  ( $m \geq n$ ).
3. The matrix  $\mathbf{H}$  have to be a column-full rank.

If  $\mathbf{s}$  and  $\mathbf{H}$  are represented as stochastic processes instead of simply random variables, additional assumptions are compulsory. At least one must assume that the stochastic processes are stationary in an exact meaning of the word.

### 1.3.2. Objective functions of Independent Component Analysis

Approximation of the data model of ICA usually consists of an objective function formulation and then its minimization (or maximization) by usage of optimization algorithms. In the following section various objective functions are presented. They are divided mainly according to their statistical properties [8].

**Likelihood.** Allows to calculate all of the independent components from the entire model at once. Firstly, a likelihood of the noise-free ICA model is formulated and then the model is estimated by a maximum likelihood method. Therefore one should consider log-likelihood, defined as in 20:

$$L = \sum_{(t=1)}^T \sum_{(i=1)}^m \log f_i(\mathbf{w}_i^T \mathbf{x}(t)) + T \ln |\det \mathbf{W}| \quad (20)$$

where:

- $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_m)^T$  – pseudo inverse of a matrix  $\mathbf{H}$
- $f_i$  – probability density function  $s_i$
- $\mathbf{x}(t)$ ,  $t=1, \dots, T$  – estimations of  $\mathbf{x}$ .

#### Mutual information and Kullback-Leibler divergence.

Allows to calculate all of the independent components from the entire model at once. Mutual information is a non-negative dependence measure between random variables. It can be determined using the differential entropy which involves estimation of the probability density function.

$$I(y_1, \dots, y_m) = \sum_i \Phi(y_i) - \Phi(\mathbf{y}) \quad (21)$$

where:  $\Phi$  denotes differential entropy, defined as in 22.

$$\Phi(y) = - \int f(y) \log f(y) dy \quad (22)$$

Mutual information is equal to zero if and only if the variables are statistically independent. Hence one has to find a transformation that minimizes mutual information between components of the  $s_i$ .

Kullback-Leibler divergence is defined for two probability densities  $f_1$  and  $f_2$  as in equation 23:

$$\delta(f_1, f_2) = \int f_1(\mathbf{y}) \log \frac{f_1(\mathbf{y})}{f_2(\mathbf{y})} d\mathbf{y} \quad (23)$$

The divergence can be considered as a kind of distance between two probability densities. If  $y_i$  in 21 are independent, their joint probability density may be decomposed into factors shown in formula 16. For that reason, using divergence, it is possible to measure independence  $y_i$  between the real density  $f(\mathbf{y})$  and the factorized one  $\hat{f}(\mathbf{y}) = f_1(y_1) f_2(y_2) \dots f_m(y_m)$ . Actually, received value is equivalent to the mutual information  $y_i$ .

Negentropy. Is an example function which optimization allows to estimate one of the components at a time. It is sufficient to find a vector  $\mathbf{w}$  which linear combination  $\mathbf{w}^T \mathbf{x}$  corresponds to one independent component of the  $s_i$ . Negentropy can be also considered as a nongaussianity's measure.

Negentropy  $\mathbf{J}$  is defined as in 24.

$$J(\mathbf{y}) = \Phi(y_{\text{gaus}}) - \Phi(\mathbf{y}) \quad (24)$$

where  $y_{\text{gaus}}$  is a Gaussian random variable of the same covariance matrix as  $\mathbf{y}$ .

Mutual information can be obtained using negentropy as shown in the equation 25:

$$I(y_1, y_2, \dots, y_n) = \mathbf{J}(\mathbf{y}) - \sum_i J(y_i) + \frac{1}{2} \log \frac{\prod_{xx} R_{xx}^y}{\det \mathbf{R}_{xx}^y} \quad (25)$$

where:

- $\mathbf{R}_{(xx^y)}$  – covariance matrix of  $y$ ,
- $\mathbf{R}_{(xx^y)}^y$  – diagonal elements of covariance matrix.

When  $y_i$  are uncorrelated, the last component in the equation 25 is equal to zero.

Finding the maximum of negentropy's directions (directions where the elements of the sum  $J(y_i)$  are maximized) is a substitute for finding representation in which mutual information is minimized. Because negentropy computation is difficult, it is often approximated by higher-order cumulants.

### 1.3.3. Algorithms used in the present paper

After selecting an objective function, one should consider the method of its implementation. In other words, it is necessary to determine how to optimize the objective function. Algorithmic properties of ICA method, such as stability or memory requirements, depend on the choice of the optimizing algorithm. Acronyms of the algorithms used in the present article along with their basic features are presented in the table 1. They are discussed in detail in the following sections.

**AMUSE – Algorithm for Multiple Unknown Signals Extraction.** This algorithm exploits properties of the second order statistics and spatio-temporal decorrelation. It does not use directly or indirectly statistical independence.

AMUSE is based on PCA algorithm, which is performed twice: first, to whiten the data and then as an appropriate algorithm for a time delayed covariance matrix. Whitening of linear transformation of the standard covariance matrix is carried out as in formula 26.

$$\mathbf{x}_1(k) = \mathbf{Q}\mathbf{x}(k) \quad (26)$$

where:  $\mathbf{Q} = \mathbf{R}_{xx}^{(-1/2)} = (\mathbf{V}\mathbf{D}\mathbf{V}^T)^{(-1/2)} = \mathbf{V}(\mathbf{D})^{(-1/2)} \mathbf{V}^T$ .

In the second step, singular value decomposition (SVD) of the time delayed covariance matrix is performed, as it is shown in equation 27.

$$\mathbf{R}_{\mathbf{x}_1 \mathbf{x}_1} = \mathbf{E}\{\mathbf{x}_1(k) \mathbf{x}_1^T(k-1)\} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \quad (27)$$

where:

- $\mathbf{\Sigma}$  – diagonal matrix with decreasing singular values,
- $\mathbf{U}, \mathbf{V}$  – orthogonal matrices.

Finally, the unmixing matrix is calculated as in 28.

$$\mathbf{W} = \mathbf{U}^T \mathbf{Q} \quad (28)$$

Tab. 1. Summary of algorithms used in the present paper and their basic properties

Algorithm	Features	Equation	Literature
AMUSE	PCA algorithm performed twice whitening of the standard covariance matrix SVD of time delayed covariance matrix	11 26 27	[2]
SOBI	an orthogonalization covariance matrix calculation use of the procedure JAD	26 30 29	[4] [4]
SONS	SOBI algorithm modification exploiting nonstationarity of the source signals division of previously whitened inputs	26, 30, 29 10	[12] [12]
JADE	based on JAD procedure optimization of previously defined objective function	29 32	[12] [12]

The main advantage of this algorithm is ability of automatic components ordering according to decreasing values of singular values (due to SVD algorithm). What is more, the components obtained from the AMUSE are uniquely defined. Unfortunately AMUSE is relatively sensitive to additional noise.

**Robust SOBI – Second Order Blind Identification.** This algorithm takes advantage of properties of second order statistics. The purpose of SOBI is to obtain reliable information (estimation of source signals of mixing matrix) from the structure of the eigenvalues from a possibly large set of data matrices. Because usually one has only a finite number of interfered signal's samples, it is inevitable to estimate the average eigenstructure by simultaneously taking into account the widest possible set of data matrices [5].

The average eigen-structure may be obtained for example by the application of JAD procedure (Joint Approximate Diagonalization). Its objective is to find the orthogonal matrix which diagonalizes a set of matrices as it is presented below in the formula 29.

$$R_x(p_i) = U D_i U^T + \epsilon_i, \quad i=1,2,\dots,L \quad (29)$$

where:

- $R_x(p_i) \in R^{(n \times n)}$  – data matrices,
- $D_i$  – real diagonal matrices,
- $\epsilon_i$  – errors or noise matrix.

The algorithm can be described by the following steps [5]:

- I Perform orthogonalization,  $\bar{x}(k) = Qx(k)$  like in equation 26.
- II Calculate covariance matrices, as described in 30:

$$\hat{R}_{(xx)}(p_i) = (1/N) \sum_{(k=1)}^N \bar{x}(k) \bar{x}(k-p_i) = Q \hat{R}_x(p_i) Q^T \quad (30)$$

for a selected set of delays  $(p_1, p_2, \dots, p_L)$ .

III Perform the JAD procedure (as in 29) using available numerical algorithm.

IV Estimate the source signals and the mixing matrix, as shown in 31

$$\begin{aligned} \hat{s}(k) &= U^T Qx(k) \\ \hat{A} &= Q+U \end{aligned} \quad (31)$$

The main advantage of the above algorithm is its robustness to additional interferences, provided that the number of covariance matrices is sufficiently large (greater than 100) [4].

**SONS – Second-Order Nonstationary Source separation.**

This is a modification of SOBI algorithm, which exploits nonstationarity of the source signals. The main difference is that SONS divides previously whitened input data on non-overlapping on each other blocks, from which time delayed covariance matrices are calculated. Afterwards, the JAD procedure is carried out according to the formula 29 [2].

**JADE – Joint Approximate Diagonalization of Cumulant Matrices.** There is also a large group of JADE algorithms, including FJADE (Fixed JADE) or QJADE (Quadratic JADE), which are based on the JAD procedure. They perform optimization of the objective function, defined as in 32.

$$\min_{(U \in R^{(n \times n)})} \sum_{(i=1)} \text{off}(UC_i U^T) \quad (32)$$

where:

- $\text{off}(F) = \sum_{(i \neq j)} (F_{ij})^2$
- $C_i$  – matrices of higher-order cumulants [12].

**2. Testing the efficiency of the chosen separation algorithms used for splitting artificially generated and real life recorded acoustic signals**

In the present chapter the results of experimental research have been presented. The analysis has been conducted in order to test and examine the functionality of algorithms described in the previous chapter. All calculations have been performed in MAT-LAB environment using one of its toolboxes, ICALAB for Signal Processing. Information regarding all of the algorithms implemented in this toolbox can be found in [12].

**2.1. Test signals**

At the beginning the separation algorithms have been tested on artificially generated signals: two sines (the first one with the amplitude 2 [cm] and frequency 10 [rad/s], the second one with

the amplitude 4 [cm] and frequency 3 [rad/s] and one saw-tooth signal (the amplitude 5 [cm] and frequency 3 [Hz]). Their time waveforms have been shown in Fig. 3. First to mix have been both sines (Fig. 3), in this process the Hilbert mixing matrix have been used. Obtained results are shown in Fig. 4. Not all

algorithms managed to solve this simple task. Algorithm SOBI and most algorithms which use higher order statistics have not separated signals correctly (Fig. 5). Algorithms which have succeeded in unbraiding the signals are AMUSE and Fixed-Point ICA, the obtained results are shown in Fig. 6.

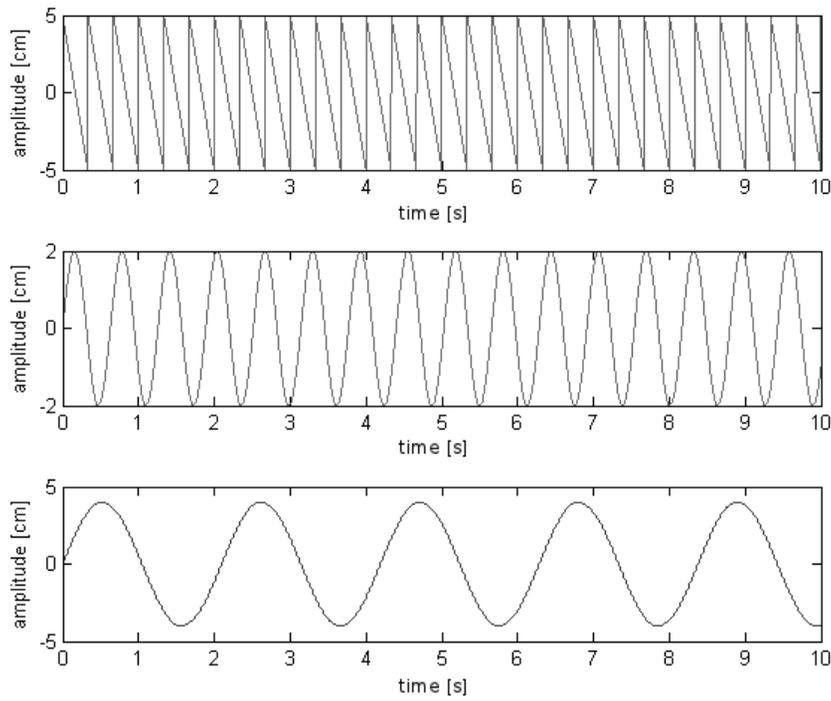


Fig. 3. Time waveforms of sinusoidal and saw-tooth signals

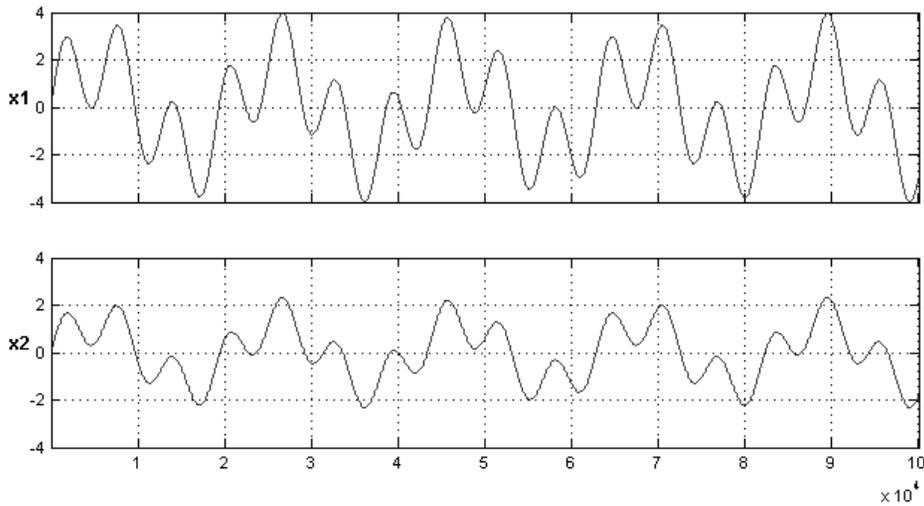


Fig. 4. Mixed sinusoidal signals

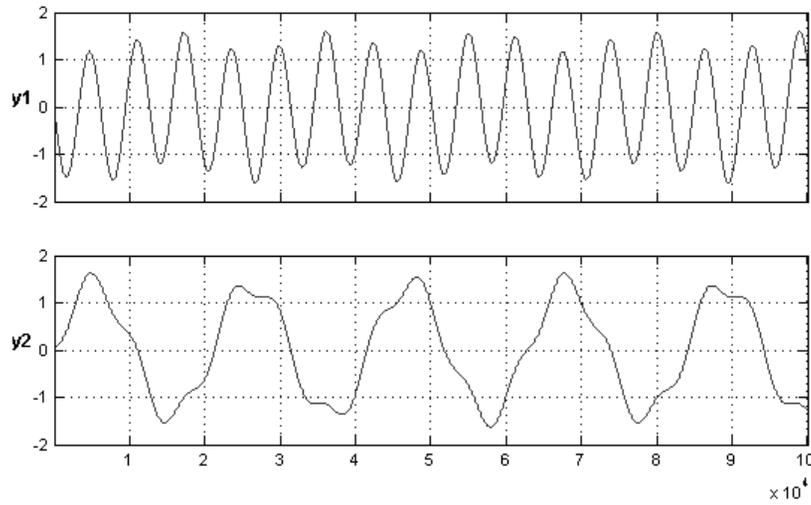


Fig. 5. Incorrectly estimated signals with SOBI algorithm

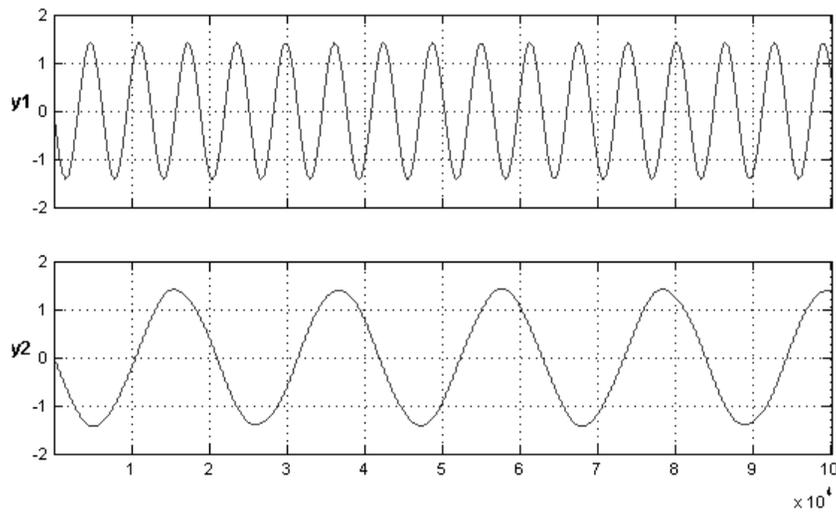


Fig. 6. Successfully deconvolved signals with AMUSE algorithm

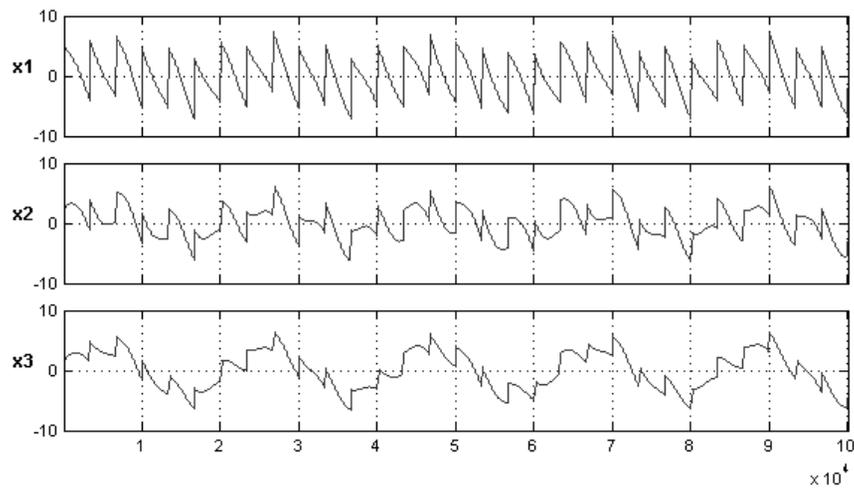


Fig. 7. Mixed sinusoidal and saw-tooth signals

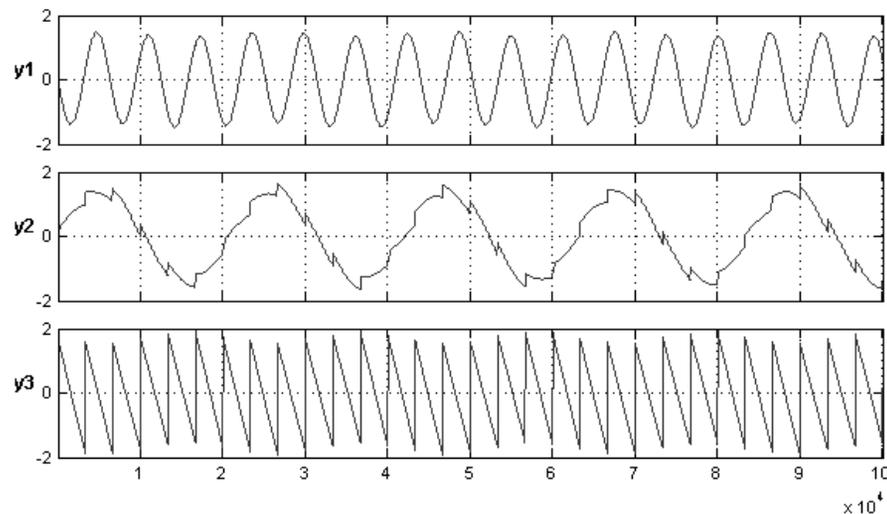


Fig. 8. Incorrectly estimated signals with the AMUSE algorithm

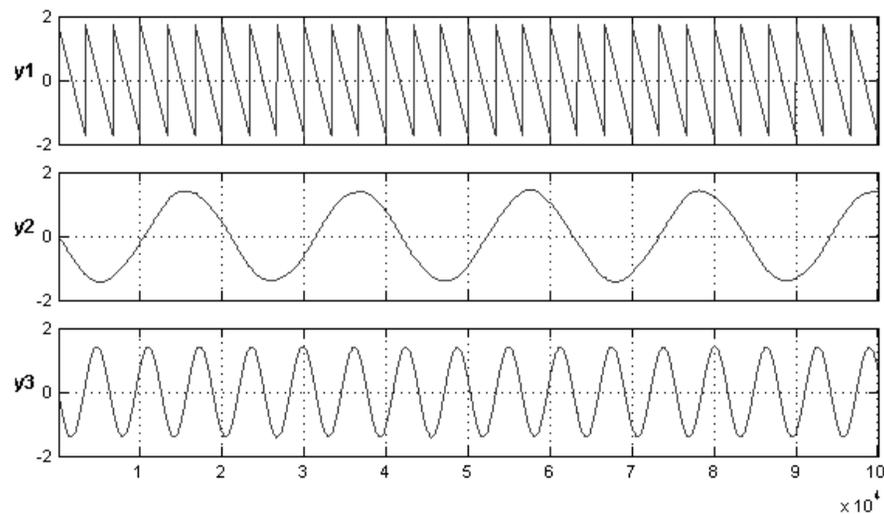


Fig. 9. Correctly estimated signals with FJADE algorithm

Subsequently all three previously generated signals (Fig. 3) have been mixed with Toeplitz mixing matrix [1]. Graphs presenting mixed signals are shown in Fig. 7. In this case algorithms AMUSE, SOBI and others did not manage to give correct results (8).

Only after using proper preprocessing (averaging), AMUSE algorithm has correctly separated the signals. Other algorithms such as FJADE, QJADE or ICA do not need preprocessing in order to successfully estimate the source signals. This phenomenon is shown in Fig. 9.

## 2.2. Human voices recordings

In the research described in the present paper, apart from artificially generated recordings, real life signals have been used

as well. The conducted research has not been extensive and exhaustive, hence presented results should be treated only as demonstrative. Three recordings of human voice were analyzed: female voice, low and high pitches of male voice. Time waveforms of these signals are shown in Fig. 10.

Because each of these voices is recorded separately, thus they had to be mixed first. In order to do that, Hankel mixing matrix has been used [11]. The signals obtained in this process are shown in Fig. 11.

Similarly to the previous research, not all algorithms allowed for correct estimation of source signals. AMUSE, FJADE, Fixed-Point ICA and other ICA algorithms distort the sound, add noise or not entirely separate the voices because of which remnants of other sounds can be heard. Even the use of preprocessing does not help. An example of incorrectly deconvoluted signals is shown in Fig. 12.

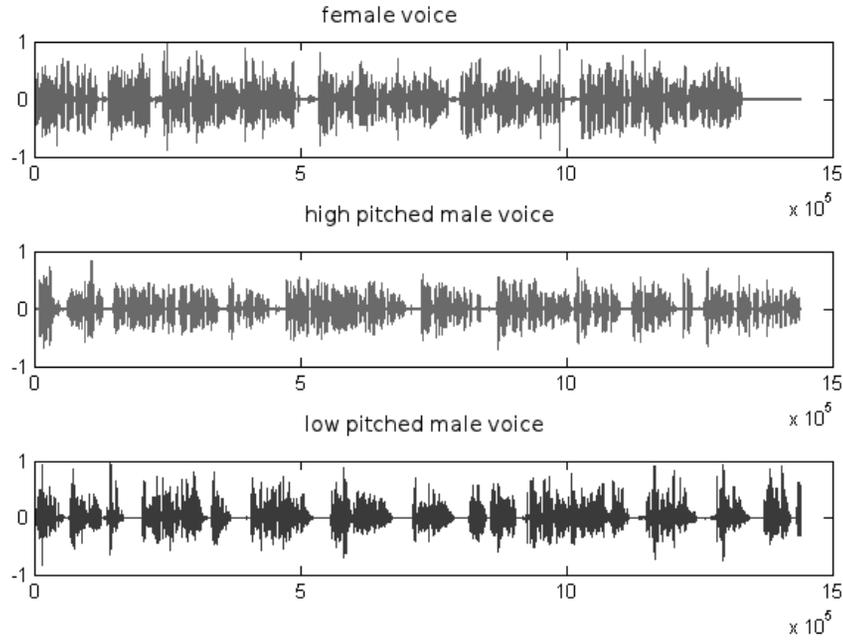


Fig. 10. Voice's waveforms

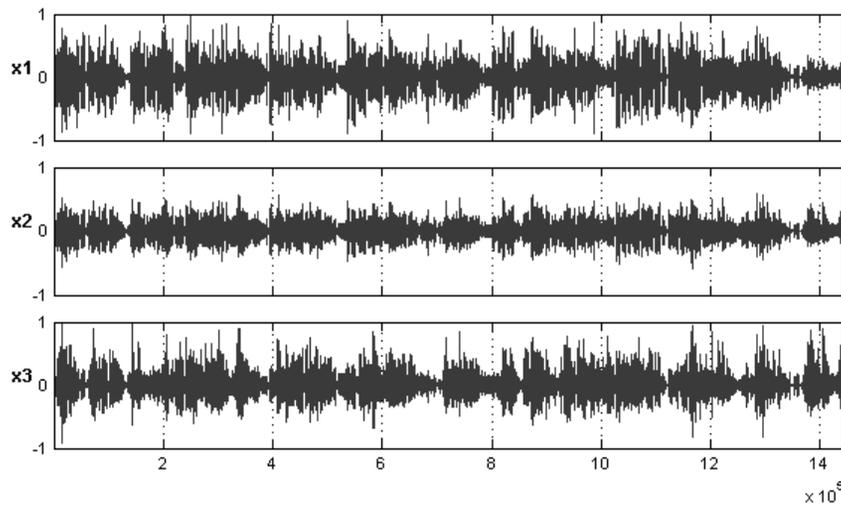


Fig. 11. Mixed signals of human voice

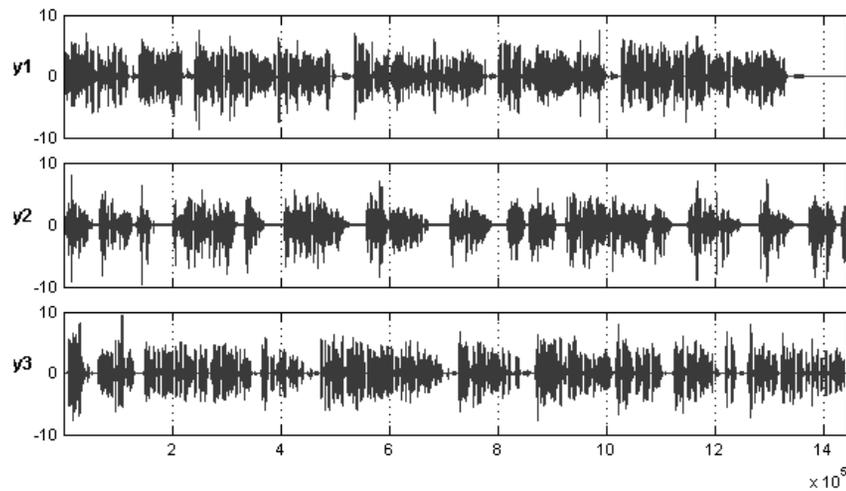


Fig. 12. Incorrectly separated signals (FJADE algorithm)

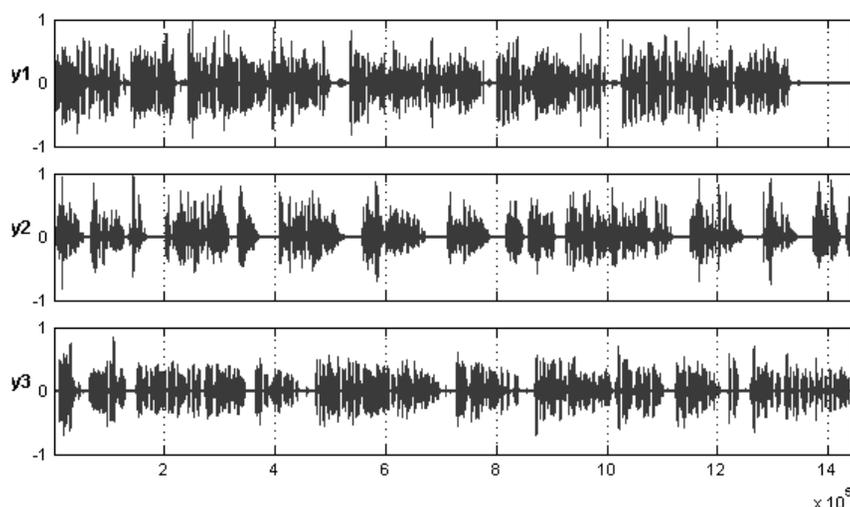


Fig. 13. Correctly estimated signals (SONS algorithm)

The best results have given SOBI and SONS algorithms, the signals have been unmixed almost perfectly. What is more, the use of preprocessing effects in disappearance of any interference in reconstructed signals. The graph presenting obtained signals is shown in Fig. 13.

### Conclusions of conducted experiments

As mentioned at the beginning of the present article, many algorithms which allow for separation of various signals exist. It is possible because some of them use information obtained from second order statistics, other from higher order statistics and some from temporary-spatial decorrelation. In the research conducted in the present paper these algorithms have been tested on artificially generated sinusoidal and saw-tooth signals as well as on recordings of human voice. It has been proved that, one universal algorithm which manages to successfully split any mixed signals does not exist. Algorithms such as AMUSE, FJADE or Fixed-Point ICA which deconvoluted sinusoidal and saw-tooth signals (see chapter 2.1) are not able to correctly split or even decreased the quality of the sound of human voices (see chapters 2.2). Algorithms SOBI and SONS on the other hand, properly estimate source signals of human voices without distorting the quality of the sound but are not able to handle mixed sinusoidal signals.

Some of the algorithms presented in this paper might be useful in many areas of human life, especially in medicine. In this discipline there is often the need to prepare a report of activities carried out by the doctor, for example during the autopsy. The doctor cannot simultaneously perform the procedure and write down the protocol, hence he records a description of the activities carried out. To avoid rewriting parts of the recordings after finishing the procedure, speech recognition systems can be employed. However, recorded additional noise has to be removed before, using automatic signals separation, in order to facilitate the work of the analyzer.

One aspect of typical ICA algorithms has to be mentioned: each obtained independent sound signal is a monophonic signal. Hence separated signals do not store information about direction, localization or spatial properties of each source of the sound. In order to comply with this issue various solutions are proposed. One of which suggests the use of ICA algorithm based on SIMO model (Single Input Multiple Output). More information on the topic can be found in [10].

### References

1. Al'pin Yu.A., Il'in S.N.: Infinite Extension of Toeplitz Matrices. *Journal of Mathematical Sciences* 2005, 127.
2. Amari S., Cichocki A.: Adaptive Blind Signal and Image Processing. Learning Algorithms and Applications. West Sussex: John Wiley & Sons, 2002.
3. Bell A.J., Lee T.-W.: Blind source separation of real world signals. *Neural Networks* 1997, 4: 9-12.
4. Belouchrani A., Abed-Meraim K., Cardoso J.F., Moulines E.: A blind source separation technique using second order statistics. *IEEE Trans. on Signal Processing* 1997, Vol. 45 (February).
5. Choi S., Cichocki A., Park H.M., Lee S.Y.: Blind Source Separation and Independent Component Analysis: A Review. *Neural Information Processings – Letters and Reviews* 2005, Vol. 6.
6. Der R.: Blind Signal Separation. Materials for Laboratory of Telecommunications & Signal Processing of the McGill University, Montreal, 2001.
7. Hyvarinen A.: Survey on Independent Component Analysis. *Neural Computing Surveys* 1999, Vol. 2.
8. Hyvarinen A., Oja E.: Independent Component Analysis: Algorithms and Applications. *Neural Networks* 2000, Vol. 13.
9. Koldovsky Z., Tichavsky P., Oja E.: Efficient Variant of Algorithm FastICA for Independent Component Analysis Attaining the Cramér-Rao Lower Bound, *IEEE Trans. on Neural Networks* 2006, Vol. 17 (September).

10. Nishikawa T., Saruwatari H., Shikano K., Takatani T.: Blind separation of binaural sound mixtures using SIMO-model-based independent component analysis. 2004 IEEE International Conference of Acoustics, Speech and Signal Processing, ICASSP, 2004.
11. Tyrtshnikov E.E.: Matrices, continued fractions, and fast algorithms. *Russ. J. Numer. Anal. Math. Modelling* 2010, Vol. 25.
12. Cichocki A., Amari S., Siwek K., Tanaka T., Phan A.H., Zdunek R.: ICALAB MATLAB Toolbox Ver. 3 for signal processing, <http://www.bsp.brain.riken.jp/ICALAB/ICALABSignalProc/>.

## IMPLEMENTATION AND SIMULATION OF THE SCALAR TIMING MODEL

MACIEJ KOMOSINSKI, ADAM KUPS

*Institute of Computing Science, Poznan University of Technology,  
Institute of Psychology Adam Mickiewicz University in Poznan*

**Abstract:** Perception of time is an important ability: timing plays a crucial role in accomplishing various real-life tasks. This work concerns judgments made about short temporal intervals. The main goal was to implement the well known Scalar Timing Model (STM) as an artificial neural network. The implementation of the STM resolves to mapping the architecture of the model into an artificial neural network. Another direction of experiments concerns the pacemaker and distribution of generated pulses in time. Manipulation of pulse distribution in time helps in explaining basic, empirical regularities concerning perception of time, and the relation between the stimulus and its sensation. The STM implementation described in this paper models the time-order error (TOE), a phenomenon often reported in experiments on living organisms. This research is a starting point for further wide-scale analysis of human time perception on psychological, psychophysical and biological levels.

**Keywords:**

### 1. Introduction

The main motivation of this research was to design and study an implementations of the Scalar Timing Model (which is derived from the Scalar Expectancy Theory – SET [4, 32, 31]) using a simulation (artificial life) environment. The idea behind this approach is that this model is relatively well established [33] by many works in psychology, and, therefore, it is very interesting to see whether it can endure limitations of artificial neural networks. If it can, then the Scalar Timing Model (STM) is one step further on the way to become an adequate scientific construct describing and explaining mechanisms of time perception in human brain and *a fortiori* in human mind. On the other hand, implementing any theoretical model in some environment gives insight into its validity according to the rules of the employed environment. Therefore such implementation is a good test whether the model is constructed well.

Another goal of this research was to simulate, at least partly, the time-order error (TOE) – the effect discovered long ago by Fechner and well described since then [12, 1, 17, 14, 34, 28, 29, 15, 16, 26, 23]. The simulation of the TOE within the designed network is the first step on the way to integrate more experimental data into this implementation.

### 2. The Scalar Timing Model

The general advantage of the STM, from the point of view of this work, is its popularity among psychologists. There are many works including experimental data concerning effects and details of human time perception, trying to link observations with this theoretical construct. Obviously, the basis of this popularity is not a matter of taste, but rather a history of development of psychology of time and the explanatory power of the SET (for discussion about theories, see [35, 9]). The fact that the STM is universal (in the sense that it was first applied in studies of animal perception of time [32]) is significant as well. The STM is relatively well developed, with many verification trials performed so far, which is important because of at least two reasons:

- huge packages of data (collected during experiments) are really helpful when it comes to checking whether some particular implementation is adequate,
- scientific value of this model justifies the need of the multi-disciplinary work on developing it.

Last but not least, the structure and cognitive psychology conceptualization frame of the STM makes it is relatively easy to implement in an Artificial Life environment.

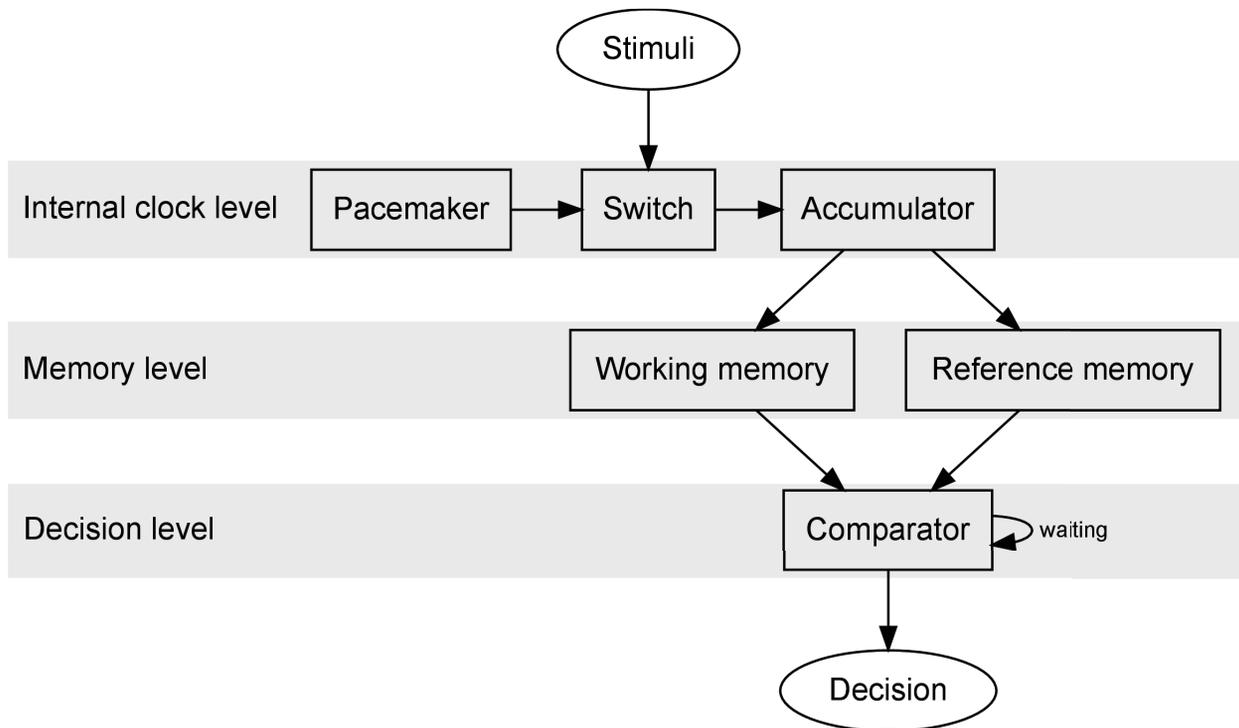


Fig. 1. The general architecture of the Scalar Timing Model (STM) [8].

The Scalar Timing Model illustrated on Fig. 1 is composed of three main parts [32]. First, the so-called internal clock level, consists of the pacemaker generating pulses and sending them to the accumulator which stores these pulses. The views on the importance of the pacemaker differ in various works [4, 27, 32]. Then, there is a switch which is considered to control the number of pulses stored in the accumulator as it is triggered by the stimulus. According to the assumptions of this model, the amount of accumulated pulses is a representation of stimulus duration. Once the representation is acquired, it is in the next step sent to the second level of the model. This one is called the memory level and consists of two components – the working memory and the reference memory. Actually, time representations from the accumulator could be sent to both of these components, depending on the timing task. In short, the goal of the reference memory is to provide useful information to compare with the content of the working memory<sup>1</sup> (which is in this case the time representation from the accumulator). The comparison process takes place on the third level of the model – the comparator. The results of the work of the comparator are the basis for further actions of an animal.

More or less precise rules that describe the way these components work have been introduced [4, 3, 32]. These rules refer to the distribution form of some constants related to the threshold of the decision rule within the comparator component, the multiplication of information before it reaches the working memory, and the switch opening/closing latency. The importance of these rules comes from the fact that they determine scalar and non-scalar sources of variability of judgments – judgments

that concern stimuli fed into the model. The constants are not directly present in the implementation described here, since the main goal was to map the architectural scheme of the STM into the artificial neural network (ANN). However, the mechanisms of the ANN allow for emergence of regularities that correspond to some of the constants. These constants can also be manually added to the design of the ANN.

The STM has two fundamental formal properties [32]:

- The mean representation of the stimulus duration should approximate the real stimulus duration (the *mean accuracy*).
- The coefficient of variation should remain constant independently from the stimulus duration. This property is similar to the Weber's law (the *scalar property*).

There are cases when these properties are not satisfied [30, 24, 19]. The second property can be explained in terms of the working/reference memory mechanism and decision thresholds, as well as by other views that ascribe this property to the pacemaker mechanism [6, 7, 27]. However, the pacemaker mechanism was shown to be problematic in explaining the second property because of the primary assumption regarding pulse distribution (*ibid.*). The next section summarizes properties of variance of the number of pulses accumulated in a given time interval, depending on the characteristics of the pacemaker. It also presents reasons that make it difficult to determine whether the second property holds or not for experimental data.

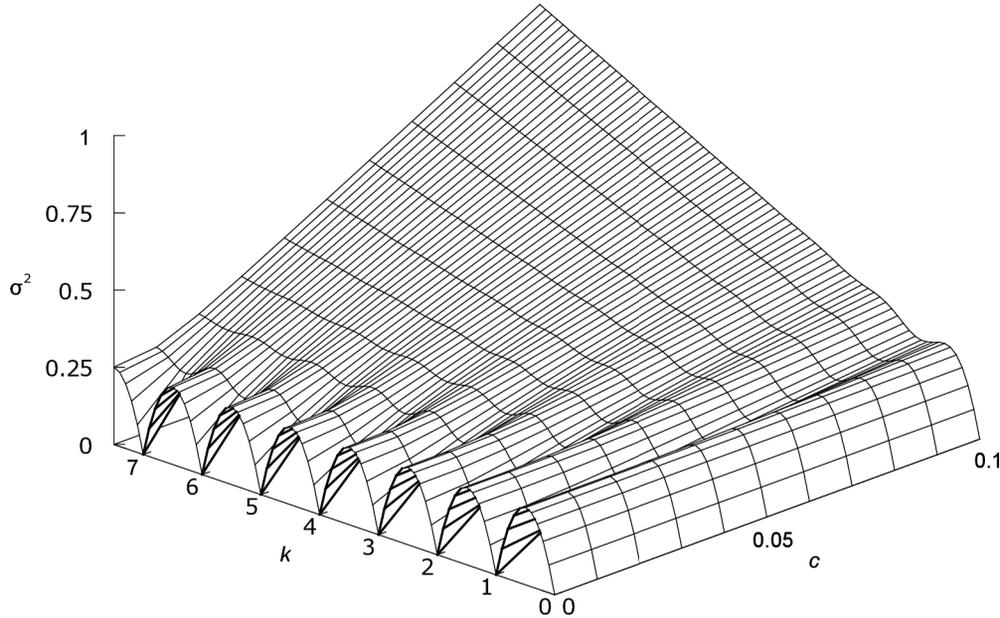
### 3. Pacemaker as the source of variability

Previous theoretical work and numerical experiments [18] confirm that for the three considered pacemaker interpulse distributions

<sup>1</sup> See [10] for more information on problems with temporal memory.

(exponential, normal, uniform), the variance  $\sigma^2$  of the number of pulses  $K$  generated during stimulus of length  $t$  depends linearly or approximately linearly on  $t$ . This relation for periodic generators is illustrated in Fig. 2. For all these distributions, also the mean

number of pulses,  $k$ , depends linearly on  $t$ :  $k=t/d$ , where  $d$  is the mean interpulse interval length from the generator. Therefore, the generator along with the pulse counter provide the mean accuracy property, but not the scalar property.



**Fig. 2.** Variance  $\sigma^2$  of the number of pulses  $K$  and its dependence on the mean number of pulses  $k$  and on the characteristics of the pulse generator  $c$ . The interpulse time  $D$  follows the normal distribution with mean  $d$ . For  $c>0$ , nonlinear components of variance decrease with increasing stimulus length  $t$ .

For a Poisson process where the pulses occur continuously and independently at a constant average rate  $\lambda$ , the times between events follow the exponential distribution,  $\text{Exp}(\lambda)$ , and  $\lambda=1/d$ . The mean number of pulses occurring in time  $t$  is  $k=t/d=\lambda t$ . The variance of the number of the pulses in time  $t$ , the variance-to-mean ratio, and the coefficient of variation are

$$\begin{aligned} \sigma^2 &= k \\ \frac{\sigma^2}{k} &= 1 \\ \frac{\sigma}{k} &= \frac{1}{\sqrt{k}} \end{aligned}$$

For normally and uniformly distributed interpulse time with mean  $d$  and variance  $\sigma_d^2$ , the variance of the number of pulses,  $\sigma^2$ , oscillates around the base value (Fig. 2). The asymptotic behavior of  $\sigma^2$  is given by

$$\begin{aligned} \sigma^2 &\approx k \cdot c + \frac{c^2}{2} + \frac{1}{6} \\ \frac{\sigma^2}{k} &\approx c + \frac{3c^2 + 1}{6k} \\ \frac{\sigma}{k} &\approx \frac{\sqrt{k \cdot c + \frac{c^2}{2} + \frac{1}{6}}}{k} \end{aligned}$$

where the  $c=(\sigma_d/d)^2$  coefficient characterizes the irregularity of the periodic oscillator and is considered constant. As  $k$  grows to infinity (which is equivalent to  $t$  approaching infinity and an oscillator with finite  $d$ ),

$$\begin{aligned} \lim_{k \rightarrow \infty} \sigma^2 &= \infty \\ \lim_{k \rightarrow \infty} \frac{\sigma^2}{k} &= c \\ \lim_{k \rightarrow \infty} \frac{\sigma}{k} &= 0 \end{aligned}$$

Depending on the characteristics of the oscillator,  $c$ , the specific requirements of an experiment, and the available precision of measurements, the time  $t$  needed for subjective stabilization of the coefficient of variation and the variance-to-mean ratio may vary, and may be considered short (i.e., not infinite as the equations above show). The performed analyses and experiments demonstrate that, given imprecise measurements and a limited range of length of stimuli, it is possible to erroneously consider the coefficient of variation constant. The risk increases for long stimulus lengths  $t$ : while  $k$  grows, the  $\sigma/k$  ratio stabilizes gradually. When one is looking for the constant coefficient of variation, the best-fit regression coefficients (note that while the Weber's law

has one parameter, its generalized form has two) may seemingly indicate a linear relationship between  $\sigma$  and  $k$ .

For the coefficient of variation to be constant (a property not fully confirmed by experiments [9, 30, 25, 24, 21, 2]), the characteristics of the pacemaker should vary [32, 27], the number of pulses measured in experiments should be multiplied by a random variable [7, 6], or other steps should be taken that concern the other components of the STM [7, 6, 22].

#### 4. Implementation of the STM

The implementation proposed in this work is an improvement over the previous architecture [19], where we aimed at using only those neuron types that were provided in the standard distribution of the Framsticks toolkit [20]. The implementation described here uses a specialized neuron, Gate, that allowed to considerably simplify the network topology. This artificial neural network, as shown in Fig. 4, additionally models the time-order error (TOE).

The simulation of the network is performed synchronously (step-by-step), so time is discrete. Signal magnitudes are represented as floating point values; neurons are not spiking, and information is not encoded as frequency. The types of neurons (i.e., available building blocks) that were used for this implementation of the STM are shown in Fig. 3.

The SeeLight receptor provides information about the magnitude of the stimulus. The Pulse neuron is the generator (it can generate a pulse train according to geometric, normal or uniform interpulse interval distributions). The N neuron is a popular AI neuron with a logistic activation function. The Thr neuron outputs either of the two values ("low" or "high") depending on whether the weighted sum of inputs exceeds its threshold value. The Delay neuron delays output to input for a given number of simulation steps. The Sum neuron, in each simulation step, adds a weighted sum of inputs to its internal state, and outputs the state.

The Gate neuron is a filtering unit with the first input acting as the control input. The sign of the control input is used as a selector: a weighted sum of the remaining inputs is propagated to the output, but the sum only concerns those inputs, whose sign is the same as the sign of the control input (positive, zero, or negative). This neuron is inspired by existing biological neurotransmitters that affect neurons locally, modulating the work of a group of neurons. The Gate neuron models this modulation in a simple way – by inhibiting one circuit and activating the other one.

The neural network structure is divided into 11 basic modules, and one group of neurons (module "A") that supports the TOE (*time-order error*) – time perception regularity described in Sec. 5. Modules described below can be divided into two groups: the components that exist in the theoretical approach (that is, in the genuine STM), and the components that are technical solutions to engineering problems absent from the theoretical model.

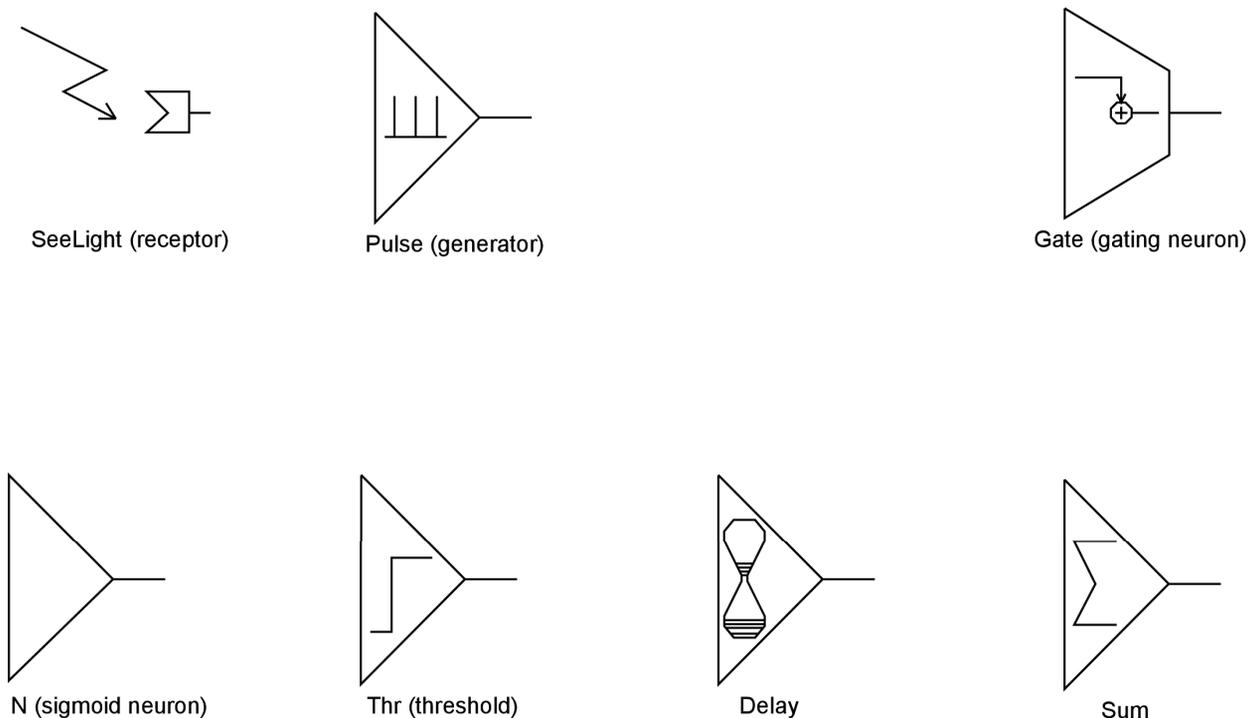
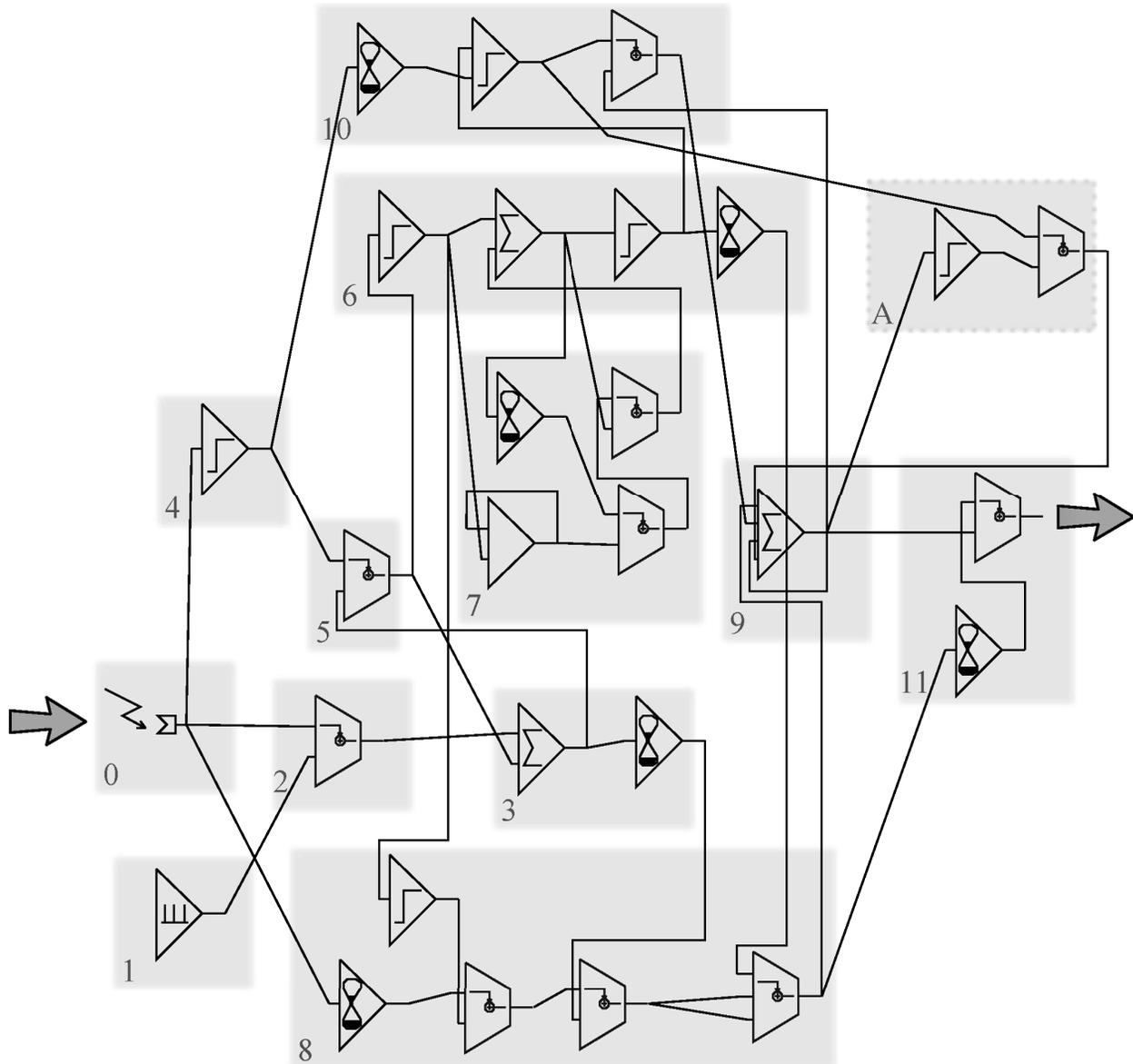


Fig. 3. Neuron types used in the implementation of the STM (Fig. 4). See text for details.



**Fig. 4.** The artificial neural network that implements the Scalar Timing Model. The numbers in this diagram correspond to numbers of paragraphs in Section 4. "0" is the stimulus receptor.

#### 1. Pacemaker

The pacemaker is a part of the STM, responsible for emitting pulses according to some distribution of the interpulse interval. Empirical data may indicate which distribution should be used in the simulation.

#### 2. Switch

The switch is another component of the STM. It lets pulses pass from the pacemaker to the accumulator while the stimulus lasts, otherwise the switch is closed. The switch is represented by the Gate neuron.

#### 3. Accumulator

The accumulator is a STM component that stores pulses coming from the pacemaker, and sends information about the collected amount of pulses to the working memory. It is implemented as a summing neuron (Sum) and one Delay neuron that helps synchronize this process with other events in the network. The state of the Sum neuron represents the number of

pulses collected by the accumulator. The Sum neuron represents a memory in a biological system; this is a simple solution and a starting point for further development.

#### 4. Scaling

Scaling, not directly present in the STM, is needed to notify regions of the network that the stimulus lasts or not. It is implemented as a simple binary threshold neuron.

#### 5. Accumulator reset loop

The scaling mentioned above controls the input of the accumulator reset loop. The loop itself is represented as a Gate neuron that is plugged into the accumulator with a negative weight from the range of  $-1..0$ . Such a connection allows to erase the accumulator contents just after exposition of any stimulus. This idea to reset a Sum neuron is used many times in the network.

#### 6. Process-information storage circuit

This module, as well as the previous one, is not considered in the STM. The task of this module is to store information about

the resetting process that takes place after the first stimulus of a pair has appeared. This information is further used to change the mode of transmission of the second stimulus, which is needed for the network to recognize it as the *second* one. It is also used to activate the working memory reset loop in the appropriate moment. This circuit is made of one Sum neuron (storage), two binary Threshold neurons and one Delay neuron.

#### 7. Process-information storage reset loop

This module is resetting the process-information storage after exposition of the second stimulus. The mechanism makes the network return to its initial state so that it is ready to process another pair of stimuli. A sigmoid neuron is used here and a recurrent connection, which allows to maintain information that initializes this loop for long enough.

#### 8. Accumulator-working memory mediator

This component is not specified in the STM. It sets the mode of information transmission from the accumulator to the working memory. The information about the exact, stored amount of pulses in the accumulator after exposition of the first stimulus is sent unchanged to the working memory. The value (length) of the second stimulus is sent in a different mode – that is, it is sent with the opposite (negative) sign in order to be subtracted from the already-stored value in the working memory. The comparison process might be more complicated [4], however, simple subtraction represents the idea. This module consists of a Delay neuron, a binary Threshold neuron, and several Gate neurons. The Gate neurons are controlled by the stimulus receptor and by the process information storage circuit. Two of the Gate neurons mediate between the accumulator and the working memory.

#### 9. Working memory

This module is an essential part of the STM – it stores and processes information about the duration of the stimuli. In the theoretical STM, comparing stimuli is ascribed to the comparator, while in our implementation, the Sum neuron that represents the working memory performs this function. The working memory has additional properties described in Sec. 5 that allow to demonstrate the time-order error effect.

#### 10. Working memory reset loop

This component is not directly present in the STM. Its goal is to reset the contents of the working memory just after exposition of the second stimulus. The module has inputs coming from the scaling module and the process-information storage circuit. Signals from those are processed in the controlling threshold neuron, which activates the loop whenever there is no stimulus and no information is detected in the process information storage circuit. The Gate neuron is responsible for performing the reset.

#### 11. Comparator

The comparator is a part of the STM. It plays a slightly different role in this implementation – see the description of the working memory. The comparator receives a signal that informs which stimulus had longer duration and what was the exact magnitude of the difference. The module consists of a Gate neuron and a synchronizing Delay neuron, which allow to output, in the appropriate moment, the information about the final result of the comparison process performed by the network.

The  $\Delta A$  module is responsible for the TOE effect described in the next section. The behavior of the module itself is explained in more detail in Sec. 5.1.

## 5. Time-order error

The time-order error (TOE) is a widely described phenomenon [13, 12, 15, 1, 11] observed in several modalities, including time perception. There are many definitions of the TOE, but in general, the TOE means that when two stimuli separated by some time interval are compared, the presentation order of these stimuli influences judgments made about them. The TOE may be either positive or negative. A positive TOE means that the first stimulus is overestimated relatively to the second one. A negative TOE is the opposite situation – the first stimulus is underestimated relatively to the second one [13, 1].

There have been many time perception studies concerning the TOE. Different methods, ranges and types of stimuli were used in research regarding this kind of phenomenon [15, 1, 11], and several factors were found to be responsible for its direction and magnitude. Within this constellation of factors, one factor is frequently reported – the duration of stimuli. When the compared pair consists of short stimuli, the TOE is positive, and it starts to decrease while the duration of stimuli increases [12, 13].

Many explanations have been proposed for the TOE, none of which could cover its every aspect [12]. Some of these explanations are based on the Gestalt psychology, while others have psychophysical origins (for example Hellström's *sensation-weighting model* [13]). There are some neurobiological approaches [11] to explain the TOE as well. One of the early explanations suggests that the negative TOE is caused by a retroactive phenomenon of fading of the memory trace of the first stimulus [12].

### 5.1. Implementation

The TOE is apparent for experiments concerning a pair of stimuli of different magnitude, but also for those concerning a pair of identical stimuli. To check whether both types of pairs implicate the TOE, different psychophysical and probabilistic measures have been applied [12]. For simplicity, only pairs of identical stimuli are shown here to illustrate that two simple assumptions might explain both directions (positive vs. negative) of the TOE in the two cases (ISI and magnitude).

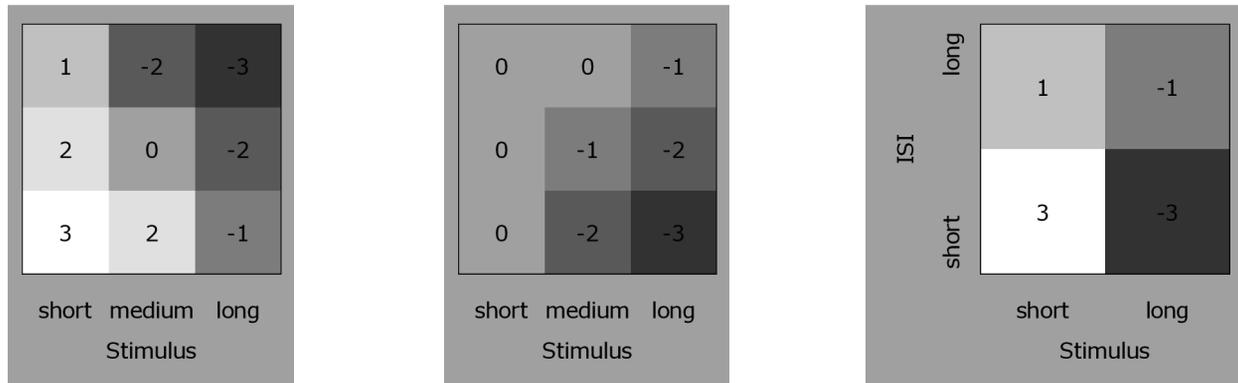
There are two assumptions made here in order to achieve the TOE during the comparison process in the network:

1. There is a continuous decrease of the value stored in the working memory, proportional to this value.
2. The working memory is not empty when the comparison process starts: it has some minimal value already stored.

The first assumption is related to the mentioned concept of a memory trace fading. The idea underlying this assumption is that the working memory is not perfect and some information is always lost during the timing process. The proportional decrease might be justified by the situation where the higher value is needed to be stored, the more biological neurons need to be activated and the loss caused by each neuron is somehow additive.

The second assumption – the positive initial value in memory – may be advantageous for the timing process and for an organism: it can help notice very short stimuli, and it can help balance the negative TOE within some critical range of durations.

The two assumptions taken together manifest themselves as a positive TOE observed for short stimuli durations, and



**Fig. 5.** An intuitive illustration of various TOE dependencies. Shades and numbers represent direction (dark  $\square$  negative, bright  $\square$  positive) and relative magnitude. Left panel: TOE according to [15]. Middle panel: TOE implemented in this work. Right panel: A variant of TOE that could be easily achieved in the current implementation (similar results were reported by [17]).

a negative TOE for long stimuli durations. The time scale in the experiments performed here is arbitrary and should be adjusted to empirical data, which may require additional (e.g., neurobiological) information.

The assumptions presented above implicate another effect: for a short ISI (interstimulus interval), the TOE is positive and becomes consequently negative with an increasing duration of the ISI, as shown in Fig. 5, middle panel. This is not, however, entirely consistent with some results reported in the literature. While such situation takes place for some modalities [12, 13], in the domain of timing the conclusions are not so clear. Jamieson and Petrusic [15, 17] demonstrated that the magnitude of TOE is rather decreasing (both positive and negative) with an increasing ISI, as shown in Fig. 5, left panel. Schab and Crowder [26], despite the fact that their research did not capture any influence of ISI on the TOE, have some doubts about the direction of change in the TOE as a function of ISI. These doubts come from the comparison of results reported by Jamieson and Petrusic and the work of Needham [23]. Schab and Crowder point up that the results by Needham are in opposition – the magnitude of the TOE is increasing as a function of ISI. Needham's results concern, however, the intensity of the auditory stimulus, not its duration.

These considerations show that the TOE implementation presented here should be considered a demonstration of this effect rather than a complete model. The implementation displays the potential of the network and confirms that the idea of the fading memory trace, together with the two assumptions (proportional decrease of value in the memory and the positive minimum value in the memory), may explain the influence of various stimuli lengths on the TOE.

A simple variant of the TOE, shown in the right panel of Fig. 5, is based on the accumulator and its reset loop, which may be responsible for the decreasing magnitude of TOE as a function of increasing ISI. When the ISI is relatively short, the accumulator may not be able to entirely erase the remains of the first stimulus on time, so the second stimulus will be added to these remains. This will result in overestimation of the second stimulus over the first one. Obviously, the increase of ISI will decrease the magnitude of the negative TOE, but the positive TOE cannot

be achieved in this approach. Consequently, if there is no TOE, increasing the ISI cannot change its magnitude.

The two assumptions mentioned earlier are sufficient for a simple TOE effect to emerge (Fig. 5, middle), and they were straightforward to implement in an artificial neural network. A continuous, proportional decrease of a memorized value is represented by the negative feedback connection of the Sum neuron (with a connection weight smaller than 1). Note that this is a representation of some process, and not a substantial explanation of the process. The second assumption is actualized by two neurons (a binary Threshold and a Gate) – that is, the module A in Fig. 4. The threshold neuron is used to check whether the working memory should be charged to some extent, and if that is true, it sends information through the Gate to the memory. The Gate is controlled by the working memory reset loop and lets the working memory be charged whenever no comparison process is running. This interplay between charging the working memory, the reset loop and the negative feedback loop results in the memory value (module 9) stabilizing, provided that the connection weights are properly adjusted.

## 5.2. Illustrative experiments

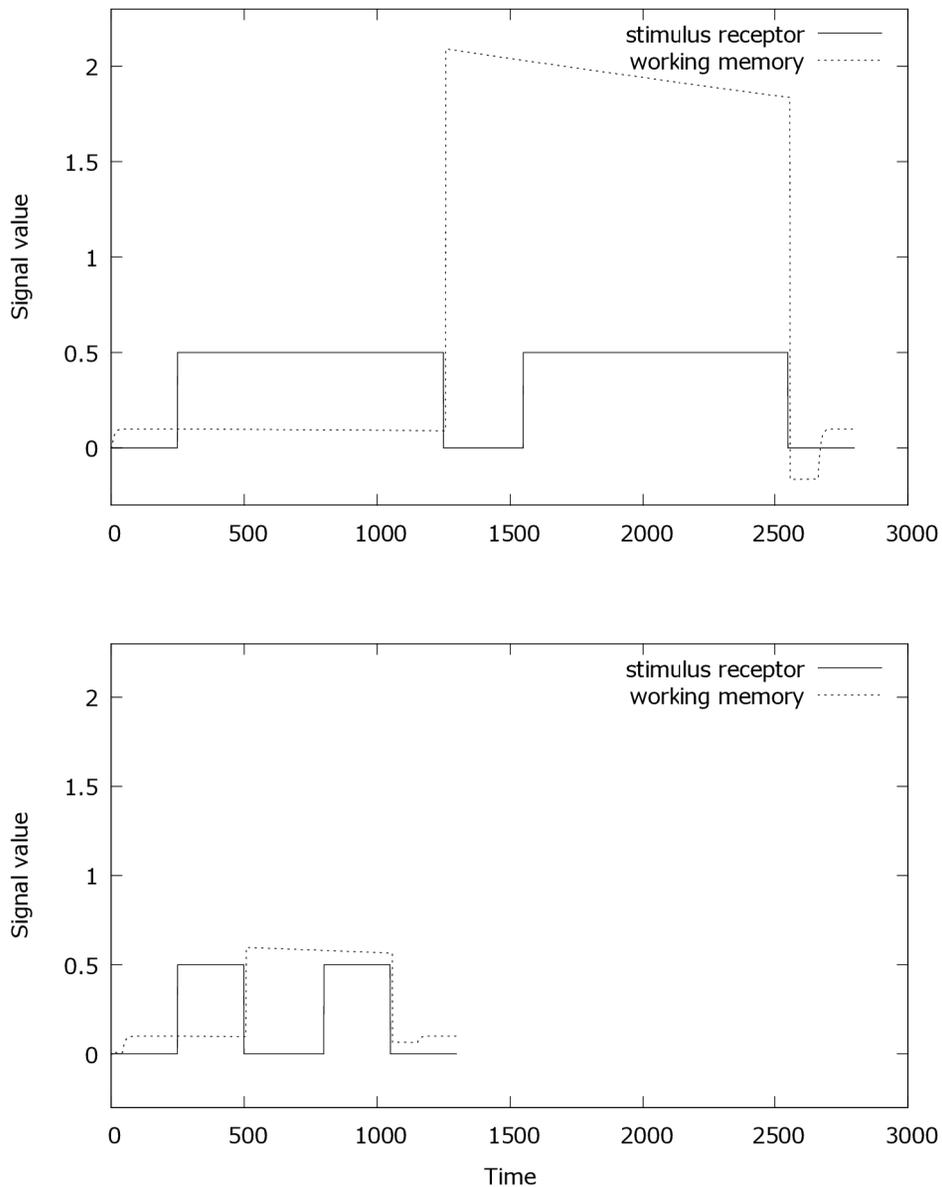
This section demonstrates two specific cases of TOE, and the way the implementation of the timing mechanism reflects the TOE effect. In each pair of stimuli, both stimuli have the same duration.

The solid line in Figs. 6 and 7 represents patterns of activation of the stimulus receptor; two stimuli separated by the ISI have been exposed. The level of activation of the receptor is constant in these examples, but it is not an obligatory condition for the comparison process to work. The receptor activation level may vary, but it has to be positive so that the stimulus can be detected. The dashed line shows patterns of activation of the Sum neuron, which represents the working memory. Note that before and during exposition of the first stimulus, the working memory emits a small positive signal, which satisfies the second assumption from Sec. 5.1. After exposition of the first stimulus, the state of the working memory rapidly increases to the level that represents the amount of pulses stored in the accumulator.

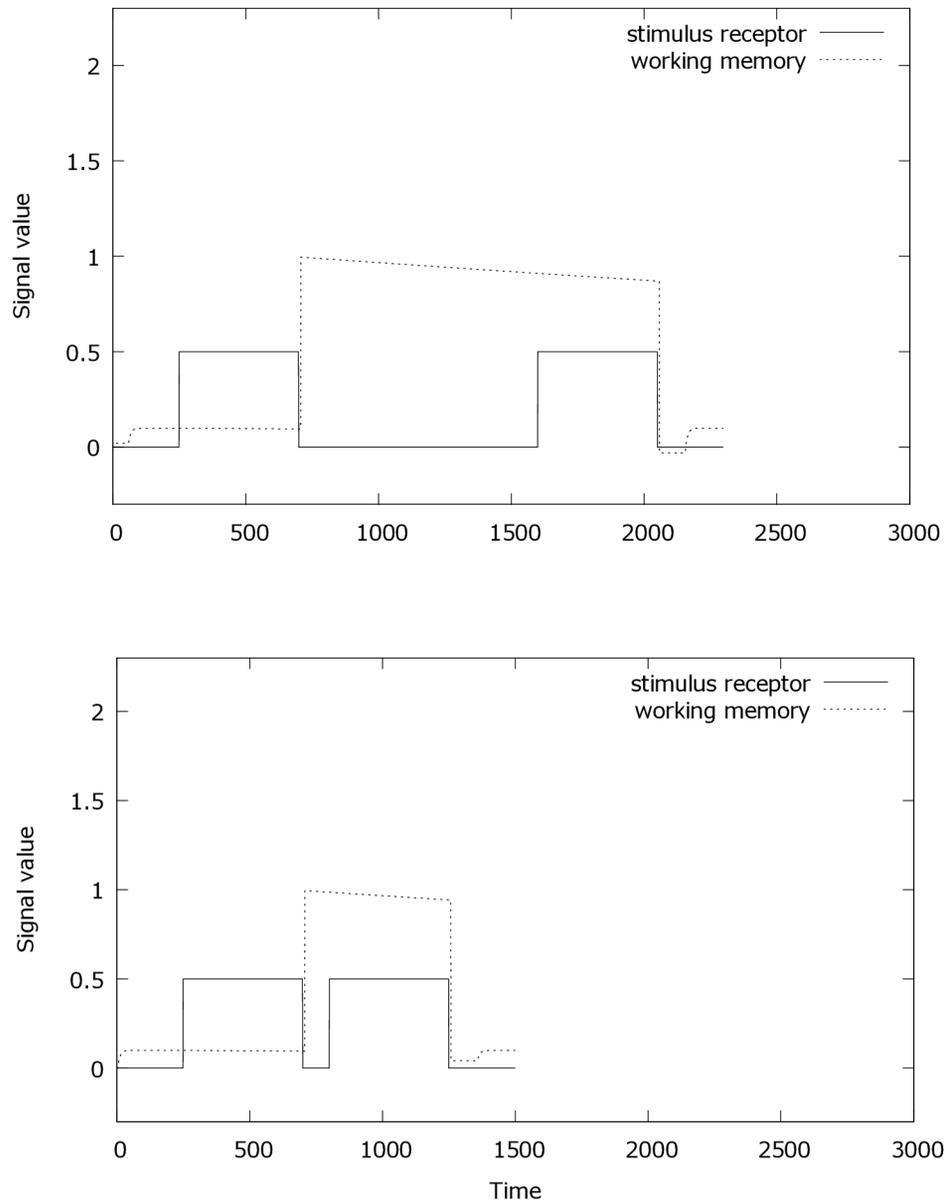
During the ISI and the exposition of the second stimulus, this memory trace fades gradually.

The first illustration (Fig. 6) concerns emergence of both positive and negative TOE depending on the magnitude (length) of the stimuli. Note that the ISI is the same in the upper and the lower panel, so the comparison concerns only one factor – the magnitude of stimuli. If the working memory fading period is long enough, the memory – after comparison – will hold a negative value, which corresponds to the negative TOE. Short fading periods lead to a positive value in memory, and therefore to a positive TOE. There exists one neutral length of the fading period that results in a zero TOE.

The second case (Fig. 7) illustrates influence of the ISI of different lengths on the direction of the TOE. The upper panel demonstrates a negative TOE as the effect of comparison of two identical stimuli separated by a longer ISI. The value of the signal stored in the working memory just after the end of the second stimulus is negative (a negative TOE). The lower panel demonstrates a sample experiment where two stimuli (the same length as in the upper panel) are separated by a short ISI. The result of the comparison is a positive value (a positive TOE) stored in the memory just after exposition of the second stimulus.



**Fig. 6.** Example of the TOE for long and short duration of stimuli. The upper panel illustrates the negative TOE for long durations, and the lower panel shows the positive TOE for short durations. ISIs are the same in both cases.

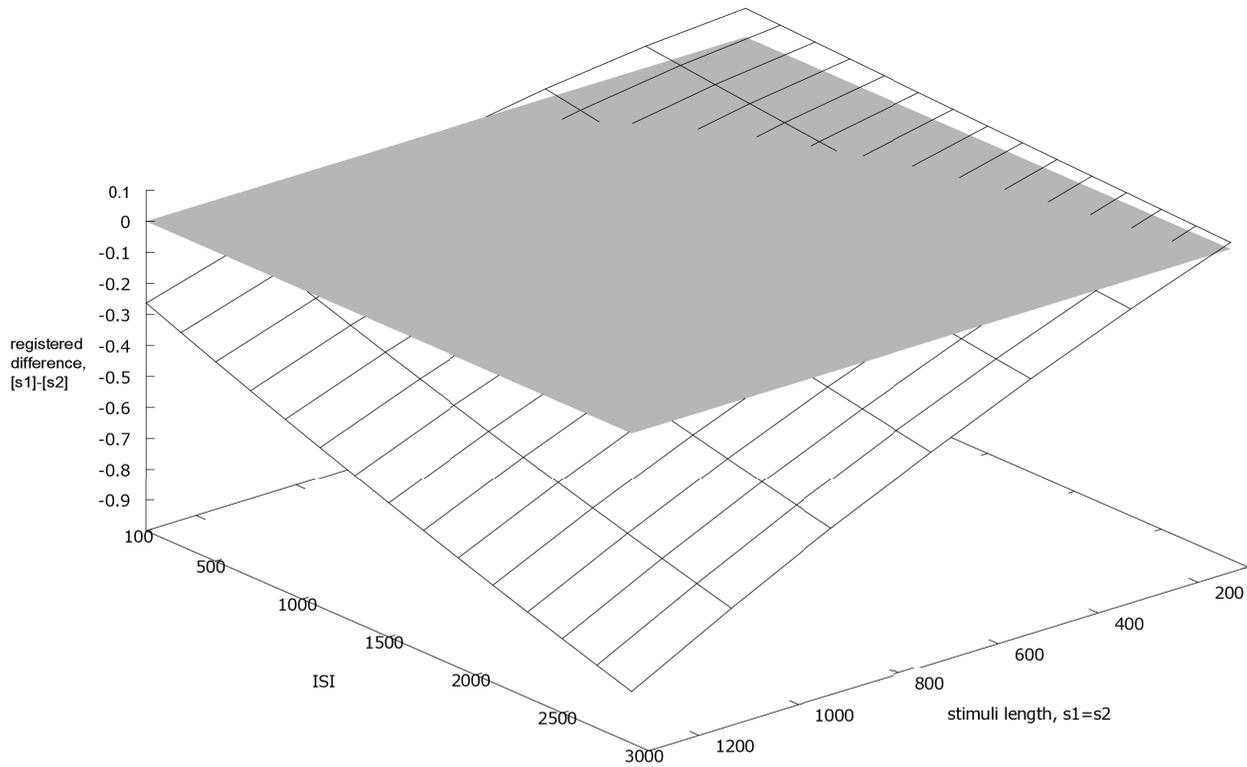


**Fig. 7.** Example of the TOE for long and short ISI. The upper panel illustrates a negative TOE for a long ISI, and the lower panel shows a positive TOE for a short ISI. Durations are the same in both cases.

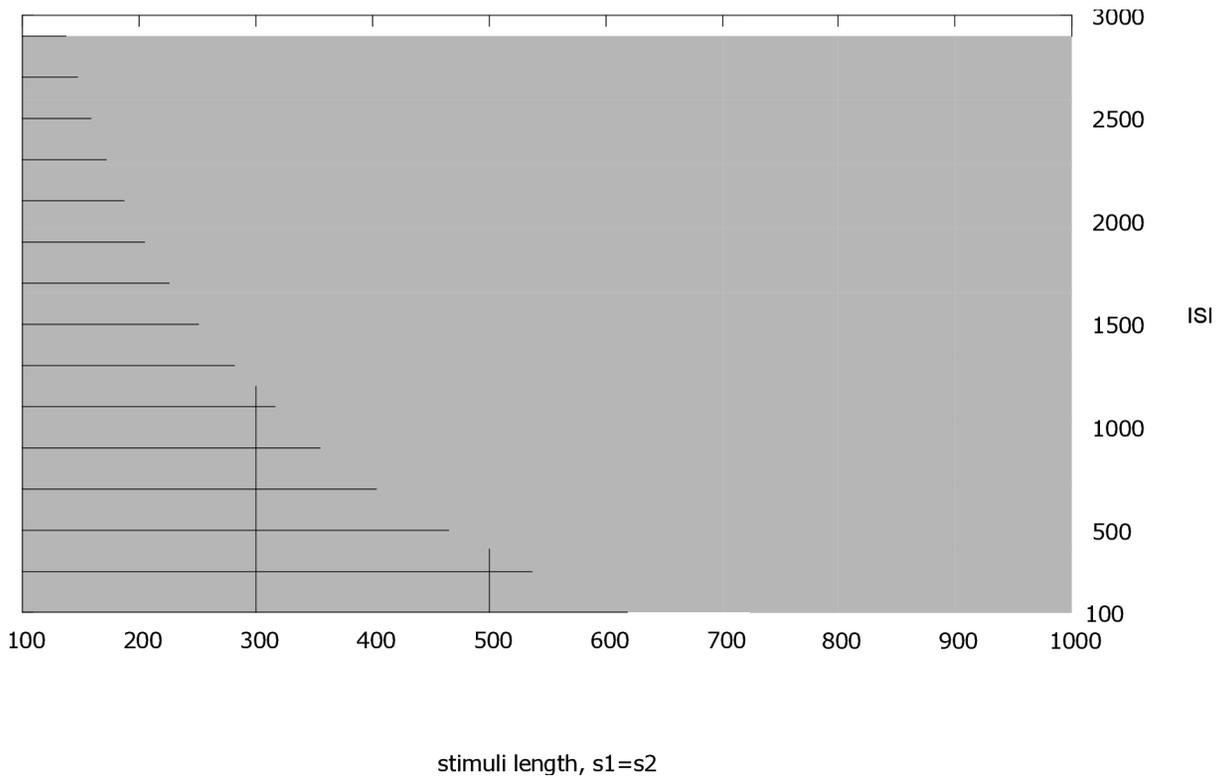
A more complete picture of the relation between the stimuli duration, the ISI and the TOE is presented in Figs. 8 and 9: a nonlinear relationship between the registered differences, the stimuli lengths and the ISI values is visible. This relationship is caused by the proportional fading process in the memory. The gray plane illustrates the points where no difference between stimuli lengths would be registered (note that in this experiment, both stimuli are *always* of the same length). Since the TOE is present, the unbiased measurement of the difference in stimuli lengths takes place only where both planes intersect.

These two elementary experiments are an example of modeling of psychological and psychophysical effects in an artificial

neural network. For an extensive simulation of the TOE, including factors like the influence of different ISIs and different stimuli lengths on the TOE, a few architectural changes are required. For example, to model a decrease in the magnitude of the positive TOE [15], additional components similar to the module "A" in the present implementation may be needed in the network. Note that modeling new phenomena does not always mean the need of adding new modules; sometimes it is sufficient to rearrange existing modules, adjust weights of connections, modify the behavior of neurons, or even simplify the architecture of the network.



**Fig. 8.** Quantitative analysis of the influence of two factors: stimuli length and ISI duration on the registered difference between the two identical stimuli.



**Fig. 9.** A projection (top view) of Fig. 8 demonstrating that in this particular implementation of TOE, no difference in stimuli lengths is registered for a nonlinear combination of stimuli length and ISI.

## 6. Conclusions

This paper describes one of the possible implementations of the Scalar Timing Model. The implementation demonstrates that psychophysical and psychological theoretical models may be successfully implemented as artificial neural networks. Simple experiments examining the TOE effect have been performed. The TOE effect is a starting point for integration of other aspects of time perception that follow from experimental data. Such integration is expected to influence the existing architecture of the STM implementation.

The implementation presented here creates opportunities to generate original research hypotheses inspired by the structure and design of the neural network. Even if some of the hypotheses inspired by the implementation is not valid, the flexibility of the structure of the implemented STM allows for easy modifications according to the newly acquired information (just as it is possible to adjust the distribution of pulses to fit experimental data). Such modifications may result from new facts that are likely to appear in experimental sciences.

The results regarding the pacemaker-accumulator module demonstrate that for the three considered interpulse distributions, the variance-to-mean ratio drops quickly for short stimulus durations and remains constant for long durations. Additional mechanisms (variable pacemaker pulse rate, specific characteristics of the switch, or others) may be introduced in this implementation to obtain the scalar property [6, 5, 7]. The simulation experiments with the current implementation demonstrate violations of the scalar property [30, 19], and illustrate potential difficulties in inferring relationships describing behavior of the model based on imperfect and scarce data [21, 18].

Directions of future research include the use of evolutionary algorithms for optimization and the use of artificial world environment for embodiment of the STM implementation. This implementation, after further development, may become a frame of exchange and integration of information on the verge of many scientific domains.

## Acknowledgement

This work has been supported by the Ministry of Science and Higher Education, grant no. N N519 3505 33.

## References

- Allan L.G.: The time-order error in judgments of duration. *Canadian Journal of Psychology* 1977, 31(1): 24-31.
- Bizo L.A., Chu J.Y.M., Sanabria F., Killeen P.R.: The failure of Weber's law in time perception and production. *Behavioural Processes* 2006, 71(2-3): 201-210.
- Church R.M.: Evaluation of quantitative theories of timing. *Journal of the experimental analysis of behavior* 1999, 71(2): 253-256.
- Church R.M.: A concise introduction to scalar timing theory. Functional and neural mechanisms of interval timing. In: Meck W.H. (ed.), *Functional and Neural Mechanisms of Interval Timing*, CRC Press, 2003, pp. 3-22.
- Getty D.J.: Discrimination of short temporal intervals: A comparison of two models. *Perception & Psychophysics* 1975, 18(1): 1-8.
- Gibbon J.: Ubiquity of scalar timing with Poisson clock. *Journal of Mathematical Psychology* 1992, 35: 283-293.
- Gibbon J.: Multiple time scales is well named. *Journal of the Experimental Analysis of Behavior* 1999, 71: 272-275.
- Gibbon J., Church R.M., Meck W.H.: Scalar Timing in Memory. *Annals of the New York Academy of Sciences* 1984, 423(1): 52-77.
- Grondin S.: From physical time to the first and second moments of psychological time. *Psychological Bulletin* 2001, 127(1): 22-44.
- Grondin S.: Overloading temporal memory. *Journal of Experimental Psychology: Human Perception and Performance* 2005, 31(5): 869-879.
- Hairston I.S., Nagarajan S.S.: Neural mechanisms of the time-order error: An MEG study. *Journal of Cognitive Neuroscience* 2007, 19(7): 1163-1174.
- Hellström Å.: The time-order error and its relatives: Mirrors of cognitive processes in comparing. *Psychological Bulletin* 1985, 97(1): 35-61.
- Hellström Å.: Comparison is not just subtraction: Effects of time- and space-order on subjective stimulus difference. *Perception & Psychophysics* 2003, 65(7): 1161-1177.
- Hellström Å., Rammsayer T.H.: Effects of time-order, inter-stimulus interval, and feedback in duration discrimination of noise bursts in the 50- and 1000-ms ranges. *Acta Psychologica* 2004, 116: 1-20.
- Jamieson D.G., Petrusic W.M.: The dependence of time-order error direction on stimulus range. *Canadian Journal of Psychology* 1975, 29(3): 175-182.
- Jamieson D.G., Petrusic W.M.: Pairing effects and time-order errors in duration discrimination. *Perception & Psychophysics* 1975, 18(2): 107-113.
- Jamieson D.G., Petrusic W.M.: Presentation order effects in duration discrimination. *Perception & Psychophysics* 1975, 17(2): 197-202.
- Komosinski M.: Analog to digital: Measuring quantities using oscillators and pulse generators. *Theory in Biosciences*, submitted.
- Komosinski M., Kups A.: Models and implementations of timing processes using Artificial Life techniques. Technical Report RA-05/09, Poznan University of Technology, Institute of Computing Science, 2009.
- Komosinski M., Ulatowski Sz.: Framsticks web site, <http://www.framsticks.com>.
- Lewis P.A., Miall R.C.: The precision of temporal judgement: milliseconds, many minutes and beyond. *Philosophical Transactions of the Royal Society B* 2009, 364(2): 1897-1905.
- Matell M.S., Meck W.H.: Cortico-striatal circuits and interval timing: coincidence detection of oscillatory processes. *Cognitive Brain Research* 2004, 21: 139-170.

23. Needham J.G.: The effect of the time interval upon the time-error at different intensive levels. *Journal of Experimental Psychology* 1935, 18: 530-543.
24. Rammsayer T., Ulrich R.: Counting models of temporal discrimination. *Psychonomic Bulletin & Review* 2001, 8(2): 270-277.
25. Rammsayer T.H., Grondin S.: Psychophysics of human timing. In: Miller R. (ed.), *Time and the brain*, Amsterdam: Harwood Academic Publishers, 2000, pp. 157-168.
26. Schab F.R., Crowder R.G.: The role of succession in temporal cognition: Is the time-order error a recency effect of memory? *Perception & Psychophysics* 1988, 44(3): 233-242.
27. Staddon J.E.R., Higga J.J.: Time and memory: Towards a pacemaker-free theory of interval timing. *Journal of Experimental Psychology: Animal Behavior Processes* 1999, 71(2): 215-251.
28. Stott L.H.: Time-order errors in the discrimination of short tonal durations. *Journal of Experimental Psychology* 1935, 18(6): 741-766.
29. Stott L.H.: The effect of practice on positive time-order errors. *Journal of Experimental Psychology* 1936, 19(6): 694-705.
30. Wearden J.H., Lejeune H.: Scalar properties in human timing: Conformity and violations. *The Quarterly Journal of Experimental Psychology* 2008, 61(4): 569-587.
31. Wearden J.H.: "Beyond the fields we know...": exploring and developing scalar timing theory. *Behavioural Processes* 1999, 45: 3-21.
32. Wearden J.H.: Applying the scalar timing model to human time psychology: Progress and challenges. In: Helfrich H. (ed.), *Time and mind II: Information processing perspectives*, Cambridge, MA: Hogrefe & Huber Publishers, 2003, pp. 21-39.
33. Wearden J.H., Norton R., Martin S., Montford-Bebb O.: Internal clock processes and the filled-duration illusion. *Journal of Experimental Psychology: Human Perception and Performance* 2007, 33(3): 716-729.
34. Woodrow H.: The effect of practice upon time-order errors in the comparison of temporal intervals. *The Psychological Review* 1935, 42(2): 127-152.
35. Zakay D., Block R.A., Tsal Y.: Prospective duration estimation and performance. In: Gopher D., Koriat A. (eds), *Attention and Performance XVII*, Cambridge, MA: MIT Press, 1999, pp. 557-580.

## A CONTINUOUS APPROACH TO THE ECG NOISEPRINT ESTIMATION

PIOTR AUGUSTYNIAK

*Institute of Automatics, AGH University of Science and Technology, Kraków, august@agh.edu.pl*

**Abstract:** Signal quality is a common problem in biomedical applications, as it impacts the reliability of electrocardiogram interpretation. The demand for a need for a dependable signal-to-noise measure is reinforced by widespread telemedical recordings in the home care conditions being interpreted automatically. Present techniques are based on baseline noise measurement and assumption about temporal noise stability. This paper presents an alternative approach to ECG noiseprint estimation technique based on a noise model calculated from seamless time-frequency representation. The key principle is recognition of area for possible cardiac components with use of temporarily adapted local bandwidth variability function. The part free of cardiac influence, above the local bandwidth of the ECG, represents background activities of any origin (muscle, mains interference etc.). Next, we consider the non-uniformly sampled time series in each particular scale as piecewise noise estimate and apply interpolation techniques to estimate the noiseprint in regions containing the ECG components. The algorithm was implemented and tested with use of the CSE Database records with the addition of the MIT-BIH Database noise patterns. The differences between the added and estimated noise show similar performance of baseline-based and noise model-based methods (accuracy of 0.69 dB and 0.64 dB respectively) as long as the noise level is stable.

**Keywords:** signal-to-noise ratio, time-frequency modeling, electrocardiogram,

### 1. Introduction

Growing number of recordings gathered in unpredictable home care conditions raises the question of revision of ECG noise level assessment methods. Unavoidable simultaneous activity of adjacent organs (e.g. muscles) and unstable recording conditions (e.g. electrodes contact) distort the record and affect the diagnostability of physiological signals. In both cases, the noise occurs randomly overlapping the ECG signal in both time and frequency domains. Many noise measurement and removal techniques were proposed in the past including signal averaging [1], adaptive noise canceling [2] or wavelet-based noise reduction [3-7]. Although relying on various principles, these techniques hardly negotiate broadband noise variation and thus are not applicable to home care recordings. Main drawback is relying on the baseline as an interval of documented electrical inactivity of the heart when picking a reference value for noise measurement. In a real ECG recordings short duration of the baseline limits the noiseprint bandwidth, and rare, irregular occurrence of the baseline limits the contribution of measurement points.

A continuous approach presented in this paper consists in recognition of cardiac-originated components and background electrophysiological components based on coordinates of each particular atom of time-frequency plane in context of the local bandwidth of the electrocardiogram (LBE). Time-frequency atoms

found out of the instantaneous ECG bandwidth are identified as noise measurement points and comprise a large percentage of values of quasi-continuous noise model. Interpolation or extrapolation was applied for only few gaps when cardiac representation fills all the time-frequency range. The continuous measure of signal-to-noise ratio (SNR) is then calculated as relation of energy of cardiac components to the energy of the noise model expressed directly in the time-frequency domain. The LBF is individually adapted to each consecutive heartbeat with use of respectively detected wave borders (Fig. 1a). Consequently the method easily adapts to the local variability of background activity, compensates for variability of the heart rate and favors the measured noise information over the estimates.

### 2. Qualification of time-frequency components

In ECG recordings a sampling frequency of 500 or 1000 Hz is commonly used, corresponding to the bandwidth of relatively short QRS complex and significantly overestimated for other cardiac components occupying the majority of recording time. This local oversampling yields a gap above the bandwidth expected for cardiac components other than the QRS and the Nyquist frequency. Similarly to the baseline the signal represents temporal inactivity of the heart in specified frequency range (scale), but

lasts significantly longer (Fig. 1b). These intervals are then used to measure the noise level at high frequency (three upper scales) directly on a scalo-temporal plane. We found reasonable to correlate the local bandwidth estimate with positions of particular component of the heart cycle (the P, QRS and T waves), rather than by an explicit value of time coordinate, because of different nature and expected bandwidth within each of these waves. We assume that the wave limits can be automatically calculated for each heart beat with acceptable reliability. Consequently, the noise measurement intervals are determined in three upper scales in relation to the diagnostic content of the corresponding electrocardiogram.

Several attempts were previously made in our research to determine the local bandwidth of the ECG:

- studies of susceptibility of diagnostic parameter values to signal distortion caused by the local random canceling of time-frequency coefficients [8],
- analysis of expert’s perception of the ECG trace revealing local signal conspicuity and thus its sectionwise relevance to the final diagnosis [9],
- individual assessment of the spectra for optimally-concatenated successive waves of selected types [10].

The local bandwidth of the electrocardiogram (LBE) is then a heuristic discrete function  $f(n)$  specifying relative bandwidth expected at the time point  $n$ :

$$f: \forall n \in \{0, 1, \dots, N\} \rightarrow f(n) \in [0; 0.5] \quad 1)$$

This function, using  $k_1 \dots k_5 \in \{0, 1, \dots, M\}$  as the representation of the standard positions of wave borders is projected to the local position of current heartbeat wave borders  $h_1 \dots h_5 \in \{0, 1, \dots, M\}$  for each point  $i = 1 \dots 5$  (Fig. 1):

$$\forall n \in [k_i, k_{i+1}], \forall m \in [h_i, h_{i+1}] \quad f(m) = P^{S_i}(F(n)) \quad 2)$$

with projection scale  $S_i$  varying from section to section:

$$S_i = \frac{h_{i+1} - h_i}{k_{i+1} - k_i} \quad 3)$$

The LBE function separates the cardiac and noise components and sets a borderline for seamless measurement of noiseprint. Time-frequency atoms of raw ECG representation are qualified as cardiac components only for scale  $j$  and time point  $m$  satisfying:  $f(m) > 2^{-j-1}$ . Otherwise they are considered as extra-cardiac components (noise representation). It is noteworthy than thanks to the reliability of wave border positions out approach yields a time-frequency model of ECG background activity including as many measured data points as possible, without a risk

of confusions of the noise with the cardiac representation. For each scale, the time-frequency noise model can be updated immediately and accurately except for the short series of atoms qualified as cardiac representation.

### 3. Seamless measurement of signal-to-noise ratio

In the sections where representation of cardiac activity is expected, the noise level measurement is not feasible, and the noise level has to be estimated from the neighboring measured values. Consecutive values of measurement noise atoms in separate scales  $N_j, j \in \{1 \dots 3\}$ , are considered as non-uniformly sampled time series  $N_j(\{n, v(n)\})$  and projected to the regular space [11] using the continuous function:

$$S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \quad 4)$$

where  $x \in [x_p, x_{i+1}], i \in \{0, 1, \dots, n-1\}$  best fitted to the time series  $N_j$  are known as cubic splines interpolation. The uniform representation of the noise, extended to the cardiac component area in three upper scales, is then obtained by sampling the  $S_i(x)$  at the time points  $m$  (Fig. 2a):

$$N'_j(m) = \sum_m S_i(x) \times \delta(x - mT) \quad 5)$$

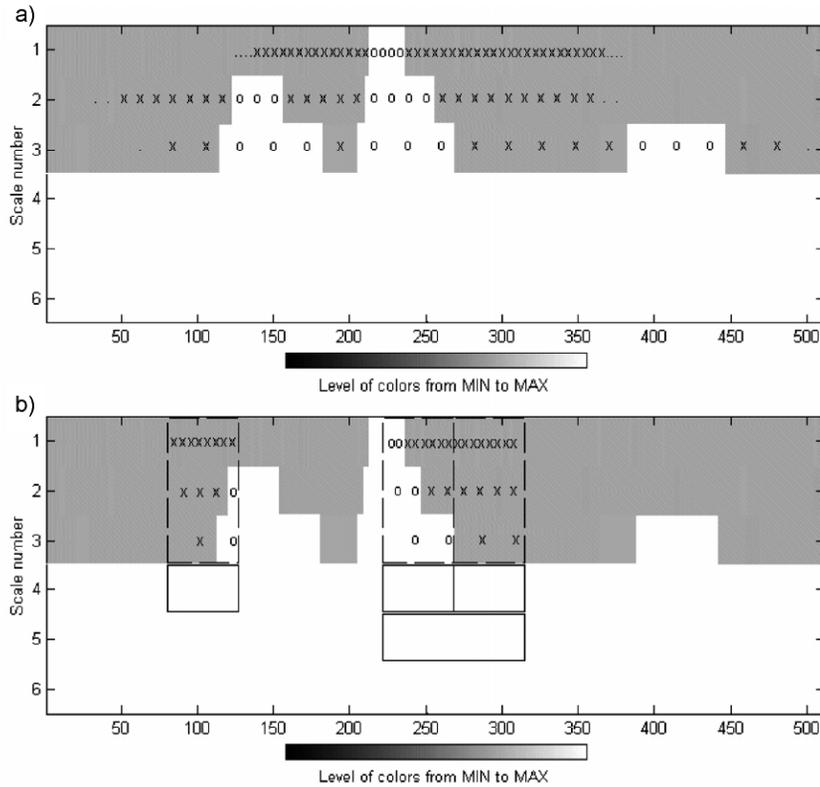
As the scale number increases, the contribution of cardiac representation groves and below third scale (32 Hz), the bandwidth is continuously occupied by the representation of cardiac activity restraining from reliable measurement of noise. In that case we applied noise extrapolation based on coefficients from first three scales estimate the noise level in lower frequencies. This extrapolation uses second-order polynomials generated by all atoms of embedded trees originating from the considered coefficient:

$$N'(k, j) = a_{k,j} \times j^2 + b_{k,j} \times j - c_{k,j} \quad 6)$$

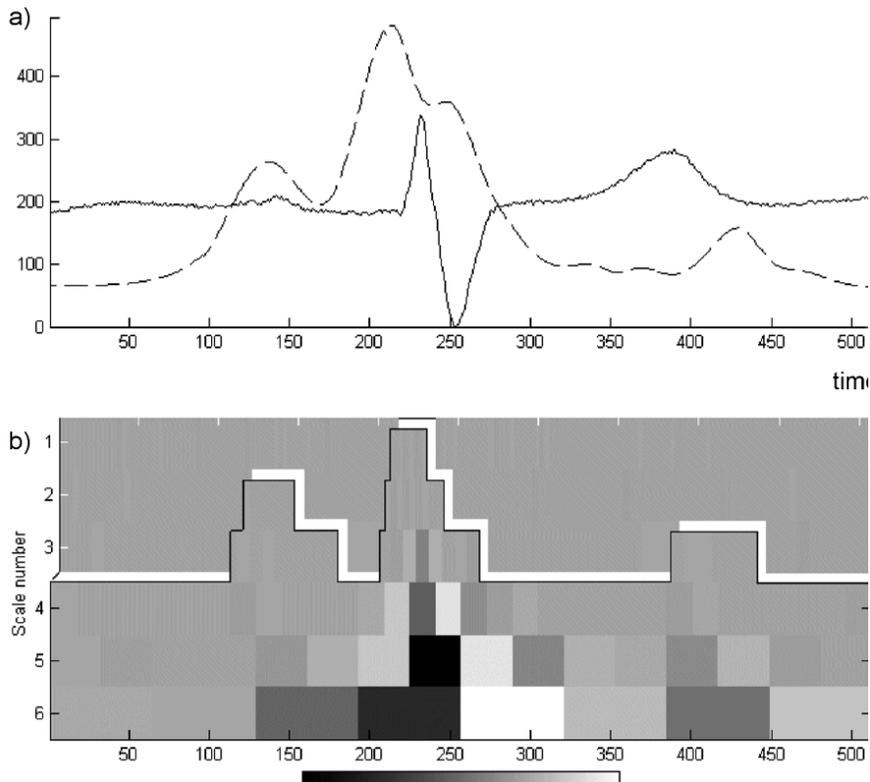
Therefore, the estimation of the noise level at a given time point  $k$  on the scale  $j$  is based on three average values  $M_j(k, i)$  from all atoms  $s(n, i)$  within a corresponding time interval on each of the first three scales (Fig. 2b):

$$M_j(k, i) = \frac{1}{2^{j-1}} \sum_{n=2^{j-1} \times k}^{2^{j-1} \times (k-1) - 1} s(n, i) \quad 7)$$

The use of parabola-based extrapolation is a compromise between accurate approximation of the shape of muscular activity spectrum and temporal stability of the noise model.



**Fig. 1.** a) The example heartbeat (solid) and the adapted bandwidth variability function (dashed). b) Corresponding time-frequency signal representation divided in the noise measurement region (above the local cut-off frequency) and the cardiac representation region (below)



**Fig. 2.** (a) Distribution of noise measurement and interpolation samples in first three scales. (b) Extrapolation of noise values to low frequency bands with averaging of the noise print in the time domain. Missing values 'o' are estimated from adjacent measured values 'x'

### 4. Testing conditions

Testing the noiseprint measurement procedure is a difficult task because of the lack of noise-free real signals. The synthesized electrocardiogram, although noise-free, differs in diagnostically meaningful details from real records. For this reason with use of stationary noise we perform two complementary experiments using real and synthesized ECG records. In all cases, three noise patterns mixed with the ECG record were representative for common noise sources:

- poor electrode contact (mainly low frequency and abrupt baseline changes),
- electromagnetic interference (mains sinus wave, 60 Hz),
- muscle fibrillation (a clipping-free section was selected).

All patterns were taken from the MIT-BIH Database (12 bit, 360 Hz) [12], resampled and mixed with the signal at the following four levels of total energy: 5%, 10%, 20% and 50% (corresponding to -13 dB, -10 dB, -7 dB and -3 dB signal-to-noise ratio). The artificial records were acquired with use of an ECG recorder (12 bit, 500 Hz) from the test generator (PS450, Fluke) and the real ECG records were taken from the CSE Multilead Database (12 bit, 500 Hz) [13] (Fig. 3). The artificial ECG signals are synthesized mathematically in the generator's low-noise hardware with no measurable random component. The time-frequency representation of the ECG perfectly matches the expected local bandwidth variability. Therefore any difference of estimated and added noise levels is interpreted as SNR-estimator inaccuracy.

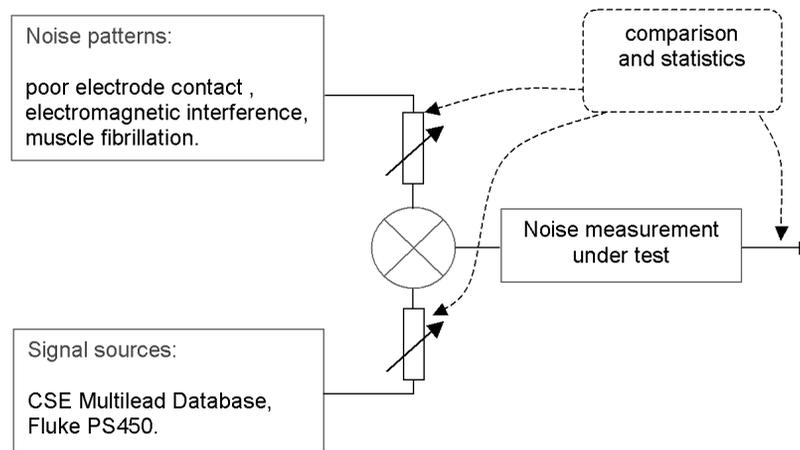


Fig. 3. Block diagram of the noise measurement testing

### 5 Test results

Performance of both baseline-based and proposed noise model-based SNR measurements is satisfactory in case of mixing the static muscular noise with the ECG's, mainly due to separate spectra of the signal and noise (Tab. 1 last row, Fig. 4). Unfortunately, results for poor electrode contact are not acceptable for both methods.

The results for real database records show lower performance, due to noise recorded in the reference signals, having the characteristics very close to the noise added artificially. This test, however, shows the behavior of SNR estimators in presence of irregular duration of ECG sections and waves.

noise pattern level [dB]	synthesized				CSE Database			
	-3	-7	-10	-13	-3	-7	-10	-13
poor electrode contact	8	45	57	58	6	35	53	46
electromagnetic interference	66	78.5	87	81	66	78.5	87	81
muscle fibrillation	80	93	92.9	93.4	72	88.5	90	83

Tab. 1. The average noise measurement accuracy [%] for synthesized and CSE-originated signals for three patterns of static noise

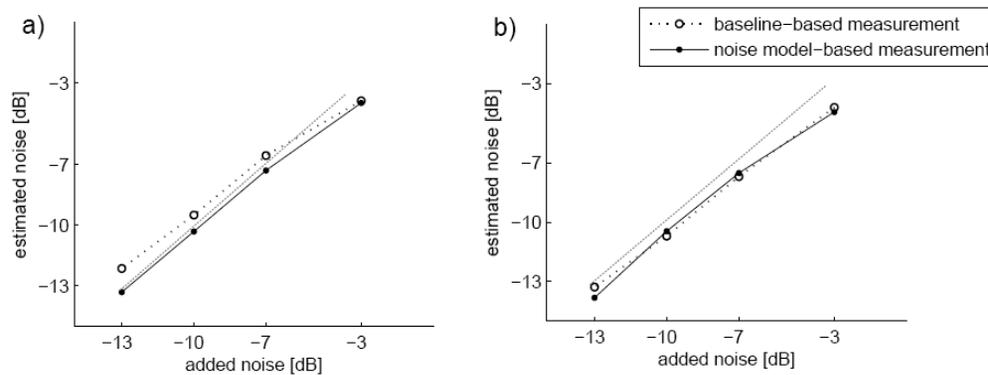


Fig. 4. SNR estimation performance for static muscular noise, plots show noise value estimated versus added (a) artificial signals, (b) CSE-originated signals

## 6. Discussion

The ECG noise recognition and calculation of signal-to-noise ratio, based on the definition of the local bandwidth of signal (i.e. cardiac components), are correctly performed in proposed algorithm. This results from our approach that maximizes the number of noise measurement points in the time-frequency plane. Comparing to baseline-based SNR measurement using at most 10% of samples for noise level estimation, the presented method performs the noiseprint measurement on ca. 70% of the time-scale atoms, and interpolates only for remaining 30%. In consequence the noise tracking is more accurate and follows the rapid changes of the noise level, except if they occur in a QRS section (typically lasting for approximately 100 ms) where the full bandwidth of signal is used by the cardiac components. The measured noise data is locally missing in a QRS section (typically lasting for approximately 100 ms) where the full bandwidth of signal is used by the cardiac components and the interpolation cannot correctly reconstruct the values of time-frequency coefficients. In practice this drawback has only marginal importance, because usually the amplitude of cardiac components on the QRS section is high and the signal masks the noise estimation errors.

The characteristics of noise patterns applied during tests differ in power spectra density and influence the performance of the noise modeling and estimation algorithm. Since the poor electrode contact noise concentrates on lower scales, it cannot be correctly estimated using values measured from the three upper scales. The characteristics of muscular activity best fit the initial assumptions: in the time domain signal changes are relatively slow and in the frequency domain the parabolic spectrum approximation is quite accurate. This is the reason for the best accuracy of the proposed algorithm in the tests with the muscle noise.

Since the time-frequency estimates of contribution from cardiac components and noise level is expressed in energy units (i.e. square values of time-frequency coefficients), comparing their values calculated over all the signal duration yields directly the general and commonly used value of signal-to-noise ratio. However, considering a non-uniform distribution of medical

information in the electrocardiogram, an interesting property of the proposed approach is the support for time-selective or event-specific assessment of the signal quality.

## 7. Acknowledgment

This Scientific work is supported by the AGH University of Science and Technology as a research project No. 11.11.120.612

## References

1. Moss A., Stern S.: *Noninvasive Electrocardiology – clinical aspects of Holter monitoring*. London: Saunders, 1996.
2. Akay M.: *Biomedical signal processing*. San Diego: Academic Press, 1994.
3. Akay M. (ed.): *Wavelets in Biomedical Signal Processing*. New York: IEEE Press, 1998.
4. Krishnan S., Rangayyan R.M.: Automatic de-noising of knee-joint vibration signals using adaptive time-frequency representations. *Med. Biol. Eng. Comput.* 2000, 38: 2-8.
5. Nikolaev N., Gotchev A.: De-noising of ECG signals using wavelet shrinkage with time-frequency dependant threshold. *Proc. European Signal Processing Conf. EUSIPCO-98*, Island of Rhodes, Greece, 1998, pp. 2449-2453.
6. Nikolaev N., Gotchev A., Egiastian K., Nikolov Z.: Suppression of electromyogram interference on the electrocardiogram by transform domain denoising. *Med. Biol. Eng. Comput.* 2001, 39: 649-655.
7. Paul J., Reedy M., Kumar V.: A transform domain SVD filter for suppression of muscle noise artefacts in exercise ECG's. *IEEE Trans. Biomed. Eng.* 2000, 47: 645-662.
8. Augustyniak P.: Controlling the Distortions Distribution in a Wavelet Packet-Based ECG Compression. *International Conference on Image and Signal Processing*, Agadir, Morocco, 2001, pp. 267-277.
9. Augustyniak P.: How a Human Perceives the Electrocardiogram. *Computers in Cardiology* 2003, 30: 601-604.

10. Augustyniak P.: Moving Window Signal Concatenation for Spectral Analysis of ECG Waves. *Computing in Cardiology* 2010, 37: 665-668.
11. Aldroubi A., Feichtinger H.: Exact iterative reconstruction algorithm for multivariate irregularly sampled functions in spline-like spaces: the  $L_p$  theory. *Proc. Amer. Math. Soc.* 1998, 126(9): 2677-2686.
12. Moody G.B.: The MIT-BIH arrhythmia database CD-ROM (Third Ed.). Harvard-MIT Division of Health Sciences and Technology, 1997.
13. Willems J.L.: Common Standard for Quantitative Electrocardiography Multilead Atlas – Measurements results Data Set 3. Commission of the European Communities – Medical and Public Health Research, Leuven, 1988.

## SSTMPROT, A POINT MUTATION SENSITIVE TOOL TO COMBINE RESULTS AND TO PREDICT THE CONSENSUS SEQUENCE, AND SECONDARY STRUCTURE OF TRANSMEMBRANE PROTEINS

JAN PAWEŁ JASTRZĘBSKI

*Department of Plant Physiology and Biotechnology, Faculty of Biology,  
 University of Warmia and Mazury in Olsztyn, ul. Oczapowskiego 1A p.113, 10-719 Olsztyn / Poland*

**Abstract:** The efficiency of predictions of protein secondary structures can be increased by treating this process as a cycle of steps, where each step is an approach to the single natural event in folding process (model and simulation of successive events). The set of simulation steps qualifies reliability of *in silico* simulation of single steps, allowing to verify the correctness of each step as well as to retain sensitivity in case of a single amino acidic substitution. For this purpose the three-part algorithm (*SSTMProt*) has been designed. This algorithm combines the results of known methods of prediction of proteins secondary structure. Furthermore, the efficiency of this algorithm has been verified using the models received from the RCSB PDB (the Research Collaboratory for Structural Bioinformatics – Protein Data Base; <http://www.rcsb.org>). The accuracy of known methods has been compared with the accuracy of designed algorithms. The accuracy has been tested by the comparison of true secondary structure with predicted secondary structure of a given protein. The results of accuracy test has been presented as percentage values of similarity between both secondary structures: predicted structure using known method vs. true structure and predicted structure using designed method vs. true structure. The results demonstrate 20-30% higher accuracy of prediction for designed algorithms then for adequate known methods.

The test of sensitivity has been done for proteins of a very conservative and stable structure (subunits of bovine cytochrome c oxidase and bacterial ATP-ase, bovine rhodopsin and human hemoglobin as a globular but alpha-helical protein). The influence of a single amino acid substitution on a resulted secondary structure predicted by *SSTMProt* algorithms has been examined. The repeatability of elaborated algorithms is 100% and each of all 12 tested combinations of methods were sensitive on a single amino acid substitution. All tests have been done for 10 models of native forms of proteins of known structure (models downloaded from the RCSB PDB 1HBB, 1HBS, 1OCC, 1U17, 1C17) and over 500 modified models; 30 known methods of prediction of secondary structure of proteins and 40 combinations of these methods included in three versions of elaborated algorithms have been examined for each protein model.

**Keywords:** Secondary structure, structure prediction, transmembrane proteins, bioinformatics, sensitivity, algorithm

### Introduction

The function and biological activity of proteins are determined in the main part by its spatial conformation. The spatial structure of a peptide chain determines: the arrangement of multiple polypeptide chains in a multi-subunit complex of different molecules (the protein quaternary structure), the shape of active sites, requires exposition to ligands, the properties of peptide environment of coenzymes, electrostatic surface, flexibility of polypeptide backbone, facility of integration of transmembrane proteins into the lipid membrane and other features which specify the function and biological activity of a molecule. The final structure of the peptide chain is determined during formation of secondary structures [1, 2]. Initial secondary structures start to reveal during

translation process, from the beginning of existence of the first part of peptide chain (first four amino acids can start to fold an  $\alpha$ -helix) [1, 2, 3, 4]. Correct estimation of the protein secondary structure can be a key-step of 3D structure prediction process. In addition, the reduction of structure prediction error to  $\pm 1\%$  allows estimating the influence of a single amino acid substitution on the final protein structure with high probability.

The accuracy of tested known prediction methods reaches 40-80%, where the upper limit is not reachable for the main part of methods. So far, homology modeling methods (using experimentally solved protein structures as a templates) are the most effective. The accuracy of homology modeling methods reach in some cases 100%, but the sensitivity of these methods for a single point mutation practically does not exist. The accuracy of methods based on the neural networks reaches 85-95% [5, 6,

7], but the problem is, that it is not possible to control precisely each step of this kind of prediction. The algorithms such as GOR, Gibrat, DPM [8, 9, 10] are information probability theory-based methods and input parameters derived from empirical studies of known protein tertiary structures [11, 12, 13]. It can be postulated the action of the theory-based method is easy and has biological basics. For instance the tool of the prediction of transmembrane regions implemented in Antheptrot program [14] utilizes on the hydrophobicity method [15] or on a more specific method such as the positive inside rule [16]. The accuracy of secondary structure predictions of these methods is low, reaches 40-55% for transmembrane proteins (see Fig. 5) and is not sufficient to precise analysis.

Designed algorithms include mechanisms of selected known methods and additional processes elaborated following described steps of protein folding (see *Material and methods*).

The algorithm has been written for transmembrane proteins. The maturation of proteins of this superfamily can be easily divided into separate processes. Finally, three similar algorithms have been elaborated. Two of them (*TM+SecStr*, *TM+(SecStr-Helix)*) have been prepared to verify the correctness of used restrictions. The third algorithm (*Aver(TM+SecStr)*) has been designed and calibrated as an proper algorithm to predict the secondary structure of a protein with high precision and high sensitivity to any changes in the amino acidic sequence. The group of designed algorithms has been named as the *SSTMProt* (**SEC**ondary **Str**ucture of **Tr**ans**M**embrane **P**rotein).

## Materials and Methods

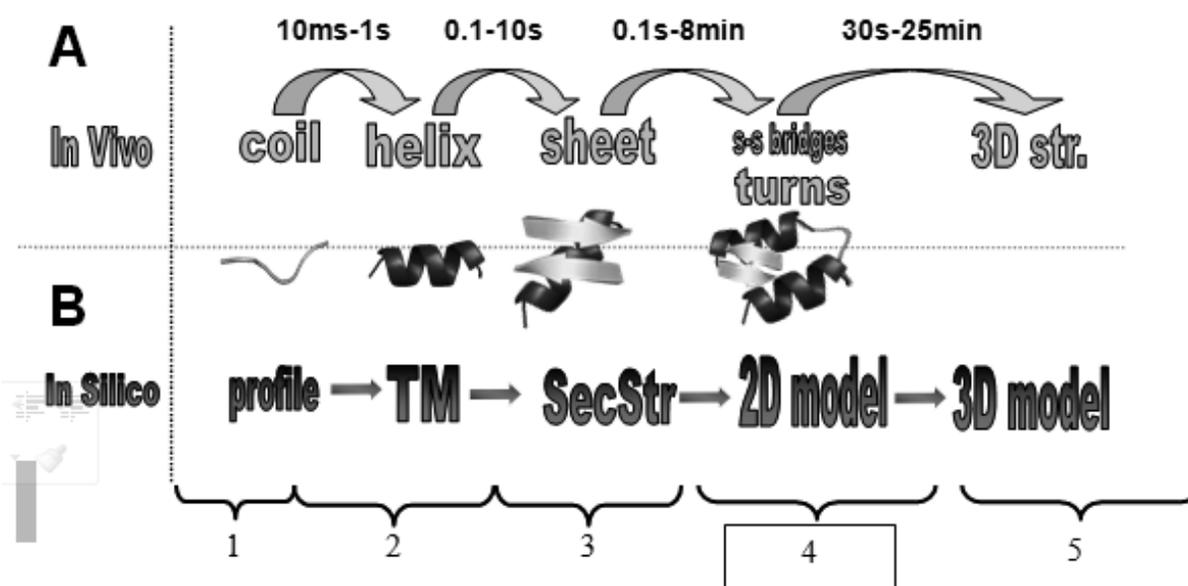
The mechanism of elaborated algorithms has been designed following the pathway of protein folding in accordance with the process occurred in nature (Fig. 1a).

Proper secondary structure prediction method has been adopted for each step of designed methodology based on the data from literature (Fig. 1b). Particular steps of forming secondary structures have been combined in three different ways by three algorithms:

1. *TM+SecStr*
2. *TM+(SecStr-Helix)* // both algorithms designed to control and verify prediction process,
3. *Aver(TM+SecStr)* // proper algorithm to predict secondary structure.

The whole designed prediction process starts from prediction of secondary structure using any method (in this case using one of the known methods) – this is the first step. The second step is to predict the transmembrane region.

1. The *TM+SecStr* algorithm summarize results of both prediction steps:
  - regions predicted as helix in any step or in both steps finally are defined as a helix
  - regions predicted as a coil in both steps finally are defined as a coil
  - regions predicted as a coil in first step and predicted as any secondary structure (excluding coils) in the second step, finally are defined as a secondary structure from the second step.



**Fig. 1.** The scheme of the empirical basis of designed algorithm. Figure A is the scheme of natural protein folding process with approximate times of each step of structure forming. B – designed procedure consists of set of steps of protein structure prediction. This methodology and designed algorithms (*TM+SecStr*, *TM+(SecStr-helix)*, *Aver(TM+SecStr)*) has been developed for transmembrane proteins. Steps: 1 — chemical and physical protein profiles; 2 — transmembrane prediction methods; 3 — secondary structure prediction methods; 4 — *SecSTMProt*: *TM+SecStr*, *TM+(SecStr-helix)*, *Aver(TM+SecStr)*; 5 — tertiary structure prediction methods.

This algorithm allows to form helices longer than even 50 amino acids (it occurs rarely in nature, because helices longer than 16-21 amino acids are going to destabilize [17, 18] and longer than 26-30 residues are not stable [19, 20, 21, 22] (excluded some examples as creatin, myosin, actin [23, 24] and other two-stranded alpha-helical proteins [25, 26]). That is why this algorithm has been used only to control the prediction and to verify the correctness of basic restrictions of designed methodology.

2. The *TM+(SecStr-Helix)* algorithm works in a very similar way to the *TM+SecStr* algorithm. The only difference is that each  $\alpha$ -helix position can fold only during the first step. Each  $\alpha$ -helix position predicted in the second step is treated as a coil. The main principle of this algorithm is that helices can form only in the first step of a folding process, the structure of helices is very stable (helical spatial

conformation of amino acidic chain is supported by numerous short chemical and physical interactions as hydrogen bonds, Van der Waals interactions, stacking interactions, hydrophobic interactions) and other secondary structures (like  $\beta$ -sheets, turns, etc.) can fold between helices during the second step. The level of accuracy of prediction using this algorithm is acceptable only for transmembrane proteins (see Fig. 6 and 7), but allows to form very short single  $\beta$ -sheets.

3. The *Aver(TM+SecStr)* algorithm combines the results of each step of prediction taking into consideration more restrictions and rules of protein folding process occurring in natural circumstances. The most important assumptions of this algorithm are:
- Positions assigned to the coil in all steps finally are defined as a coil (Fig. 2).

1st step – cccccchhhhhhhhhhhhhcccccccccccccccccc  
 2nd step – ccchhhhhhhhhhhccccccccccccccchhhhhhhc

result\_1 – ccc---hhhhhhh-----cccccccc-----c

**Fig. 2.** The scheme of basics of the mechanism of helix and coil regions determination by the *Aver(TM+SecStr)* algorithm; c – coil, h – helix, dash – moot region, result\_1 – temporary outcome sequence.

- Positions assigned to the helix in first step and to the coil in another step (order of steps is not important in his case) are defined as a “moot region” (in this case “coil/helix moot region”) in temporary outcome sequence – **result\_1** (signed by dashes in Fig. 2 and 3).
  - Final positions assigned to the helix occupy a half of “coil/helix moot region” from existing helix site (existing helix on **result\_1** sequence) (Fig. 3).
  - In the case of odd residues of “mood region” the number of positions to be signed as helix is round up to the nearest integer (Fig. 3).
  - If “mood region” borders with coil positions on both sides, helix region will occupy half a number of positions rounding up to the nearest integer of “mood region” (Fig. 3).
  - Single helix region should not be longer than 21 positions; helix region can be longer than 21 amino acids only if prediction process fulfills restriction no.1;
  - single helix region cannot be longer than 30 positions
    - each helix longer than 30 amino acids in the second
- temporary outcome sequence have to be shortened to maximum 30 positions in accordance with restrictions no. 9 and 10.
- If one helix region is too long (more than settled value – default 30 positions) it will be shortened following the criteria:
    - average length of transmembrane region of helix occupies about 30 positions [27, 28];
    - hydrophobicity and hydrophilicity thresholds should be high enough to determine helices shorter than 21 positions and has to be high enough to determine helices shorter than 30 positions [15];
    - if hydrophobicity profile is not high enough to shorten helix region maximum to 30 residues, the helix should be broken in the region of the lowest hydrophobicity or by the helix brokers (aminoacids as proline or “histidine triplet”).
  - Space between two helices cannot be shorter than 3 positions.

1st step – cccccchhhhhhhhhhhhhcccccccccccccccccc  
 2nd step – ccchhhhhhhhhhhccccccccccccccchhhhhhhc  
 result\_1 – ccc---hhhhhhh-----cccccccc-----c  
 complement – ...cchh.....hhhc.....cchhhcc.  
 result\_2 – ccccchhhhhhhhhhhhhccccccccccccccchhhhhcc

**Fig. 3.** The scheme of the mechanism of complement “mood regions” on the temporary outcome sequence – result\_1 by the *Aver(TM+SecStr)* algorithm; c – coil, h – helix, dash – moot region, dot – primary defined region, result\_1 – first temporary outcome sequence, result\_2 – second temporary outcome sequence.

```

1st step      – cccccchhhhhhhhhhhhhhhcccccccccccccccc
2nd step     – cccttcchhhhhhhhhheeeeeccccchhhhhhhc

result_1     – cccttcchhhhhhhhhhh----eecccc-----c
complement  – .....hhee.....cchhhcc.
result_2     –      cccttcchhhhhhhhhhhheeeeeccccchhhhhcc

```

**Fig. 4.** The scheme of basics of the mechanism of beta sheet and turn regions determination by the Aver(TM+SecStr) algorithm; c – coil, h – helix, t – turn, e – beta sheet, dash – moot region, dot – primary defined structure, result\_1 – first temporary outcome sequence, result\_2 – second temporary outcome sequence.

- Positions assigned to the coil in the first step and assigned to the beta sheet in the second step finally are defined as beta sheet (Fig. 4).
- Positions assigned to the coil in the first step and assigned to the turn in the second step finally are defined as turn (Fig. 4).
- A single beta sheet region should not be shorter than 3 residues.
- Positions labeled as a helix in the first step and labeled as a beta sheet in the second step are defined as a „helix/sheet mood region” on the first temporary output sequence (Fig. 4).
- Each “helix/sheet mood region” is analyzed in accordance with restrictions no. 3-9 with changed coils for beta sheet symbols (Fig. 4).
- If “helix/sheet mood region” consists of odd number of residues and around this region (-10 position from the beginning of analyzed region and +10 position from the end of analyzed region) minimum one position of another beta sheet region exists, in this “mood region” beta sheet occupies half a number of positions round up to the nearest integer value.
- Regions predicted as helix in the first step and predicted as turn in the second step are defined as a „helix/turn mood region” on the first temporary output sequence; each “helix/turn mood region” is analyzed in accordance with restrictions no. 3-9 with changed coils for turns symbols (Fig. 4).

Five models of transmembrane proteins have been used to design and verify the SSTMProt algorithm. These are the models of three first subunits of bovine heart cytochrome c oxidase (1OCC.pdb and 1V54.pdb) at fully oxidized state (cox1, cox2 and cox3 chains A, B and C), rhodopsin (1U19.pdb) and transmembrane subunit of bacterial ATPase (1C17.pdb) (equivalent of human subunit ATP6). Additionally, the algorithm has been verified using a few globular proteins – three models of subunits of human hemoglobin (1HBA.pdb, 1HBB.pdb and 1HBS.pdb):  $\alpha$  and  $\beta$  subunit of a native form of protein and  $\beta$  subunit of mutated hemoglobin S. Secondary and tertiary structure of mutated and native form of human hemoglobin is known and precisely described in RCSB PDB. Using these models it is possible to verify sensitivity of elaborated algorithm and methodology on a single amino acidic substitution in polypeptide chain.

All the models used to the algorithm verification have been downloaded from the RCSB PDB:

- the first subunit of Cox: 1OCC\_A.pdb[29] and 1V54\_A.pdb [30]
- the second subunit of Cox: 1OCC\_B.pdb[29] and 1V54\_B.pdb [30]
- the third subunit of Cox: 1OCC\_C.pdb[29] and 1V54\_C.pdb [30]
- “M” subunit of C12 subcomplex of F1FO ATP synthase: 1C17\_M.pdb [31]
- “A” subunit of bovine rhodopsin: 1U19\_A.pdb [32]
- alpha subunit of native form of human Hemoglobin: 1HBB\_A.pdb [33] and 1HBA\_A.pdb [33]
- beta subunit of native form of human Hemoglobin: 1HBB\_B.pdb and 1HBA\_B.pdb
- beta subunit of human Hemoglobin S: 1HBS\_B.pdb [34]
- models of human subunits of cox have been prepared using homology modeling method; 1OCC.pdb has been used as a template. The similarity between the sequences of subunits of bovine protein (1OCC) used as template and human proteins (GI in the NCBI Protein database: cox1 – 506829, cox2 – 337192, cox3 – 2052365) allows to build precise 3D models of the human proteins:
  - **cox1\_GI:506829 vs. 1OCC\_A.pdb**
  - Score = 906 bits (2342), Expect = 0.0, Method: Compositional matrix adjust. Identities = 468/512 (91%), Positives = 493/512 (96%), Gaps = 0/512 (0%)
  - **cox2\_GI:337192 vs. 1OCC\_B.pdb**
  - Score = 325 bits (832), Expect = 3e-91, Method: Compositional matrix adjust. Identities = 164/220 (74%), Positives = 196/220 (89%), Gaps = 0/220 (0%)
  - **cox3\_GI:2052365 vs. 1OCC\_C.pdb**
  - Score = 466 bits (1199), Expect = 9e-134, Method: Compositional matrix adjust. Identities = 227/260 (87%), Positives = 239/260 (91%), Gaps = 0/260 (0%).

The following tools were used for prediction of the secondary structure of protein, and its transmembrane regions and physicochemical parameters:

- Prediction of transmembrane segments:
1. SPLIT Server (<http://split.pmfst.hr>)
  2. SOSUI (<http://bp.nuap.nagoya-u.ac.jp/sosui/>) [27]
  3. Tmpred 1 (<http://www.ch.embnnet.org>)
  4. METApp (<http://cubic.bioc.columbia.edu>)
  5. hmmtop (<http://www.enzim.hu/hmmtop/>)

6. TOPpred (<http://bioweb.pasteur.fr>)
7. Antheptrot (<http://antheptrot-pbil.ibcp.fr/>) [14]
8. BPROMPT (<http://www.jenner.ac.uk>)

Prediction of secondary structures:

1. Predator (<http://www.bioweb.pasteur.fr>)
2. Predict7 [35]
3. Garnier – Antheptrot (<http://antheptrot-pbil.ibcp.fr/>)
4. Gibrat – Antheptrot (<http://antheptrot-pbil.ibcp.fr/>)
5. DPM – Antheptrot (<http://antheptrot-pbil.ibcp.fr/>)
6. Levis – Antheptrot (<http://antheptrot-pbil.ibcp.fr/>)
7. JPred-jhmm (<http://www.compbio.dundee.ac.uk>)
8. JPred-jpred (<http://www.compbio.dundee.ac.uk>)
9. JPred-jnet (<http://www.compbio.dundee.ac.uk>)
10. JPred-jfreq (<http://www.compbio.dundee.ac.uk>)
11. JPred-jpssm (<http://www.compbio.dundee.ac.uk>)
12. JPred-jalign (<http://www.compbio.dundee.ac.uk>)
13. UCSC (<http://www.soe.ucsc.edu>)
14. PsiPred (<http://bioinf.cs.ucl.ac.uk>)
15. ROSS DONALD KING (<http://npsa-pbil.ibcp.fr>)
16. Antheptrot/SOPMA (<http://npsa-pbil.ibcp.fr/>) [14]
17. PAT – SIPMA96 (<http://bioserv.cbs.cnrs.fr>)
18. SOPM (<http://searchlauncher.bcm.tmc.edu>)
19. PAT-dsc (<http://bioserv.cbs.cnrs.fr>)
20. NNpredict – none class (<http://www.cmpharm.ucsf.edu>)

Analysis of physicochemical parameters of sequences:

1. Antheptrot [14]
2. JalView [36]
3. BioEdit [37]
4. Predict7 [35]

The source codes were written in C++, Turbo Pascal and Delphi. NAMD [38], VMD [39], SwissPDBViewer and SwissModel [40] have been used for precise analysis of proteins structure and to compare numerous 3D models simultaneously (calculation of RMSD during models optimization). The input files and author's own protocols have been written using the script languages of the above programs and have been coded in Object Pascal.

RasMol [41], DeepView [40], PyMOL [42], VMD [39], Antheptrot3D [14], PovRay [43] programs have been used to visualize the models and to prepare high-quality pictures.

SwissPDBViewer and SwissModel [40], BLAST [44], ClustalW [45] tools have been used to homology modeling and threading.

## Results

Two features of the *SSTMProt* algorithms have been tested: accuracy and sensitivity. The accuracy has been presented as a percentage value of similarity between predicted structure and native structure (the accuracy is 100% when a sequence of predicted secondary structure is identical with true secondary structure of protein). The sensitivity has been calculated and presented as a percentage value of similarity between predicted secondary structure of a native form of protein and predicted secondary structure of a protein with a single amino acidic substitution.

The *SSTMProt* combines minimum two different types of secondary structure prediction methods (in this research: transmembrane region prediction methods and simple secondary structure prediction methods). The following features have been analyzed during the selection of available and known secondary structure prediction methods: accuracy of prediction, repeatability, simplicity of algorithm (more simple algorithm is easier to recode it, to implement into its own tool and easier to adapt to any new restrictions), independency of any homology modeling methods and other methods based on the analysis of known protein 3D structures (homology modeling is insensitive to any single and even small amino acidic mutations), accessibility of an algorithm source code (it was necessary to implement efficiently of the code into its own programs). Finally, twelve combinations of tested known methods have been selected to the next research:

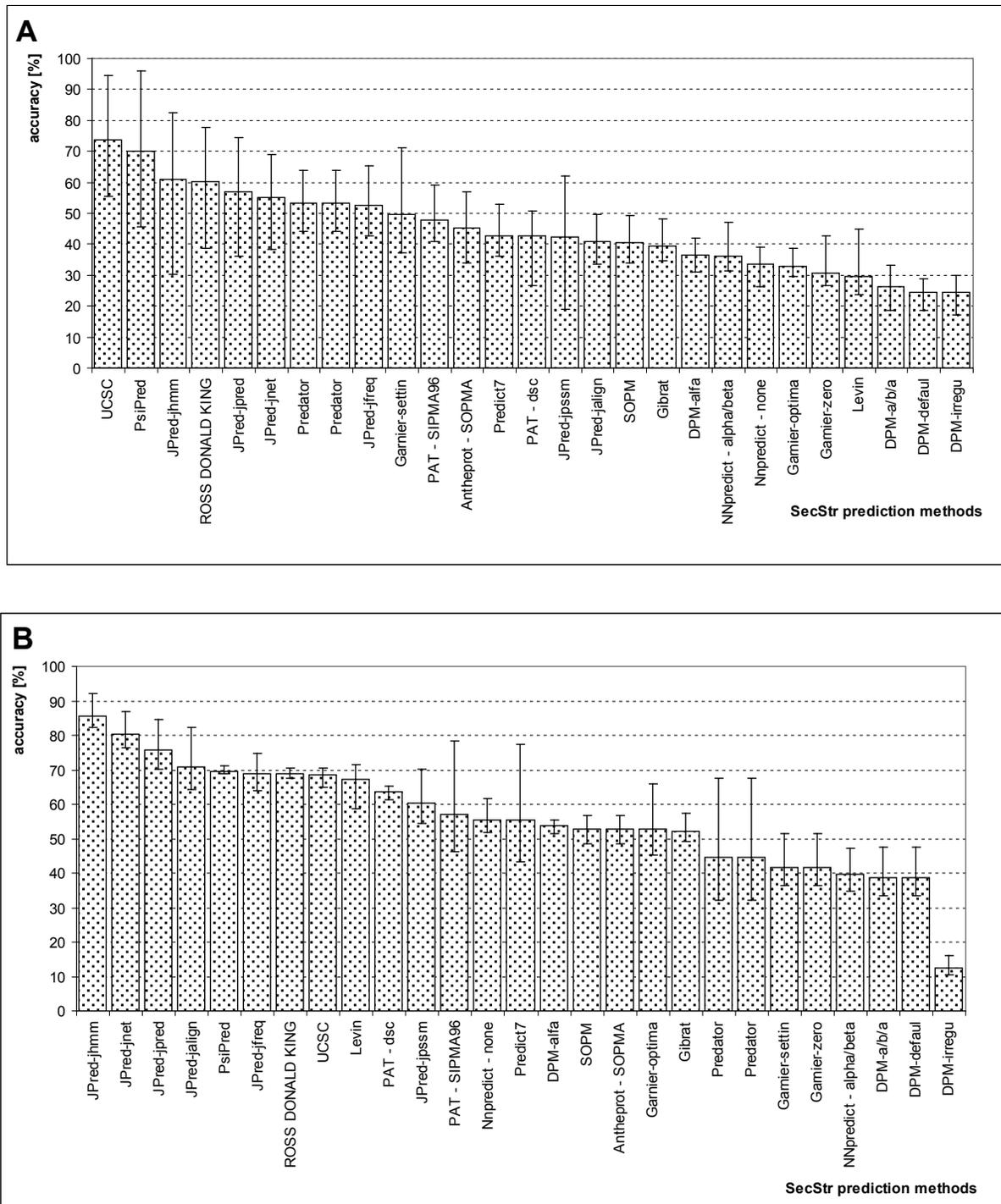
1. Split and JPred
2. Split and PsiPred
3. Split and UCSC
4. Split and Garnier
5. SOSUI and JPred
6. SOSUI and PsiPred
7. SOSUI and UCSC
8. SOSUI and Garnier
9. TMpred and JPred
10. TMpred and PsiPred
11. TMpred and UCSC
12. TMpred and Garnier

The accuracy tests of the *SSTMPred* algorithms have been done for five transmembrane proteins (cox1, cox2, cox3 and subunit of F0 part of ATP synthase) and three globular proteins (alpha and beta subunit of the human hemoglobin and beta subunit of human mutated hemoglobin S). Original secondary structures of analyzed proteins (the data from RCSB PDB) have been aligned with secondary structures of the same proteins predicted by *SSTMPred* algorithm. The results of this alignment have been presented as a percentage value of a rate of identical positions and all positions in the sequence.

The results of accuracy of known tested methods have been arranged in two groups: the group of transmembrane proteins – cox, F0 and rhodopsin (Fig. 5A) and the group of alpha-helical globular proteins – subunits of human hemoglobin (Fig. 5B). Each value is the average accuracy of given method for the whole group of proteins with maximum and minimum value.

The results of accuracy of designed methods have been presented as two groups of graphs (transmembrane and globular proteins), where each group consist of the separate graphs for each protein (transmembrane: cox1, cox2, cox3, F0, rhodopsin; globular: alpha subunit, beta subunit, beta subunit of hemoglobin S) (Fig. 6-7). Each graph contains values of average, maximum and minimum accuracy of six types of methods: three algorithms of *SSTMProt*, all tested known methods ('tested available'), known methods selected for next research ('selected available') and methods to predict transmembrane regions.

The accuracy of well-known transmembrane region prediction methods is much higher than the accuracy of secondary structure prediction methods, because this measurement has been done only for helices, excluding other secondary structures.

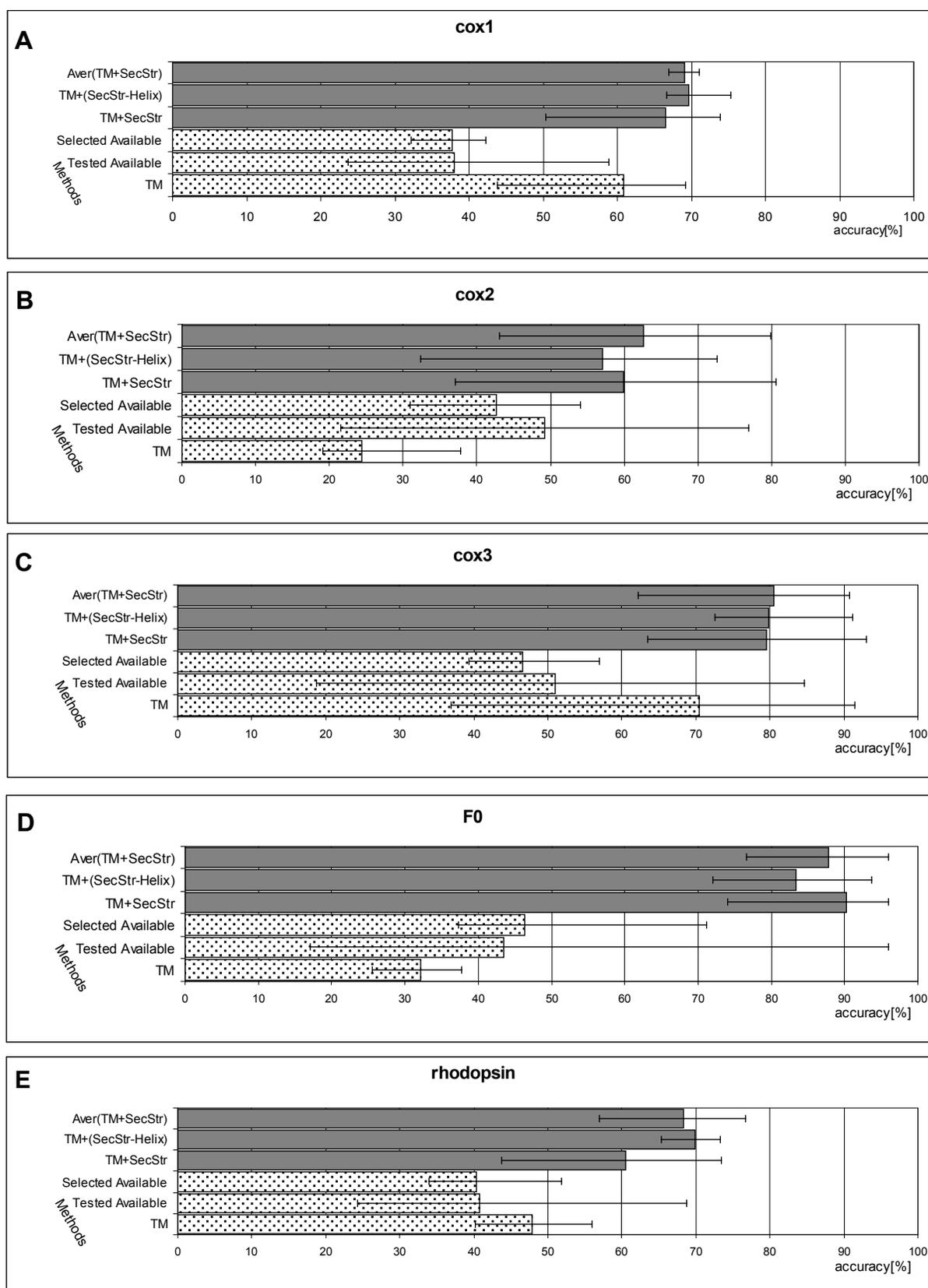


**Fig. 5.** The average, maximum and minimum accuracy of all tested methods of protein secondary structure prediction. The tests have been done for eight proteins of known structure: A – transmembrane proteins (I, II and III subunit of cox, bacterial subunit F0 and rhodopsin) and B – globular proteins (alpha and beta subunit of hemoglobin and beta subunit of hemoglobin S). The results have been set in accordance with average accuracy value form the highest to the lowest value.

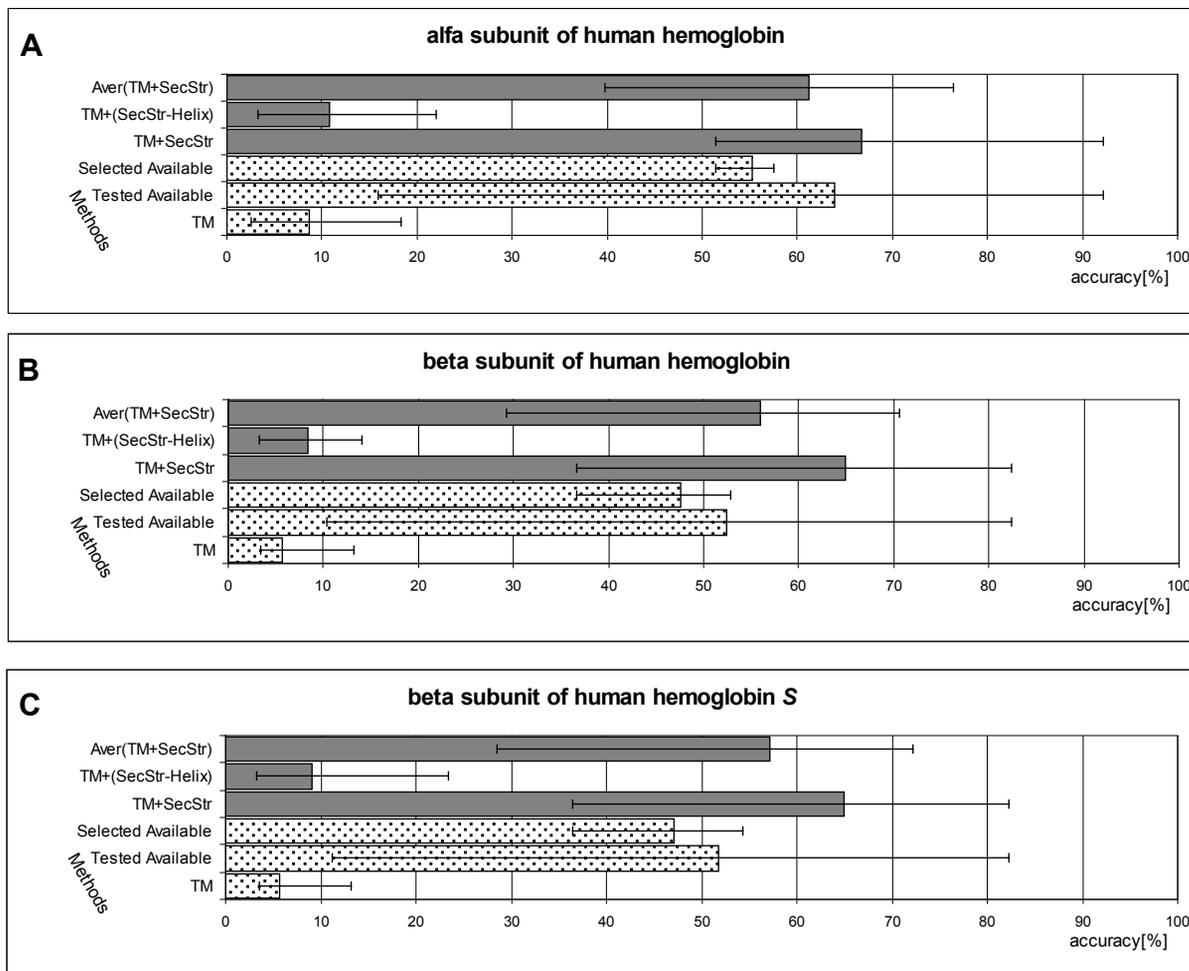
Results of the *SSTMProt* algorithm predictions of both models of  $\beta$  subunit of human hemoglobin (native form and hemoglobin S) have been compared in this research. These models have been selected to the analysis to test the accuracy of designed method for globular, alpha-helical proteins and to test the sensitivity of designed algorithms on a single amino acidic substitution in polypeptide chain (beta chain of

hemoglobin S contains the single amino acid substitution; this mutation causes stickle cell anemia).

Calculated percentage values of similarity between predicted (using the *SSTMProt* algorithms) secondary structure and a native form and mutated protein (hemoglobin S) have been collected in the Tab. 1. These results are presented in order of combinations of known predicted methods.



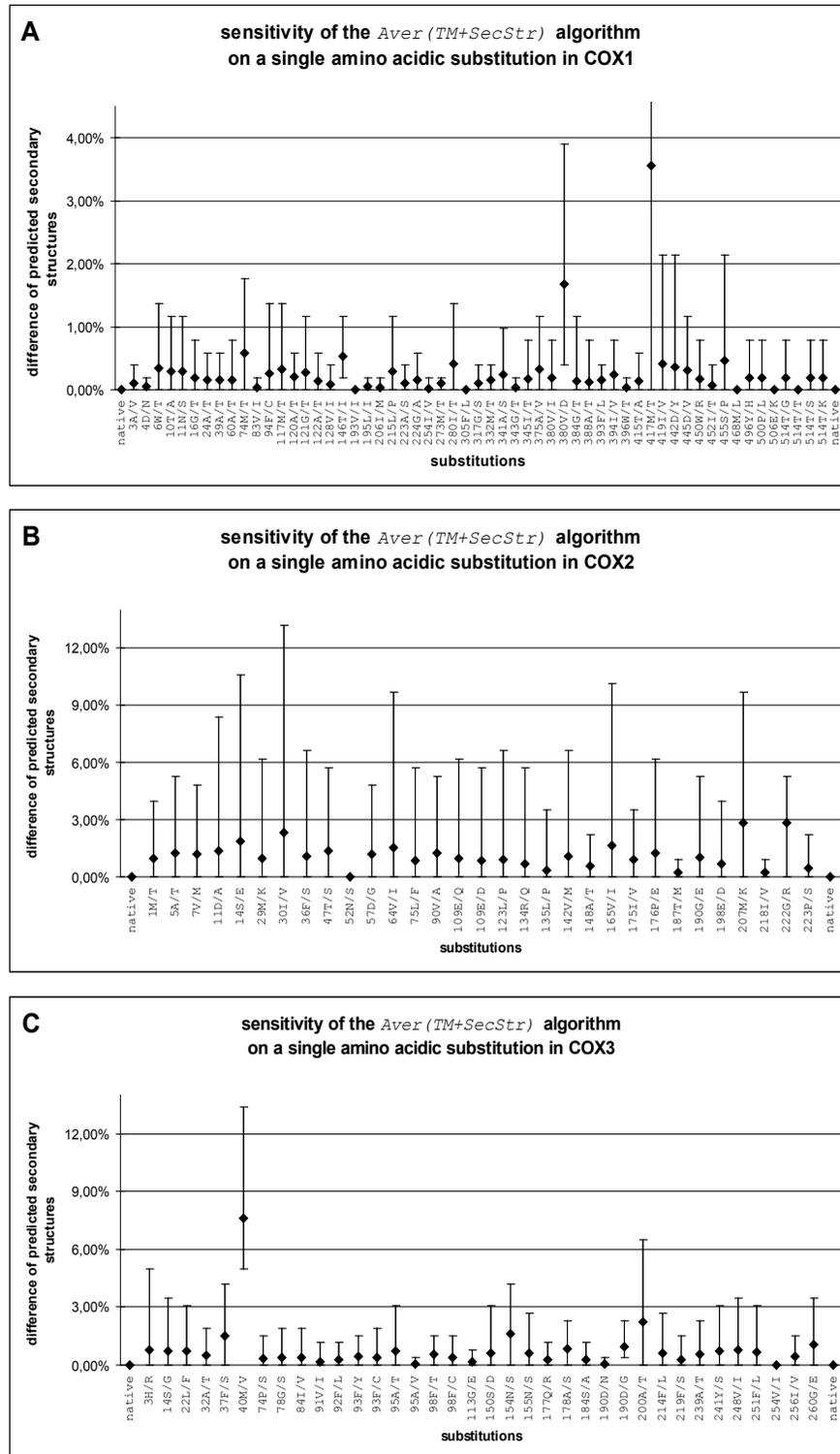
**Fig. 6.** The comparison of average accuracy of developed algorithms (SSTMProt: TM+SecStr, TM+(SecStr-helix), Aver(TM+SecStr)) with average accuracy of prediction by all tested known methods for five transmembrane proteins of known structure (I, II and III subunit of cytochrome c oxidase (graphs: A – cox1, B – cox2 and C – cox3), graph D – bacterial F0 and graph E – rhodopsin). The results has been presented as percentage values of similarity between the result of prediction and real structure. The results of predictions using known methods have been presented in three groups: “selected available” – average accuracy of all methods selected to twelve studied combinations, “tested available” – average accuracy of all tested methods, “TM” – average accuracy of transmembrane region prediction by all tested methods.



**Fig. 7.** The comparison of average accuracy of developed algorithms [SSTMProt: TM+SecStr, TM+(SecStr-helix), Aver(TM+SecStr)] with average accuracy of prediction by all tested known methods for three globular proteins of known structure [alpha (A) and beta (B) subunit of native form of hemoglobin and beta subunit of hemoglobin S (C)]. The results has been presented as percentage values of similarity between the result of prediction and real structure. The results of predictions using known methods have been presented in three groups: “selected available” – average accuracy of all methods selected to twelve studied combinations, “tested available” – average accuracy of all tested methods, “TM” – average accuracy of transmembrane region prediction by all tested methods.

**Tab. 1.** Similarity of original and predicted secondary structures of two forms of human hemoglobin: beta subunit of a native form (1HBB\_B PDB model) and mutated form – hemoglobin S (1HBS\_B PDB model). Combinations of used methods has been listed in column “combinations”. The difference between predicted secondary structure of both forms of hemoglobin proofs enough sensitivity of designed algorithm. The SSTMProt algorithm has been tested in 12 combinations of known methods. Calculated identity has been presented as a percentage values (“Percentage similarity” column) and numerical complicity (“SecStr pos.” column) of identical positions in compared sequences.

Combinations of methods	Secondary structure identity between native form of human hemoglobin and S mutant	
	SecStr pos. [ident/all]	Percentage similarity[%]
Native PDB Hbb_b vs. Hbs_b	128/146	87,67%
Split+Jpred	145/146	99,32%
Split+UCSC	145/146	99,32%
Split+PsiPred	145/146	99,32%
Split+Garnier	145/146	99,32%
SOSUI+Jpred	145/146	99,32%
SOSUI+UCSC	145/146	99,32%
SOSUI+PsiPred	145/146	99,32%
SOSUI+Garnier	144/146	98,63%
TMpred+Jpred	145/146	99,32%
TMpred+UCSC	145/146	99,32%
TMpred+PsiPred	145/146	99,32%
TMpred+Garnier	144/146	98,63%



**Fig. 7.** The results of sensitivity of designed algorithms to detect single amino acidic substitution. The results have been presented as percentage differences between secondary structures of predicted native form and predicted substituted form. Each substituted position has been presented separately. All samples have been predicted using *Aver*(TM+SecStr) algorithm. The range of values mean the lowest and the highest difference between predicted structures (each substituted form has been predicted twelve times – twelve combinations of methods). A – results for cox1, B – cox2, C – cox3.

The tests of sensitivity of the *SSTMProt* algorithms have been done for a few transmembrane proteins of a very conservative structure (mtDNA coding subunits of bovine heart cytochrome c oxidase: *cox1*, *cox2*, *cox3*). SNP positions collected in HmtDB [46] and MITOMAP [47] databases have been used in tests as a point mutations.

The temporary results of low sensitivity and very simple known methods (described above) have been analyzed using the *SSTMProt* algorithm. Both, high repeatability of final results of tested algorithms and insignificant influence of SNP's on three dimensional structure of proteins allow to establish the lowest criteria of positive results (at least the difference of only one position between original and predicted secondary structure means a positive result of the sensitivity test – the success of the sensitivity test). The following combinations of used methods have been done in this test:

1. Split and DPM
2. Split and Garnier
3. Split and Gibrat
4. Split and SOPMA
5. SOSUI and DPM
6. SOSUI and Garnier
7. SOSUI and Gibrat
8. SOSUI and SOPMA
9. Tmpred and DPM
10. Tmpred and Garnier
11. Tmpred and Gibrat
12. Tmpred and SOPMA

All information about the SNP's of the transmembrane proteins has been taken from HmtDB [46] and MITOMAP [47]. This sensitivity test was very restrictive because single nucleotide polymorphisms occur quite often and should not have any significant influence on the final protein structure. The results of the sensitivity tests have been presented on the Fig. 5. The values define differences between predicted by *SSTMProt* algorithm secondary structures of both native form defined by Anderson sequence or the Cambridge Reference Sequence [48] and mutated form (with single amino acidic substitution). The values are the percentage differences between predicted structure for the sequence with mutation and predicted structure of Anderson sequence (the sequence used as the wild form).

## Discussion

The results presented on the Fig. 6-7 demonstrate that the accuracy of prediction of the secondary structure of selected transmembrane proteins using the *SSTMProt* algorithms is higher than using known, commonly used methods. The *SSTMProt* algorithm has been designed on the basis of the general stages of natural folding process of the transmembrane proteins and basing on the methods of prediction of the transmembrane regions. That is why the accuracy of prediction is very high for proteins an-

chored in the membrane. This algorithm has to be adapted to predict the structure of globular proteins. This "reconstruction" mainly concerns the module of simulation of the helices folding process (the low values of accuracy on the charts 7A, B and C for the *TM+(SecStr-Helix)* algorithm). In this case the accuracy for *cox2* has the lowest value because the most important and a large domain of the protein (there is the cytochrome c binding place) in the globular part occurs in the aquatic environment (inter membrane space) out of the inner mitochondrial membrane. The secondary structure of this part in the main part is  $\beta$ -sheet.

More precise prediction and better repeatability of results for designed algorithms are clearly presented on the Fig. 5-7. The maximum deviations of accuracy for *SSTMProt* reaches about 15%, while the same parameter for known methods for the same group of proteins reaches about 50% (Fig. 6-7 – 'Tested Available'). The lowest results abbreviation, the maximum repeatability and the regular high accuracy have been received using the *Aver(TM+SecStr)* algorithm. It proves that this algorithm has been developed using the most accurate biological properties among all three algorithms (*TM+SecStr*, *TM+(SecStr-Helix)* and *Aver(TM+SecStr)*). That is why the results of the two first algorithms (*TM+SecStr*, *TM+(SecStr-Helix)*) have been used as references to design and develop the *Aver(TM+SecStr)* algorithm.

The first tests of algorithm sensitivity have been done for subunits of human hemoglobin. The results of these tests have been presented in the Tab. 1. The difference between values of results for HBB\_B and HBS\_B proves that designed algorithm is sensitive enough to detect an influence of a single amino acidic substitution on a final protein structure. In this case estimated accuracy deviation is much higher (approximately 40% – stages 10-12) than the real influence of a single amino acidic substitution on a structure of a native form of analysed protein (12,3% for downloaded models of native form – 1HBB\_B and mutated 1HBS\_B). These tests show only sensitivity or insensitivity of tested algorithm and the values are not adequate to the real level of secondary structure changes.

More precise sensitivity tests (Fig. 8A, B and C) show that the *Aver(TM+SecStr)* algorithm is very sensitive for single substitution in analyzed chain. High average accuracy, relatively low deviations results and high repeatability of values of *Aver(TM+SecStr)* algorithm results encourage to further research and development presented algorithm.

At present our team are working to develop a current version of an algorithm and to build online tool to analyze primary structure and to predict secondary structure of globular and transmembrane proteins.

## Acknowledgement

Dr Jan Pawel Jastrzebski is very grateful to Prof. Marcella Atimonelli for possibility of making this kind of research and a lot of patient and anonymous referees for criticism and suggestion.

## References

- Weissman J.S., Kim P.S.: Reexamination of the folding of BPTI: predominance of native intermediates. *Science* 1991, 253(5026): 1386-1393.
- Radford S.E., Dobson C.M., Evans P.A.: The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature* 1992, 358(6384): 302-307.
- Kanaya E., Ishihara K., Tsunasawa S., Nokihara K., Kikuchi M.: Indication of possible post-translational formation of disulphide bonds in the beta-sheet domain of human lysozyme. *Biochem. J.* 1993, 292(2): 469-476.
- Clarke D.T., Doig A.J., Stapley B.J., Jones G.R.: The alpha-helix folds on the millisecond time scale. *Proc. Natl. Acad. Sci. USA* 1999, 96(13): 7232-7237.
- Protein Structure Prediction, [http://cmgm.stanford.edu/WWW/www\\_predict.html](http://cmgm.stanford.edu/WWW/www_predict.html).
- Hovijitra N.: Computational methods for predicting transmembrane alpha helices. Computational Molecular Biology Final Project. Stanford University, Biochemistry BioMedical Informatics, 2002, 218-231.
- Frishman D., Argos P.: Seventy-five percent accuracy in protein secondary structure prediction. *Proteins: Structure, Function, and Genetics* 1997, 27(3): 329-335.
- Garnier J., Osguthorpe D.J., Robson B.: Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *Journal of Molecular Biology* 1978, 120: 97-120.
- Garnier J., Gibrat J.F., Robson B.: GOR method for predicting protein secondary structure from amino acid sequence. *Methods in Enzymology* 1996, 266: 540-553.
- Gibrat J.F., Garnier J., Robson B.: Further developments of protein secondary structure prediction using information theory. *Journal of Molecular Biology* 1987, 198: 425-443.
- Sen T.Z., Jernigan R.L., Garnier J., Kloczkowski A.: GOR V server for protein secondary structure prediction. *Bioinformatics* 2005, 21(11): 2787-2788.
- Kloczkowski A., Ting K.L., Jernigan R.L., Garnier J.: Combining the GORV Algorithm With Evolutionary Information for Protein Secondary Structure Prediction From Amino Acid Sequence. *Proteins: Structure, Function and Genetics* 2002, 49: 154-166.
- Thompson M.J., Goldstein R.A.: Predicting protein secondary structure with probabilistic schemata of evolutionarily derived information. *Protein Science* 1997, 6(9): 1963-1975.
- Deléage G., Combet C., Blanchet C., Geourjon C.: ANTHE-PROT: An integrated protein sequence analysis software with client/server capabilities. *Computers in Biology and Medicine* 2001, 31(4): 259-267.
- Kyte J., Doolittle R.F.: A simple method for displaying the hydrophobic character of a protein. *Journal of Molecular Biology* 1982, 157: 105-132.
- von Heijne G.: Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *Journal of Molecular Biology* 1992, 8: 249-254.
- Graff D.K., Pastrana-Rios B., Venyaminov S.Y., Prendergast F.G.: The Effects of Chain Length and Thermal Denaturation on Helix-Forming Peptides: A Mode-Specific Analysis Using 2D FT-IR. *Journal of the American Chemical Society* 1997, 119(46): 11282-11294.
- Klein C.T., Mayer B., Kohler G., Wolschann P.: Influence of solvation on helix formation of poly-alanine studied by multiple annealing simulations. *Journal of Molecular Structure: THEOCHEM* 1996, 370(1): 33-43.
- Hendry P., Mccall M.J., Lockett T.J.: Influence of helix length on cleavage efficiency of hammerhead ribozymes. *Australian Journal of Chemistry* 2005, 58(12): 851-858.
- Bertaud J., Hester J., Jimenez D.D., Buehler M.J.: Energy landscape, structure and rate effects on strength properties of alpha-helical proteins. *Journal of Physics: Condensed Matter* 2010, 22(3): 035102.
- Stryer L.: Biochemia. Warszawa: Wydawnictwo Naukowe PWN, 1999.
- Bowie J.U.: Helix packing in membrane proteins. *J. Mol. Biol.* 1997, 272: 780-789.
- Honnappa S., Gouveia S.M., Weisbrich A., Damberger F.F., Bhavesh N.S., Jawhari H., Grigoriev I., van Rijssel F.J.A., Buey R.M., Lawera A., Jelesarov I., Winkler F.K., Wuthrich K., Akhmanova A., Steinmetz M.O.: An EB1-Binding Motif Acts as a Microtubule Tip Localization Signal. *Cell (Cambridge, Ma)* 2009, 138: 366-376.
- Yang Y., Gourinath S., Kovacs M., Nyitray L., Reutzler R., Himmel D.M., O'Neill-Hennessey E., Reshetnikova L., Szent-Gyorgyi A.G., Brown J.H., Cohen C.: Rigor-like structures from muscle myosins reveal key mechanical elements in the transduction pathways of this allosteric motor. *Structure* 2007, 15: 553-564.
- Su J.Y., Hodges R.S., Kay C.M.: Effect of Chain Length on the Formation and Stability of Synthetic alpha-Helical Coiled Coils. *Biochemistry* 1994, 33(51): 15501-15510.
- Lau S.Y., Taneja A.K., Hodges R.S.: Synthesis of a model protein of defined secondary and quaternary structure. Effect of chain length on the stabilization and formation of two-stranded alpha-helical coiled-coils. *The Journal of Biological Chemistry* 1984, 259: 13253-13261.
- Hirokawa T., Boon-Chieng S., Mitaku S.: SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* 1998, 14: 378-379.
- Mitaku S., Hirokawa T.: Physicochemical factors for discriminating between soluble and membrane proteins: hydrophobicity of helical segments and protein length. *Protein Engineering* 1999, 11: 953-957.
- Tsukihara T., Aoyama H., Yamashita E., Tomizaki T., Yamaguchi H., Shinzawa-Itoh K., Nakashima R., Yaono R., Yoshikawa S.: The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 Å. *Science* 1996, 272: 1136-1144.
- Tsukihara T., Shimokata K., Katayama Y., Shimada H., Muramoto K., Aoyama H., Mochizuki M., Shinzawa-Itoh K., Yamashita E., Yao M., Ishimura Y., Yoshikawa S.: The low-spin heme of cytochrome c oxidase as the driving element of the proton-pumping process. *Proc. Natl. Acad. Sci. USA* 2003, 100: 15304-15309.
- Rastogi V.K., Girvin M.E.: Structural changes linked to proton translocation by subunit c of the ATP synthase. *Nature* 1999, 402: 263-268.

32. Okada T., Sugihara M., Bondar A.N., Elstner M., Entel P., Buss V.: The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *J. Mol. Biol.* 2004, 342: 571-583.
33. Kavanaugh J.S., Rogers P.H., Case D.A., Arnone A.: High-resolution X-ray study of deoxyhemoglobin Rothschild 37 beta Trp---Arg: a mutation that creates an intersubunit chloride-binding site. *Biochemistry* 1992, 31: 4111-4121.
34. Padlan E.A., Love W.E.: Refined crystal structure of deoxyhemoglobin S. I. Restrained least-squares refinement at 3.0-Å resolution. *J. Biol. Chem.* 1985, 260: 8272-8279.
35. Cármenes R.S., Freije J.P., Molina M.M., Martín J.M.: Predict7, a program for protein structure prediction. *Biochem. Biophys. Res. Commun.* 1981, 159(2): 687-693.
36. Clamp M., Cuff J., Searle S.M., Barton G.J.: The Jalview Java Alignment Editor. *Bioinformatics* 2004, 20: 426-427.
37. Hall T.A.: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* 1999, 41: 95-98.
38. Phillips J.C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R.D., Kale L., Schulten S.: Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* 2005, 26: 1781-1802.
39. Humphrey W., Dalke A., Schulten K.: VMD – Visual Molecular Dynamics. *J. Molec. Graphics* 1996, 14: 33-38.
40. Guex N., Peitsch M.C.: SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 1997, 18: 2714-2723.
41. Sayle R., Milner-White E.J.: RasMol: Biomolecular graphics for all. *Trends in Biochemical Sciences* 1995, 20(9): 374.
42. The PyMOL Molecular Graphics System, <http://www.pymol.org>.
43. PovRay: Persistence of Vision Raytracer, <http://www.povray.org>.
44. Altschul S.F., Madden T.L., Schaffer 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.* 1997, 25: 3389-3402.
45. Thompson J.D., Higgins D.G., Gibson T.J.: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994, 22: 4673-4680.
46. Attimonelli M., Accetturo M., Santamaria M., Lascaro D., Scioscia G., Pappada G., Russo L., Zanchetta L., Tomaseo-Ponzetta M.: HmtDB, a Human Mitochondrial Genomic Resource Based on Variability Studies Supporting Population Genetics and Biomedical Research. *BMC Bioinformatics* 2005, 6(4): S4.
47. Ruiz-Pesini E., Lott M.T., Procaccio V., Poole J., Brandon M.C., Mishmar D., Yi C., Kreuziger J., Baldi P., Wallace D.C.: An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nucleic Acids Res.* 2007, 35: D823-D828.
48. Anderson S., Bankier A.T., Barrell B.G., de Bruijn M.H., Coulson A.R., Drouin J., Eperon I.C., Nierlich D.P., Roe B.A., Sanger F., Schreier P.H., Smith A.J., Staden R., Young I.G.: Sequence and organization of the human mitochondrial genome. *Nature* 1981, 290: 457-465.

## ANALYSIS OF DATA GENERATED BY INSULIN PUMP USING LABVIEW ENVIROMENT

BARTŁOMIEJ MATEJKO<sup>1,2</sup>

<sup>1</sup> *Multidisciplinary School of Engineering in Biomedicine, AGH- University of Science and Technology, Kraków, Poland;* <sup>2</sup> *Department of Metabolic Diseases, Jagiellonian University Medical College, 15 Kopernika St., Kraków, Poland, b.matejko@yahoo.com*

**Abstract:** Diabetes mellitus is a group of metabolic diseases characterized by impaired insulin secretion or action affecting around 200 million people worldwide. Type 1 diabetes makes up 10% of all diabetes cases. Currently, the most physiological way of insulin delivery involves the use of an insulin pump, which administers insulin in a continuous manner. During an outpatient visit insulin pump data are read and analyzed to optimize treatment and improve patients metabolic control. There are commercial software companies producing insulin pumps which present data in a report format. The aim of this project was to write a program in a LabVIEW environment which would analyze the insulin pump data. The program includes structured bookmarks, statistics on glucose (different systems of measurement), insulin, nutritional information, uses advanced tools in the pump, and presents the entire recorded history of pump data (commercial version statistics use only two week time intervals). Implemented as an innovative feature that compares the blood glucose meter to indicate the continuous glucose monitoring system. The software built for this purpose can be used in diabetes clinics to analyze insulin pump data and to research on the application of new diabetes technology. The current version can be treated as a foundation for possible further expansion of the program functionality.

**Keywords:** diabetes, insulin pump, CareLink Medtronic

### Introduction

In the 21<sup>st</sup> century we observed a dynamic development of medicine, particularly in the fields of new medical devices, therapies, and pharmaceutical drugs. Thanks to development we can live longer than our predecessors. However, many chronic diseases such as hypertension, cancer, or diabetes remain incurable. According to the World Health Organization (WHO) in the year 2000 nearly 171 million people suffered from diabetes and this number continues to grow. Forecasts predict that by 2030 there will be 366 million diabetics worldwide [1, 2]. In Poland, diabetes affects 2,6 million people with half of them being unrecognized or untreated cases [3, 4].

Diabetes is one of the main reasons for disability and untimely death [4]. Good education and the best therapeutic methods can help to avoid diabetic complications and lead to a normal life. Type 1 diabetes makes up 10% of all diabetes cases. Because of the autoimmune destruction of  $\beta$  pancreatic islet cells these patients need to take insulin for the rest of their lives. Two insulin regimens commonly are used: multiple daily injection (MDI) and continuous subcutaneous *insulin infusion* (CSII) with personal insulin pump. Many researches have demonstrated that CSII is more effective than MDI in optimizing glicemic control [5-11].

American Diabetes Association (ADA) and European Association for the Study of Diabetes (EASD) made a common statement that insulin pump therapy is the most physiological method of insulin delivery in children and is recommend to every diabetic who accepts this method of therapy [12, 13].

### The goal and general idea

In Poland there are three main insulin pump producers (Medtronic, Roche and Dana). Every company provide his own software to manage diabetes. This software allows to analyse data only from his pumps. In Poland, the most popular pumps are Medtronic products.

Insulin pump producers possess the ability to save data from pumps and glucometers and to analyse them using their professional software (eg CareLinkProfessional, AccuChek 360, Dana manager) [14, 15, 16]. All available programs provide various options of visualization and analysis of the measurements. The idea was to create practical and helpful software to analyse data from Medtronic pumps (MiniMed 508, Paradigm 712, Paradigm 722, VEO) and to extend the functionality of the commercial program CareLinkProfessional. This software can be used by

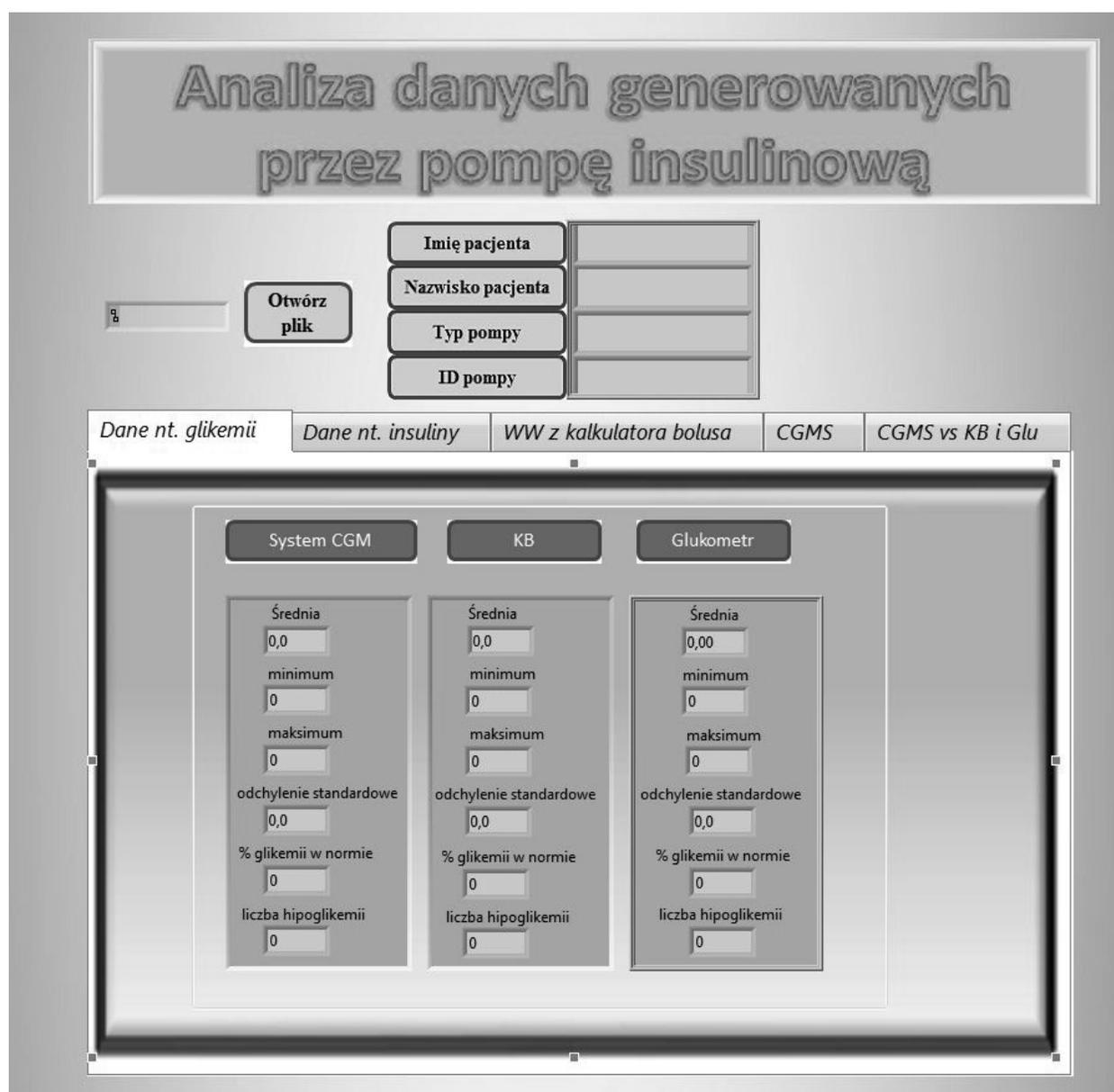
medical doctors to assess patients glycemic control and to find situation when patients make mistakes to avoid it in the future. Additionally, this program can be used by patients to check different parameter levels which will facilitate better diabetes management. This program will also be used by the author for his PhD thesis to obtain parameter statistics and research their connection to patients metabolic control and daily insulin dose. The purpose was also to implement an innovative feature that compares indication of glucometer to the continuous glucose monitoring system.

## Methods

The software was designed for a Windows operating system. The application was created in LabView 2009 from the National

Instruments company. LabVIEW includes the functionality of general-purpose programming languages and is at the same time easy to use. LabVIEW environment also allows the user to construct a variety of applications, starting from its design through prototyping and further development. It contains many useful components ready for collection, analysis, presentation, and reporting of the obtained data. It also has extensive function library for most programming tasks [17, 18]. Its greatest advantage is the ability to compile code fragments in real time during application development [19]. LabVIEW provides control measures built program (debugger) during application development, set interrupts, tracking the transition of signals.

Downloaded data from insulin pump have different unit. The blood glucose level is expressed in mg/dL, Basal, Bolus Insulin in U. The state of the too low concentration of glucose in the blood is called the hypoglycemia (threshold is set arbitrary on



**Fig. 1.** Main page of the “Insulin Pump Analyser” application with patients information and bookmarks which presents relevant data (glucose values, insulin values, exchanges from Bolus Calculator, CGM graph and function that compares glucose values from CGMS to Glucometer and to Glucose from Bolus Calculator).

55 mg/dL) and a state of the increased blood glucose level is a hyperglycemia (threshold set on 180 mg/dL).

## Results and discussion

The created application is based on data with extinction \*.csv (Comma Separated Values) generated by the software Medtronic CareLink Professional. The program consists of several basic functions such as reading data from a file, presentation of patient data and presents medical information about the insulin pump including (Fig. 1):

- daily insulin dose,
- amount of insulin given as a bolus on a daily basis,
- what percentage of all boluses was normal type,
- what percentage of all boluses bolus was with a function bolus calculator (KB),
- amount of carbohydrate exchangers per day obtained from the function KB,
- glucose from the meter, the bolus calculator, and from a CGMS (Continuous Glucose Monitoring System),
- CGMS glucose measurements presented in a graph.

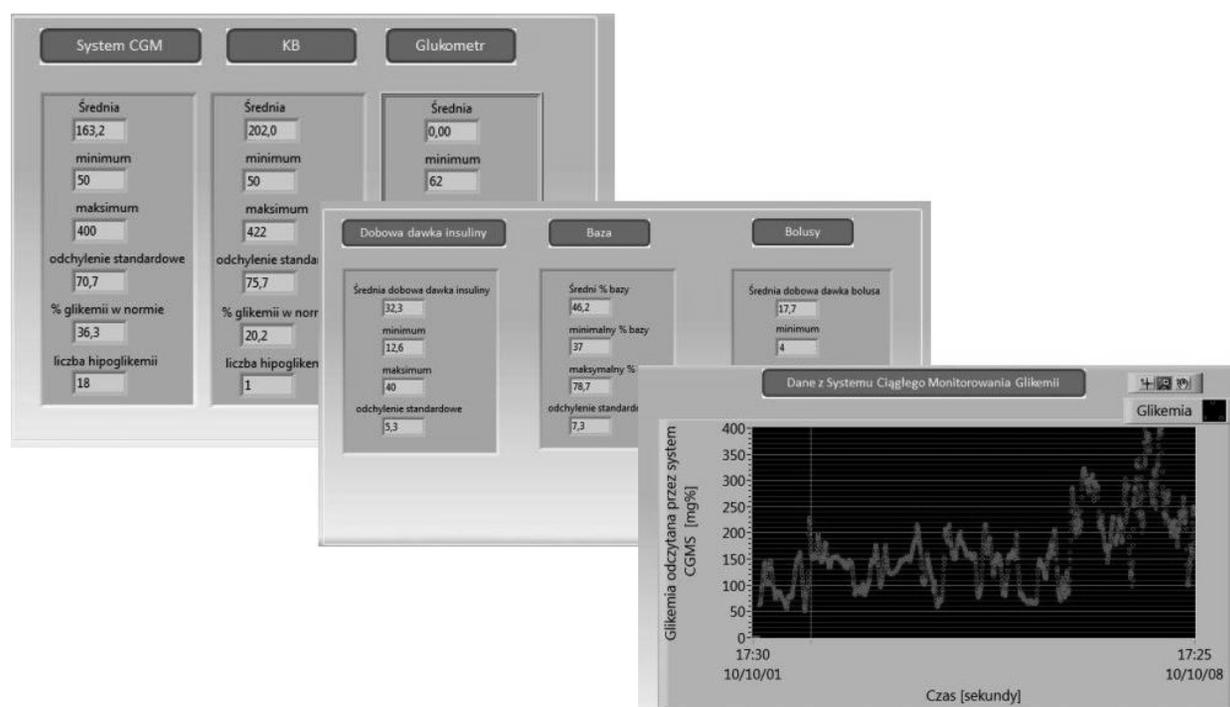
Analysis lasts a few seconds. Exemplary patients data is presented in Fig. 2. To view the compiled data one needs to simply click on an interesting bookmark which presents relevant data in the form of statistics. Other implemented innovative features

include the comparison of CGMS glucose with glucose bolus calculator or with the glucometer and generating corresponding statistics. Built versions of the installation allow the program to run on a computer without National Instruments LabView software.

In order to validate the operation of all functions of the application at every stage, the results were compared with data from a text file (generated from CareLink Professional) using MS Excel and Statistica (Statsoft, Poland). Test was performed on 20 patients files with the extension .csv containing different combinations of data from the Medtronic insulin pump (from pumps: Paradigm 712, Paradigm 722, Veo; glucometers: Contour Ts and Contour Link). No difference was found. A weakness of the program is a lack of an elaborate data structure of the program. However, all implemented functions work as intended. The created program meets the essential assumptions of the project.

Physicians get a complete picture of patient treatment and a means to analyze the current therapies and suggest modifications. Important for the physician shall also be monitored episodes of hypoglycemia and hyperglycemia. By understanding the context of these events it is possible to consider and suggest how to avoid them in future. Another parameter that will allow doctors to guide patient treatment is the number of glucose measurements per day.

In general commercial and presented software can help to save time needed for raw data interpretation. This time can be used by the doctor to search for solutions to optimize therapy.



**Fig. 2.** Bookmarks with exemplary insulin pump data. Exemplary patient has mean glycemia level in CGMS 163mg/dL, 36,3% of measurements was in range, and the number of hipoglikemia was 18 episodes. Minimum, maximum glycemia and standard deviation for mean is shown. Second and the third column are for Bolus Calculator tool data and for glucometer data. In the middle statistics for total daily insulin dose (32,4 U), mean % basal insulin (46,2%), mean daily bolus level (17,7 U) and their standard deviation are presented. The last picture presents data from CGM system. In program, user can zoom in or zoom out parts of graph.

## Conclusions

The "Insulin Pump Analyser" application has been successfully applied in the analysis of data generated from insulin pumps. Its intuitive interface can be easily used by medical doctors and patients. It was implemented in LabView in its educational version. This program allows the processing of statistical data for the entire period remembered by the pump, a major advantage compared to the commercial version where statistics are generated for each week separately. Data presented in this format can be easily interpreted by the user. The obtained results indicate a potential usefulness of the created program for data analysis generated from insulin pumps and the program can prepare easy to understand reports.

The developed application can be considered as the foundation for further improvements. At the moment the application is in the working stage. New functionalities will be added including PDF report production, visualization of more statistics (number of boluses per day, number of self glucose monitoring per day), functions (choice of dates for the analysis) and login functionality. It will also be possible to add a "send mail" option that will inform physicians if life-threatening parameters are measured. The next development stages will be reported in future issues of BAMS.

*This work was part of the Engineer Diploma Project in Multidisciplinary School of Engineering in Biomedicine 2011, AGH-UST*

## References

1. World Health Organization, International Diabetes Federation. Definition and Diagnosis of Diabetes Mellitus and Intermediate Hyperglycemia: Report of a WHO/IDF Consultation. Geneva: World Health Organization Document Production Services, 2006
2. Yach D., Stuckler D., Brownell K.: Epidemiologic and economic consequences of the global epidemics of obesity and diabetes. *Nature Medicine* 2006, 12(1): 62-66.
3. Korzeniowska K., Jabłeczka A.: Cukrzyca (Część I), Diabetes (Part I). *Farmacja Współczesna* 2008, 1: 231-235.
4. 14 listopada – Światowy Dzień Walki z Cukrzycą, [http://naukawpolsce.pap.pl/palio/html.run?\\_Instance=cms\\_naukapl.pap.pl&\\_PageID=1&s=szablon.depesza&dz=szablon.depesza&dep=367789&data=&lang=PL&\\_CheckSum=1213615981](http://naukawpolsce.pap.pl/palio/html.run?_Instance=cms_naukapl.pap.pl&_PageID=1&s=szablon.depesza&dz=szablon.depesza&dep=367789&data=&lang=PL&_CheckSum=1213615981) (2011 Jul 21).
5. Diabetes Control and Complications Trial. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin dependent diabetes mellitus. *N. Engl. J. Med.* 1993, 329: 977-986.
6. Jeitler K., Horvath K., Berghold A. et al.: Continuous subcutaneous insulin infusion versus multiple daily insulin injections in patients with diabetes mellitus: systematic review and meta-analysis. *Diabetologia* 2008, 51(6): 941-51.
7. Pickup J.C., Sutton A.J.: Severe hypoglycaemia and glycaemic control in Type 1 diabetes: meta-analysis of multiple daily insulin injections compared with continuous subcutaneous insulin infusion. *Diabetic medicine: a Journal of the British Diabetic Association* 2008, 25(7): 765-774.
8. Pickup J.K.H.: Continuous subcutaneous insulin infusion at 25 years: evidence base for the expanding use of insulin pump therapy in type 1 diabetes. *Diabetes Care* 2002, 25(3): 593-598.
9. Bruttomesso D., Costa S., Baritussio A.: Continuous subcutaneous insulin infusion (CSII) 30 years later: still the best option for insulin therapy. *Diabetes/metabolism research and reviews* 2009, 25(2): 99-111.
10. Aberle I., Zimprich D., Bach-Kliengel B., Fischer C. et al.: Continuous subcutaneous insulin infusion leads to immediate, stable and long-term changes in metabolic control. *Diabetes, obesity & metabolism* 2008, 10(4): 329-335.
11. Jankovec Z., Hahn M., Grunder S. et al.: Analysis of continuous patient data from the Czech National Register of patients with type 1 and type 2 diabetes using insulin pump therapy. *Diabetes research and clinical practice* 2010, 87(2): 219-223.
12. Phillip M., Battelino T., Rodriguez H. et al.: Use of insulin pump therapy in the pediatric age-group: consensus statement from the European Society for Paediatric Endocrinology, the Lawson Wilkins Pediatric Endocrine Society, and the International Society for Pediatric and Adolescent Diabetes, endorsed by the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes Care* 2007, 30: 1653-1662.
13. ADA Clinical Practice Recommendations, *Diabetes Care* 2000, 23 (suppl.).
14. Care Link Professional, <http://www.medtronic.com/for-healthcare-professionals/products-therapies/diabetes/diabetes-management-software> (2011 Sep 21).
15. Accu Chek 360, <http://www.accu-chek.pl/pl/produkty/zarządzaniem/360.html> (2011 Sep 21).
16. Dana manager, [http://www.sooil.com/NEW/eng/m3\\_ma\\_00.html](http://www.sooil.com/NEW/eng/m3_ma_00.html) (2011 Sep 21).
17. Product Information: What is NI LabVIEW?, <http://www.ni.com/labview/whatis/> (2011 Sep 21).
18. Strona producenta oprogramowania LabView, <http://poland.ni.com/> (2011 Sep 21).
19. Chruściel M.: LabVIEW w praktyce. Legionowo: Wydawnictwo BTC, 2008.

## THE MOTION CAPTURE SYSTEM TO CONDUCT AN ELECTRICAL MUSCLE STIMULATION EXPERIMENT ON UPPER LIMB MOVEMENT

ANDRZEJ SKALSKI<sup>1</sup>, MIROSŁAW SOCHA<sup>1</sup>, ŁUKASZ MALICKI<sup>2,3</sup>, MAREK IWANIEC<sup>3</sup>

<sup>1</sup>AGH University of Science and Technology, Department of Measurement and Instrumentation,  
Al. Mickiewicza 30, Krakow, Poland;

<sup>2</sup>AGH University of Science and Technology, Multidisciplinary School of Engineering In Medicine,  
Al. Mickiewicza 30, Krakow, Poland;

<sup>3</sup>AGH University of Science and Technology, Department of Process Control,  
Al. Mickiewicza 30, Krakow, Poland;

{skalski@agh.edu.pl, socha@agh.edu.pl, lukaszm@agh.edu.pl, iwaniec@agh.edu.pl}

**Abstract:** The article describes a novel method of conducting measurements and analysis of upper limb movement. The system is applied to validate quality of electrical muscle stimulation (EMS). The paper describes the most important parameters of designed equipment and tools. The process of recording images using a standard webcam and the precise description of measurement post, used methods, image processing and hardware calibration are shown. The main goal of the research is to diagnose effects of rehabilitation or training of the muscle biceps brachii.

**Keywords:** EMS, markers tracking, muscle biceps brachii, segmentation

### 1. Introduction

Biceps brachii muscle is a two-headed muscle located on the upper limb. The long head is attached along bicipital groove, while the short head is attached to the coracoid process of scapula. During the movement the muscle contracts and bends the arm at the elbow, therefore the biceps brachii muscle is called 'the flexor'. Moreover the muscle is responsible for rotation of the forearm. Muscle contracts thanks to electric impulses generated by EMS device. The stimulation signal is delivered by 2 electrodes connected to the skin. The parameters of electric impulses: impulse width, intensity and frequency can be modulated. Different values enable to control an angular velocity of the arm movement process.

As the peripheral unit cheap webcam is used. This solution enables to cut down costs what allows to make the proposed solution to be commonly available. The full process is recorded on a specially built measurement stand. The analysis is based on upper limb movement measurement software written in Matlab by the authors. The software estimates markers location on the hand what enables the system to calculate trajectory and angles between the arm and the reference line. Therefore, these parameters can be used to assess stimulation quality.

### 2. The EMS measurement system ideas

The measurements of EMS effectiveness are the most popular verification methods in sport science e.g. [1, 2].

The known method to measure upper limb movement was proposed as a complex system including designed devices. A purpose-built device was developed to measure the isometric strength of wrist extension and the active and passive ranges of motion at the wrist. This device consisted of a fixed base, clamped to a specially designed table, and a moveable arm with a handle at its free end. An accelerometer was coupled to the axis of the moveable arm (to measure angular displacement) and the handle was strain-gauged (to measure passively applied and actively generated moments) [3].

The other method to register upper limb movement requires at least three cameras. This system allows to measure the displacement along all three axis XYZ in Cartesian coordinate system. Limb tremor is registered on the basis of point light source movement. Light source is located on a special wristband placed on a patient's body. It is possible to use one light source or two sources at the same time. Three cameras record muscles' tremor simultaneously [4]. With regard to hand movement in

one plane, it is enough to use only one camera as in solution described in the paper.

In 2002 Porcari et al. [2] published the results of their research related to muscle stimulation. They intended among other things to improve muscle firmness by using EMS. After 8 week-training of using EMS to stimulate the muscles: biceps femoris, quadriceps femoris, biceps brachii muscle, triceps brachii muscle, and muscles of abdomen (rectus abdominus and obliques) the authors made verification tests including measurements of body weight, body fat (via skinfolds), girths, isometric and isokinetic strength (human biceps brachii, triceps brachii muscle, quadriceps femoris muscle, hamstrings), and appearance (via photographs from the front, side, and back). The test based on photographs was not proceeded automatically: Subjects were photographed from the front, side, and back using a digital camera. Men were clothed in a trunk-style swimming suit and women were clothed in a 2-piece swimsuit. All photographs were reviewed and graded for firmness and tone by 1 of the researchers using a 1-10 Leikert-type scale (with 10 being highly firm and toned and 1 being the least firm and toned) [2].

Another example of long term EMS process validation was measurement of isometric force. The isometric force (flexion or extension) was gained by a transducer (Universal Force-Moment Sensor System, JR3, Woodland, CA) mounted in series between the wrist-hand orthosis and a custom-designed support that allowed an adjustment of the arm position for each subject. The force signal was displayed on digital oscilloscopes (model TDS 460A, Tektronix, Pittsfield, MA) and recorded on a digital tape (DAT Sony PC 116, Sony Data Recording, Montvale, NJ). A pulse output from the stimulator was also recorded on tape to indicate the onset of stimulation. The force signal was displayed on two digital oscilloscopes. One oscilloscope was used for visual feedback for the subject including display of the target force. After each trial, the level of force exerted by the subject at the time of stimulation was measured from the oscilloscope display. On the second oscilloscope, the force signal was offset and amplified to increase the resolution of the evoked force [5].

Muscle work could be also checked by recording EMG activity. Electrodes were placed on the mid-belly of the muscle and over the distal tendon. The EMG signal was amplified, band-pass filtered (16-1000 Hz), digitized and sampled (rate 10 kHz) using a CED 1401 interface (Cambridge Electronic Design, Cambridge, UK). Single motor unit activity was obtained using bipolar wire

electrodes (75  $\mu\text{m}$  Teflon-coated stainless steel) inserted into the muscle via a hypodermic needle (23 gauge). The needle was extracted once the wires were in place. The intramuscular EMG activity was amplified, band-pass filtered (60-3000 Hz), digitized and sampled with sampling rate equal 10 kHz. To help the subjects maintain the activity of a single motor unit, visual and auditory feedback of the EMG activity was provided [6].

### 3. Estimation of stimulation effectiveness

#### 3.1. Proposed method

Presented methods enable to examine physiological parameters of the limb by using electrostimulation with video-recording. Pictures are received in fixed time intervals what makes analyzing at each part of training possible. In proposed solution a typical EMS devices were used that met the Medical Devices Directive: 93/42/EEC [7].

The training session uses special designed equipment. Its idea is shown on figure 1. Number 1 indicates a rear wall which is used as a background with a calibration grid (described in the chapter 3.4). The angle frame, (Fig. 1, no. 2) which other components are attached to, is stationary and has the ability to assert, for example, on the table in front of the patient. Hand rail is constructed to allow a person to base the test on behalf of the permanent and stationary parts and profiles. To avoid shivering and losing stability the base (Fig. 1, no. 3) is connected to all components. Additionally, a specially designed part for webcam location (Fig. 1, no. 4) to preserve constant distance from webcam to patient is used.

#### 3.2. Localization of electrodes and muscles stimulation

Two electrodes with dimension about 40x40 mm were used in order to cause upper limb movement. Electrodes are localized

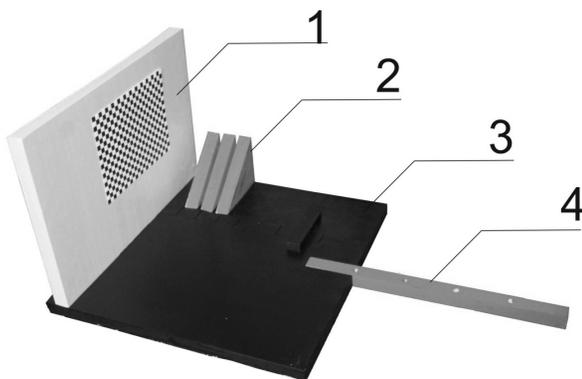


Fig. 1. The measurement construction (description in text)

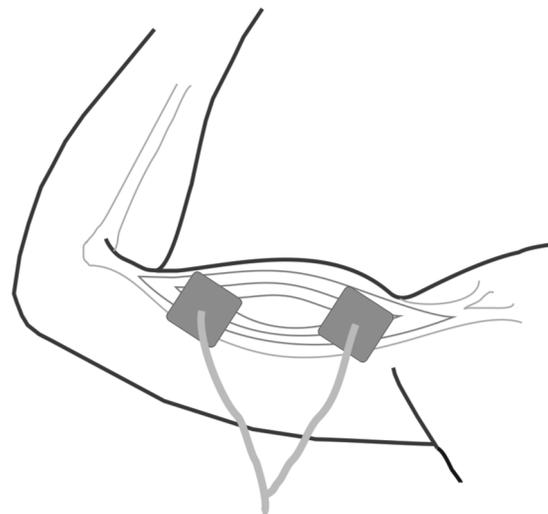


Fig. 2. Localization of electrodes

on the front side of the arm at about 5 cm distance from each other (Fig. 2). Localization of electrodes is specified as 1/3 and 2/3 length of humerus on biceps brachii muscle.

During algorithm validation frequency not lower than 80Hz and intensity not lower than 14mA are applied. These parameters allow to record high velocity and measure angles to 120 degrees.

### 3.3. Images recording

The process of recording is constant, therefore the used algorithm takes only these parts of frames, where the move appears. In this case, movement is defined as an angle change more than 1 degree (see 4<sup>th</sup> paragraph).

The webcam (2MPix, high-definition), which was used, has an appointed place on the base about 700 mm from the recorded arm. The rear part of the position has a bonded black and white checkerboard pattern with known dimensions, in order to carry out calibration. The patient's limb is placed in such way that the elbow is always located in a designated area. This arrangement allows the repetition of maintenance items for different patients and the confidence that the movement of the hand will always be in the field of camera view.

Most common image sensors in web and low-cost industrial camera are single-chip digital Bayer sensor. They use a color filter array that passes red, green, or blue light to selected subpixels. Full color pixels in captured image are interpolated from subpixels thanks to a demosaic algorithm. The natural-space color in Bayer sensor data is RGB color model.

The bit rate of high resolution image in RGB color model can be very high: for 2Mpixel \* 3 components of color \* 30 frame per second = 180MB/sec. The maximum bit rate of USB 2.0 interface is 60 MB/s. Comparing these quantities the flow of image information must be limited by data reduction or compression.

Based on limited perception of color in comparison with the perception of contrast, it is possible to reduce the number of transferred data without significant loss of data quality. In that method, the color description is transformed from RGB model to YCbCr or YUV color model. In these models, the chroma information (channels: Cb,Cr,U,V) is separated from the luminance (Y). In the next step, the chroma components are spatially decimated and then transmitted. The luminance components are transmitted without spatial decimation. One "chroma pixel" (chrominance sample) can cover 1, 2, 4 or 8 "luminance pixel". Decimation in chrominance data can reduce total bit stream of image data. After the image transmission the image data can be saved to file or shown in transmitted YCbCr color space or after re-transformation in RGB color space.

The described technique of color data reduction, called "chroma subsampling" can reduce image resolution which may affect the results of the segmentation process of color images. In the worst case, with a maximum data reduction, color information can be shared by four pixels. This reduces the bit rate of an uncompressed video signal by one-third with no visual difference.

Applied markers, high resolution and quality of recorded video stream allow to limit an error to maximum one pixel.

An important issue to perform a test is necessity of using markers such as colored circles. Two markers allow to locate

precisely the hand and compare successive movements (Fig. 2). Tags are attached to the elbow and the wrist. What is more, specially designed markers allow to distinguish them from electrodes.

The presented solution could also be equipped with accelerometers, what enables to designate the deflection angles of hand based on the signal representing the acceleration recorded by the sensors. Accelerometers respond to translational acceleration and the deviation from the direction of the gravitational field. What is important, accelerometers give us only relative displacement which in consequence conducts to error propagation.

### 3.4. Calibration

Before using the camera to record processes it is necessary to carry out the calibration procedure [8].

The calibration was done for the webcam, thanks to the Camera Calibration Toolbox for Matlab [9]. Calibration procedure enables to determine coefficients of the model image distortion introduced by the optics of the camera.

The software uses the Extended DLT (Direct Linear Transform) camera model [10] and maps e.g.: the camera perspective translation, radiation, perpendicular axis of the image etc. The results of calibration process is vector field of describing distortion introduced by the optics of the camera recorded images. For used webcam, distortions, calculated as the distance of each pixel between the image before and after correction, have a value about of 2 pixels which allows not to take calibration results into consideration (pixel error: [2.128; 1.455]) with regard to high resolution of recorded sequences (1600 x 1200 pixels).

## 4. The markers localization algorithm

In order to conduct the localization process, circular tracks and a reference line of reference markers enabling the appointment of the angle  $\gamma$  between the table and the hand are designed. Markers are taken as a circular sticker, which significantly increases the efficiency of estimation of the markers location. Colors of elements are chosen in such a way that they can be easily separated in different areas of the environment and human skin color in the analysis (markers [R: 155, G: 6, B: 158], Line [R: 51 G: 204, B: 255] in RGB format).

Estimation of pixels representing markers and reference line depend on thresholding procedure. The sequence of images is recalculated in two spaces HSV (Hue Saturation Value) and YCbCr. Permissible ranges of variation are: HSV: [0 – 1], YCbCr: [0 – 255] for each component.

The thresholds are chosen experimentally in order to match the color of markers and lines separately. Pixels with coordinates  $x, y$ , are classified as belonging to the markers if they are in the range of extent values. The range depends on a few factors such as brightness, daylight, artificial light, shadows etc. Conjunctions (1) and (3) show values used during high brightness; (2) and (4) – during shadow dominance. Otherwise the algorithm would find only one marker or failure-markers. Classification is based on the following equation:

Markers:

$$I(x, y) : H(x, y) > 0.7 \wedge S(x, y) > 0.7 \wedge Cb > 135 \wedge Cb > 165 \wedge Cr > 110 \tag{1}$$

$$I(x, y) : H(x, y) > 0.8 \wedge S(x, y) > 0.9 \wedge Cb > 165 \wedge Cb > 170 \tag{2}$$

Reference line

$$I(x, y) : H(x, y) > 0.5 \wedge H(x, y) < 0.6 \wedge S(x, y) > 0.3 \wedge Cb > 70 \wedge Cb < 150 \wedge Cr > 120 \tag{3}$$

$$I(x, y) : H(x, y) > 0.5 \wedge H(x, y) < 0.65 \wedge S(x, y) > 0.9 \wedge Cb > 140 \wedge Cb < 170 \wedge Cr > 85 \tag{4}$$

where  $H, S$  are components of the HSV and  $Cb, Cr$  of the YCbCr format.

After thresholding, morphological operations are used in order to fill every hole in the resulting objects. Each object in a frame is determined by coordinates of the center of its gravity, size and eccentricity. In case of detection of larger number of objects, these parameters allow to select 2 objects representing markers automatically.

Identification of individual markers is performed on the basis of distribution of individual pixels in a frame. Due to the fact that the markers are placed on the hand (displacements of markers relative to each other can take place only in a certain way), it is possible to locate markers on the basis of the objects coordinates and the distance from the origin. Marker no 1 (located on the wrist) is chosen as the one which distance to the upper left corner of the image is the smallest. Last object is identified as a marker 2 (located on the elbow) automatically.

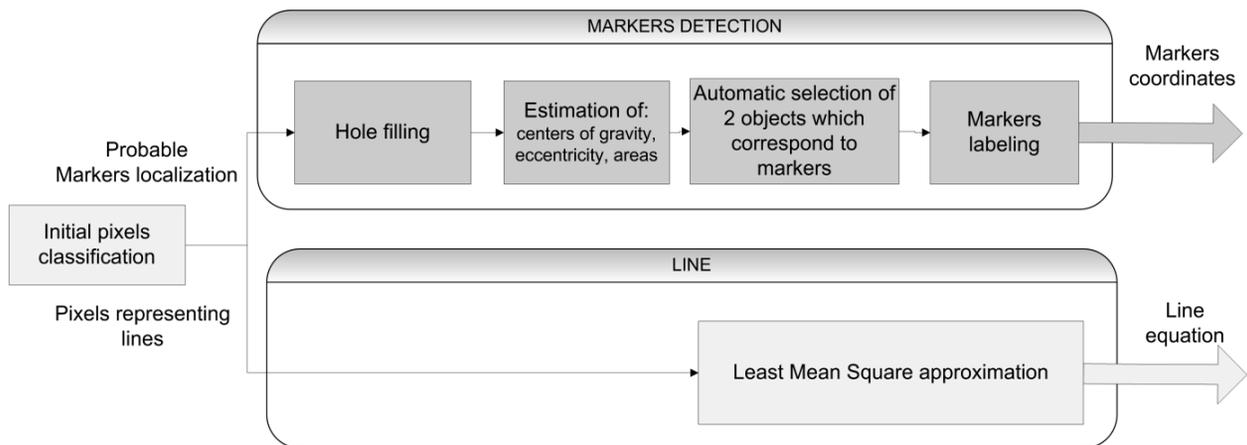


Fig. 3. Block diagram of searching markers and reference line

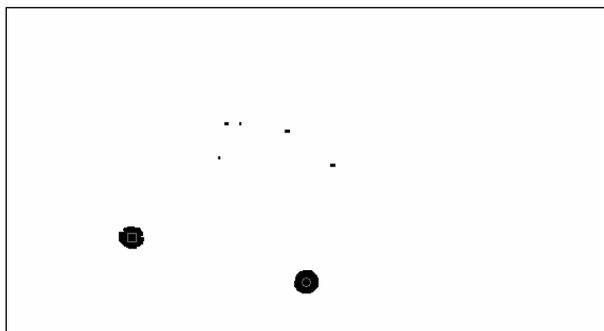


Fig. 4. Graphical results showing separated markers

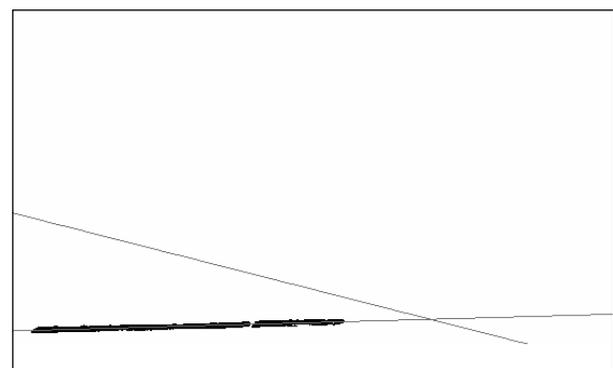


Fig. 5. Graphical results showing two lines pose an angle between separated markers and reference line

Information about the location of markers 1 and 2 allows to determine the equation of a straight, which is used to determine the searching angles. The idea of the method is presented in figure 3, whereas results of markers and line detection in figures 4 and 5.

Moreover the algorithm of the process markers separation eliminates wires and electrodes from the pictures. Cables and others accessories in pictures do not have influence on separation process. The only important issue is to guarantee that anything else in the vision field is similar in terms of markers' colors.

## 5. The angle between upper limb and frame of reference

Determination of angle between the arm and the table is still necessary to specify the equation of a straight on the table reference. Due to the large number of pixels per line, shown in the sequence,

the authors proposed the calculation of the corresponding straight line equation using least squares approximation.

Intersection of two straights: the first one passing through the markers 1 and 2:  $y = m_1x + b_1$ , and the second one, the corresponding line on the table:  $y = m_2x + b_2$ ; allows to determine the angle between the arm and the table of equation in the following form:

$$y = \arctg \left( \frac{m_1 - m_2}{1 + m_1 \times m_2} \right) \quad (5)$$

After the analysis for all selected frames, the program draws the graphs showing the trajectory of motion and the angles between the arm and the table. Exemplary results are presented in figures 6 and 7. No 1 and 2 in figure 7 indicate localization of markers stuck on arm.

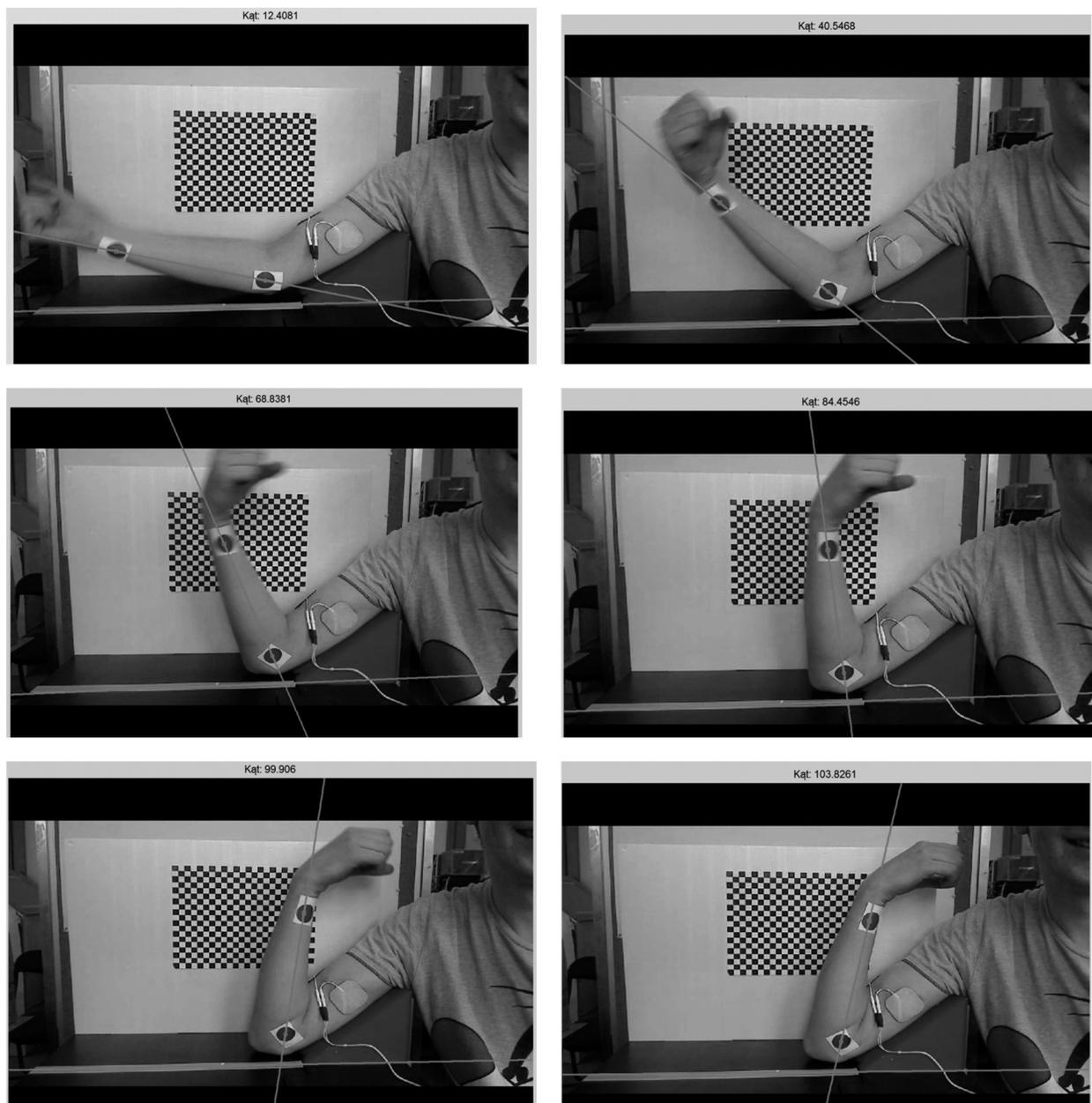


Fig. 6. Exemplary frames with marked results

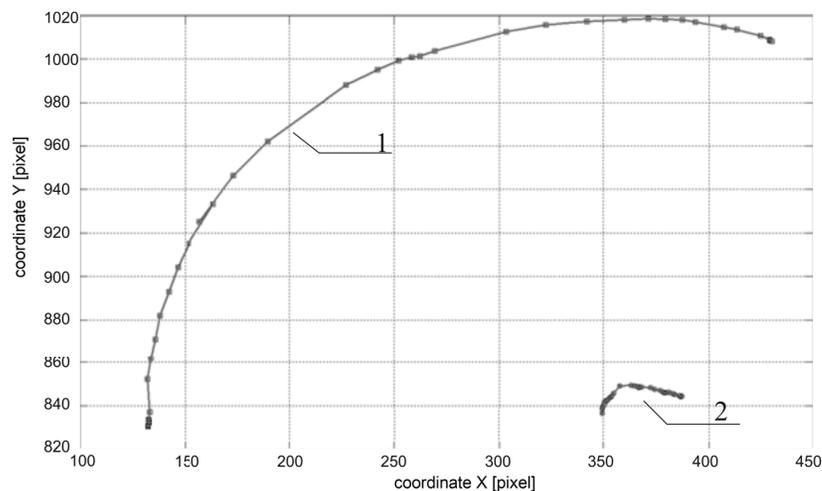


Fig. 7. Trajectory of stimulating arm

## 6. Summary and Discussion

The article presents the concept of a system for measuring angles of deflection of the arm and hand movement trajectory for the evaluation of muscle electrostimulation using EMS device and simple vision system.

The electrical muscle stimulation as a therapy becomes cheaper and cheaper and needs a detailed verification of success. The fear of electrical stimulation appears less often than before. In consequence this method of treatment is used more often.

Incurred costs of described method are the main advantage. Every component, that was used in this research, was bought to reduce expenses and prepare a project of cheap and precise system. Thanks to using a typical printer that enables to get checkboard and markers with known color model it is possible to use the proposed method everywhere where access to webcam and computer is allowed.

In order to increase space of comparable methods the authors would like to upgrade the system in further research by using accelerometers and gyroscopes.

## 7. Acknowledgments

Scientific research was partially financed from Polish means for science as the research project N N501 241138: 'Magnus effect in mechanics and biomechanics' and by Polish Ministry of Science and Higher Education, project no 11.11.120.774 ("Działalność Statutowa").

## 8. References

- Hortobágy T., Lambert N.J.: Influence of electrical stimulation on dynamic forces of the arm flexors in strength-trained and untrained men. *Scandinavian Journal of Medicine & Science in Sports* 1992, 2(2): 70-75.
- Porcari J.P. et. al.: Effects of Electrical Muscle Stimulation on Body Composition, Muscle Strength, and Physical Appearance. *Journal of Strength and Conditioning Research* 2002, 16(2): 165-172.
- Powell J., Pandyan D., Granat M., Cameron M., Stott D.: Electrical Stimulation of Wrist Extensors in Poststroke Hemiplegia. *Stroke* 1999, 30: 1384-138.
- Izworski A., Michałek M., Rudzińska M.: Innovative ways of measurement and analysis of pathological tremor of upper limbs (In Polish: Nowe metody pomiaru i analizy patologicznego drżenia kończyn górnych). *Bio-Algorithms and Med-Systems* 2005, 1(1/2): 297-300.
- De Serres S.J., Enoka R.M.: Older adults can maximally activate the biceps brachii muscle by voluntary command. *Journal of Applied Physiology* 1998, 84(1): 284-291.
- Petersen N., Taylor J.L., Gandevia S.C.: The effect of electrical stimulation of the corticospinal tract on motor units of the human biceps brachii. *The Journal of Physiology* 2002, 544: 277-284.
- Council Directive 93/42/EEC of 14 June 1993 concerning medical devices, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31993L0042:EN:HTML>.
- Socha M., Duda K., Zieliński T.P., Duplaga M.: Algorithmic correction of geometric distortion of bronchoscope camera (In Polish: Algorytmiczna korekcja zniekształceń geometrycznych kamery bronchoskopu). XV Sympozjum Modelowanie i Symulacja Systemów Pomiarowych, 18-22 September 2005, Krynica, Poland.
- Bougue J.-Y.: Camera Calibration Toolbox for Matlab, [http://www.vision.caltech.edu/bouquetj/calib\\_doc/](http://www.vision.caltech.edu/bouquetj/calib_doc/).
- Bakstein H.: A Complete DLT-based Camera Calibration with a Virtual 3D Calibration Object. Thesis, Charles University, Praga, 1999.

# PYTHON BASED SIMULATOR OF ICU PATIENTS HYPERGLYCEMIA

JERZY BARANOWSKI, WALDEMAR BAUER, AND PAWEŁ PŁUCIENNIK

*Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie, Katedra Automatyki,  
 jb@agh.edu.pl, wbauer@student.agh.edu.pl, pplucien@student.agh.edu.pl*

**Abstract:** In this paper a Python based simulator of ICU patients glycemia is presented. Paper includes an underlying mathematical model along with its parameters. Simulator is described along with description of used software solutions. Its operation is also described and illustrated with appropriate screenshots.

**Keywords:** Python, ICU-MM, hyperglycemia, LQ controller, simulation

## 1. Motivation

One of the more important problems in the Intensive Care Units is the control of blood glucose level of patients. Many afflictions like heart attack or multi-organ disorder cause sudden rises in the glucose level. This situation is very dangerous because it increases the risk of infection, hampers blood coagulation and disturbs the metabolic balance. Studies shown, that the rigorous blood glucose level control substantially reduces the mortality rate in ICUs. Nowadays either frequent testing and insulin shots or automatic insulin pumps are used to avoid hyperglycemia even for patients without history of diabetes.

In this paper we present an application that allows analysis of ICU patients glycemia and also observation of potential treatment with chosen glucose level stabilisation strategy. Our goal was to create a user friendly application, that would not require expensive software packages. In the following sections we present the mathematical model of patients glucose dynamics along with its parameters, we highlight the stabilisation strategy that was the topic of authors earlier works and finally we present the simulator describing technical solutions and functionalities implemented.

## 2. Mathematical model

We consider a so called ICU-Minimal Model, given by following four, nonlinear ordinary differential equations

$$\begin{aligned} \dot{G}(t) &= P_1(G_b - G(t) - X(t)G(t) + \frac{F_G(t)}{V_G}) \\ \dot{X}(t) &= -P_2X(t) + P_3(I_1(t) - I_b) \\ \dot{I}_1(t) &= \alpha \max(0, I_2(t) - n(I_1(t) - I_b) + \frac{F_I(t)}{V_I}) \\ \dot{I}_2(t) &= \beta (G(t) - h) - \dot{I}_2(t) \end{aligned}$$

where  $G$  and  $I_1$  are the glucose and the insulin concentrations in the blood plasma. The variable  $X$  describes the effect of insulin on net glucose disappearance. The variable  $I_2$  does not have a strictly defined clinical interpretation and was introduced for mathematical reasons, as a way of modelling the functional pancreatic insulin system (as usually ICU patients are not diabetic). This model can also be represented with a diagram which is presented in the figure 1.

The parameters  $G_b$  and  $I_b$  denote the basal value of plasma glucose and plasma insulin, respectively. The model consists of two input variables: the exogenous insulin flow ( $F_I$ ) and the carbohydrate (glucose) calories flow ( $F_G$ ), both administered intravenously. The glucose distribution space and the insulin distribution volume are denoted as  $V_G$  and  $V_I$ , respectively. The coefficient  $P_1$  represents the glucose effectiveness (i.e., the fractional clearance of glucose) when insulin remains at the basal level;  $P_2$  and  $P_3$  are the fractional rates of net remote insulin disappearance and insulin-dependent increase, respectively. Endogenous insulin is represented as the insulin flow that is released in proportion (by  $\gamma$ ) to the degree by which glycemia exceeds a glucose threshold level  $h$  that is corresponding to the

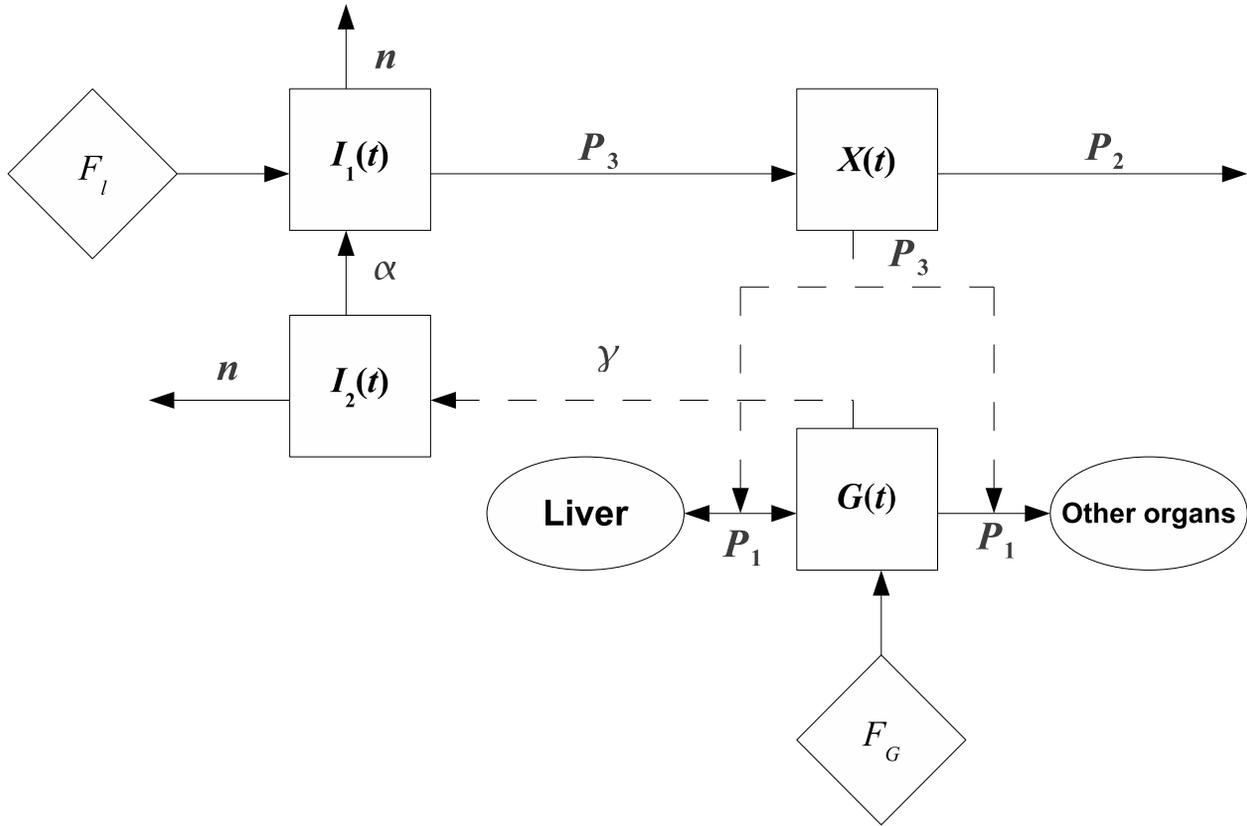


Fig. 1. Diagram representation of model (1)

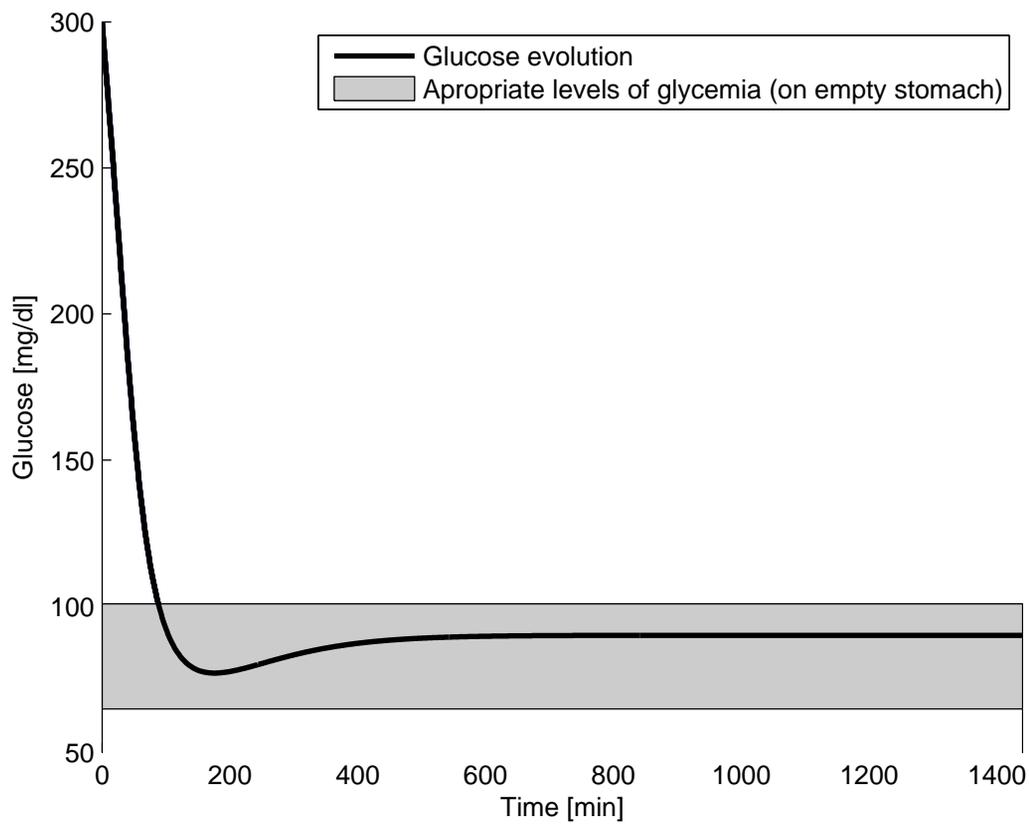


Fig. 2. Example of unassisted stabilisation of blood glucose level

normal glucose level of 7.5 mmol/dl. The time constant for insulin disappearance is denoted as  $n$ . In order to keep the correct units, an additional model coefficient,  $\beta = 1$  min, is added. The coefficient  $\alpha$  is a scaling factor for the second insulin variable  $I_2$ . For more details see [22]. The model (1) is slightly modified in order to improve the clarity of equations (all parameters are now positive and grouping in first equation was changed. Parameters of the model are collected in table 0 and were determined by Van Herpe et. al. [22].

**Tab. 1.** Explanation of used symbols

Sym.	Initial value	Unit	Description
G	300	$\frac{\text{mg}}{\text{dl}}$	blood glucose concentration
X	0.02	$\frac{1}{\text{min}}$	effectiveness of insulin on blood glucose disappearance
$I_1$	56	$\frac{\mu\text{U}}{\text{ml}}$	blood insulin concentration
$I_2$	0.1	$\frac{\mu\text{U}}{\text{ml}}$	mathematical variable
$F_G$	0	$\frac{\text{mg}}{\text{min}}$	administered glucose
$F_I$	0	$\frac{\mu\text{U}}{\text{min}}$	administered insulin
BM	80	kg	body mass
$V_G$	1.6 BM	dl	glucose distribution volume
$V_I$	120 BM	ml	insulin distribution volume
$G_b$	90	$\frac{\text{mg}}{\text{dl}}$	basal value of plasma glucose
$I_b$	56	$\frac{\mu\text{U}}{\text{ml}}$	basal value of plasma insulin
$P_1$	$1.31 \times 10^{-2}$	$\frac{1}{\text{min}}$	glucose effectiveness
$P_2$	$1.35 \times 10^{-2}$	$\frac{1}{\text{min}}$	net remote insulin disappearance
$P_3$	$2.9 \times 10^6$	$\frac{\text{ml}}{\text{min}^2 \mu\text{U}}$	insulin-dependent increase
h	136	$\frac{\text{mg}}{\text{dl}}$	glucose threshold level
n	0.13	$\frac{1}{\text{min}}$	insulin disappearance time constant
$\alpha$	3.11	$\frac{1}{\text{min}}$	scaling factor for $I_2$ variable
$\beta$	1	min	time scaling coefficient
$\gamma$	$5.36 \times 10^{-3}$	$\frac{\mu\text{U dl}}{\text{ml mg min}^2}$	proportionality factor

This model is an extension of so called minimal developed by Bergman et al. in 1981 [13] for modeling of glucose metabolism after intravenous glucose tolerance test (IVGTT), it was then adapted by Furler et al. in 1985 [7] to represent the diabetic state. This version, for purposes of Intensive Care Unit patient analysis was developed by Van Herpe et al. in 2007 [22] and is called ICU-MM (Intensive Care Unit Minimal Model). More details regarding development of minimal model can be found in [15, 16] and [14] p. 73. Different approaches to glucose dynamics modeling see [20]. Different approach to modelling of ICU patients see [8] and comparison of approaches can be found in [17]. More medical approach to the problem [23].

Using this model one can observe how glucose level stabilises from an initial high level to the norm. Example of such

behaviour can be seen in figure 2. As it can be seen patient reaches acceptable level, but a noticeable “hump” can be seen when glucose drops below desired level of 90 mg/dl (18 mmol/dl). That is why it is desired to improve stabilisation to avoid drops in glucose levels at the same time increasing the speed of reaching the steady state.

### 3. Application of control theory for treatment design

Using model (1) one can search for a method of administering insulin and glucose in such way that it would improve stabilisation of patient’s glucose level leading to quicker reaching of steady state and avoidance of drops below the desired threshold.

One way of determining a treatment strategy is to use a so called Linear Quadratic (LQ) optimal control. The main idea behind this treatment is to balance the quickness of stabilisation with a minimal exogenous input to the patient (doses of insulin and glucose). It can be represented as problem of optimisation of a performance index

$$J = \int_0^{\infty} (\Delta x(t)^T W \Delta x(t) + \Delta u(t)^T S \Delta u(t)) dt \quad (2)$$

where  $\Delta x$  represents the deviation from the desired steady state and  $\Delta u$  is the deviation from exogenous inputs corresponding to that steady state. It can be proven (see for example [21]) that under certain assumptions such treatment takes the following form

$$\Delta u(t) = K \Delta x(t) \quad (3)$$

that is in our case, the doses of insulin and glucose are proportional to deviation of state variables ( $G, X, I_1, I_2$ ) from the steady state. In this paper we do not intend to present this treatment scheme in detail as it was a subject of earlier works. More precise description of determining and application of LQ control to for stabilisation of patient glucose levels can be found for example in [2, 4-6].

### 4. Construction of the simulator

Our intent was to create a simulator that would be easily accessible for potential users. That is why we had to consider different computing environments allowing computation and analysis of dynamical model.

#### 4.1. Choice of the computing environment

Currently, the most popular environment to perform simulations in control engineering is Matlab. It offers many easy-to-use packages to numerical calculations which allows quick modeling of any phenomenon. User also has access to Simulink package with which he can create flowcharts in easy way. These two features caused that the package become very popular in the world of science. Despite of many positive features, Matlab is very expensive tool. For that reason, it is only used at technical universities and rich companies. In addition, applications written with Matlab are executable only on this specific environment.

High price of Matlab was the main reason for seeking alternative solutions. For many years the only alternative was GNU Octave and Scilab which are free applications for the Unix systems. But as well as Matlab, they only allow users to create software strictly connected with environment. If developing software with SciPy [19], NumPy [12] and Matplotlib [11] libraries, which for Python programming language [18], this issue has been resolved. SciPy, NumPy and Matplotlib (called PyLab) are open source libraries which work together to provide an alternative to Matlab. With the solutions used, they are able to provide full functionality of Matlab, including the Simulink packages. Python packages are based on the same libraries for calculations as Matlab. In summary, PyLab is excellent multiplatform replacement for the MathWorks product.

Choice of the Python language allowed also to construct the simulator with a web interface allowing use from a simple web browser that is why certain technological solutions had to be used.

## 4.2. Technologies used

The glycemia simulator was developed with Python programming language. As it was intended to have a web interface, Django [1] – one of the most popular Python web frameworks – was used. This combination of cPython (the most common Python implementation written in C) and Django provides great environment for building modern web applications. Python is object-oriented, high level multi-paradigm [10] programming language with lots of useful third party libraries. Its syntax, simple and intuitive, lets developers concentrate on the subject without being distracted by language complexities. The additional scripts and modules mentioned – often intensely developed – covers many areas of programming: from simple system scripts, websites, application servers to complex scientific computing and real-time algorithms [9]. The application presented below uses NumPy and SciPy libraries to work with matrices and complex data structures. Control Systems Library [3] is used in order to calculate infinite-horizon, continuous-time Linear-Quadratic Regulator needed for some simulations. Graphical representation of data is prepared with 2D plotting library – matplotlib. This one provides tools to generate “publication quality figures” in many hardcopy formats.

## 4.3. Implemented functionalities

It should be noted that the simulator was created in Polish as it was a part of engineer's degree thesis of the second and the third author. The simulator itself is quite simple. The space is divided into two sections: the left one for specifying simulation details and the right one for viewing results. Core functionality is based on one xHTML form used to set all of the simulation parameters like levels of glucose and insulin at the beginning of the treatment, desired glucose level patient's body mass and much more. Values for each one of these parameters can be put in using standard text input fields or with special sliders chained with each of them (see screen-shot in the figure 3).

For every slider-input pair, if one of the elements changes its value, the second one is updated instantly. If the value put in the text input (integer or float) is out of the specified range for the

parameter that is being modified, the range value is considered during calculations.

The figure shows a web interface for setting simulation parameters. It consists of five rows, each with a slider and a text input field. The parameters and their values are:
 

- Masa pacjenta: 80.0 [kg] (slider range 50-150)
- Glukoza bazowa: 90.0 [mg/dl] (slider range 80-110)
- Insulina bazowa: 56.0 [µU/ml] (slider range 50-60)
- Glikemia początkowa: 300.0 [mg/dl] (slider range 110-300)
- Insulinooporność początkowa: 0.020 [1/min] (slider range 0-0.05)

 Each slider has a diamond-shaped handle and a corresponding text input field containing the current value.

Fig. 3. Choice of the model parameters in the simulator

In figure 4 there is a block where variants of ICU-MM controlling can be chosen. At the moment of writing the paper there are four options of simulation implemented and available: autonomous ICU-MM, insulin / insulin & glucose LQR controlled systems and the one based on intravenous insulin infusion. If the last of these is chosen, additional customizations (time period of dosage and the insulin portion) become available.

Wybierz symulacje:

- Model podstawowy
- Model sterowany jedynie insuliną
- Model sterowany glukozą i insuliną
- Model z wlewem dożylnym

Fig. 4. Choice of the desired type of simulation

The submitted form runs scripts generating Matlab's plot-like graphs with matplotlib library and calculating simulation results.

These are divided into three sections. In each one of them there are several informations prepared: model stabilization time, minimal level of blood sugar and of course administered glucose and insulin. Of course, there are only considered simulations specified with the input data form.

First one is the summary table (see figure 5) that gathers all of the calculated values in one place. It makes further comparisons easier and provides an overview of used methods.

Second one presents results of all simulation variants separately. The graphs thumbnails for glucose and insulin trajectories are added (example in figure ). Full-size transparent plot can be obtained by clicking the small one.

The last but not least one is a functionality of exporting simulation results into PDF document. The format mentioned works great with reports that needs to be printed. In this case, important data is being prepared in readable and well-formatted way. This allows users to keep simulation informations with them

even if they are offline. Example of such report is presented in the figure 7.

## 5. Conclusions

In this paper a functional dedicated simulator of ICU patient glycemia was presented. This simulator has a web interface allowing usage with a simple internet browser giving it a great accessibility. It allows observation of unassisted glucose level stabilisation but also an analysis of different treatment strategies. As an example an automated treatment (using for example an insulin pump) with a so called LQ controller is included. Simulator allows selection of different patient parameters, choice of simulation and provides a possibility of generating reports in a form of a pdf file.

Metoda regulacji	Stabilizacja po [min]	Podana glukoza	Podana insulina	Cukier min.
Model bazowy	375	---	---	64.0mg/dl
Insulina + glukoza	53	10.83g	0.18 $\mu$ U	87.5mg/dl

Fig. 5. Summary table generated by the simulator

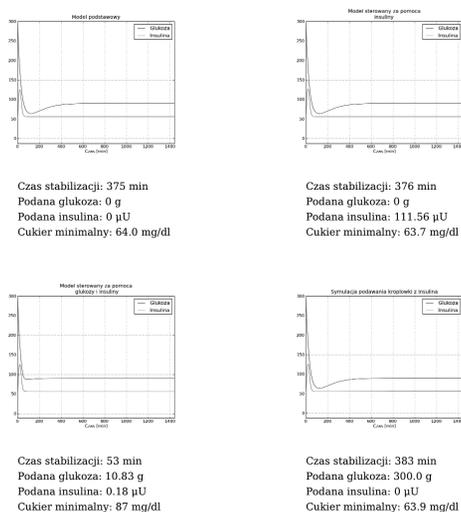


Fig. 6. Example of a plot generated by the simulator



Symulator glikemii pacjenta OIOM

## Raport z symulacji



Raport wygenerowano dnia 21.12.2010, o godzinie 01:34. Wszelkie prawa zastrzeżone.

Fig 7. Example of a final report generated by the simulator

## References

- Django | The Web framework for perfectionists with deadlines, <http://www.djangoproject.com/>.
- Baranowski J., Mitkowski W.: Optimal glucose level regulation for Intensive Care Unit patients: CSEDOM approach. Papers. 24th IFIP TC7 Conference on System Modelling and Optimization, Palais Rouge, Buenos Aires, Argentina, 2009, pp 46-47.
- Python Control Systems Library, [http://sourceforge.net/apps/mediawiki/python-control/index.php?title=Main\\_Page/](http://sourceforge.net/apps/mediawiki/python-control/index.php?title=Main_Page/).
- Bauer W., Pluciennik P.: Stabilizacja glikemii pacjentów OIOM przy użyciu regulatora LQ. *PAR – Pomiar, Automatyka, Robotyka* 2011, (4): 76-81.
- Bauer W., Pluciennik P.: Symulator glikemii pacjenta Oddziału Intensywnej Opieki Medycznej (OIOM). Praca Inżynierska. Akademia Górniczo-Hutnicza im. Stanisława Staszica, Wydział Elektrotechniki, Automatyki, Informatyki i Elektroniki, Kraków, 2011.
- Bauer W., Pluciennik P., Baranowski J., Mitkowski W.: LQ based glucose stabilisation for Intensive Care Unit patients. In: Tadeusiewicz R., Ligęza A., Mitkowski W., Szymkat M. (eds), CMS'09: Computer Methods And Systems, Kraków, 2009, pp. 307-312.
- Furler S.M., Kraegen E.W., Smallwood R.H., Chisolm D.J.: Blood glucose control by intermittent loop closure in the basal mode: Computer simulation studies with a diabetic model. *Diabetes Care* 1985, 8(6): 553-561.
- Hovorka R., Chassin L.J., Ellmerer L., Plank J., Wilinska M.E.: A simulation model of glucose regulation in the critically ill. *Physiological Measurement* 2008, 29: 959-978.
- Wittber S.: Hard Real-time Python, or Giant Robots of Doom. PyCon Australia, 2010, <http://blip.tv/pycon-australia/hard-real-time-python-or-giant-robots-of-doom-3856630>.
- Ziadé T. et al.: Expert Python Programming. *American Journal of Physiology* 2008.
- Matplotlib: Python plotting, <http://matplotlib.sourceforge.net/index.html>.
- Numpy – Scientific Computing Tools for Python, <http://numpy.scipy.org/>.
- Bergman R.N., Phillips L.S., Cobelli C.: Physiologic evaluation of factors controlling glucose tolerance in man: measurement of insulin sensitivity and beta-cell glucose sensitivity from the response to intravenous glucose. *J. Clin. Invest.* 1981, 68(6): 1456-1467.
- Chee F., Fernando T.: Closed-Loop Control of Blood Glucose of Lecture Notes in Control and Information Sciences. Berlin-Heidelberg: Springer-Verlag, 2007.
- Kovács L.: Extension Of The Bergman Minimal Model For The Glucose-Insulin Interaction. *Periodica Polytechnica Ser. El. Eng.* 2006, 50(1-2): 23-32.
- Makroglou A., Li J., Kuang Y.: Mathematical models and software tools for the glucose-insulin regulatory system and diabetes: an overview. *Applied Numerical Mathematics* 2006, 56: 559-573.
- Van Herpe T., Haverbeke N., Van den Berghe G., De Moor B.: Prediction Performance Comparison between three Intensive Care Unit Glucose Models. Proceedings of the 7th IFAC Symposium on Modelling and Control in Biomedical Systems, Aalborg, Denmark, 2009.
- Python Programming Language – Official Website, <http://python.org>.
- Scipy – Scientific Tools for Python, <http://www.scipy.org/>.
- Clausen W.H.O., De Gaetano A., Vølund A.: Within-patient variation of the pharmacokinetics of subcutaneously injected biphasic insulin aspart as assessed by compartmental modelling. *Diabetologia* 2006, 49(9): 2030-2038.
- Mitkowski W.: Stabilizacja systemów dynamicznych. Warszawa: WNT, 1991.
- Van Herpe T., Espinoza M., Haverbeke N., De Moor B., Van den Berghe G.: Glycemia Prediction in Critically Ill Patients Using an Adaptive Modeling Approach. *Journal of Diabetes Science and Technology* 2007, 1(3): 348-356.
- Tatoń J., Czech A., Bernas M.: *Diabetologia kliniczna*. Warszawa: Wydawnictwo Lekarskie PZWL, 2008.

## AN APPLICATION SOFTWARE FOR ANATOMICAL STRUCTURES SEGMENTATION FROM TOMOGRAPHIC DATA OF ABDOMINAL CAVITY

GRZEGORZ TOPOREK<sup>1,2</sup>, MIROSLAW SOCHA<sup>1</sup>, ANDRZEJ SKALSKI<sup>1</sup>

<sup>1</sup> AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY, DEPARTMENT OF MEASUREMENT  
AND INSTRUMENTATION, Al. A. Mickiewicza 30, KRAKOW, POLAND;

<sup>2</sup> AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY, MULTIDISCIPLINARY SCHOOL  
OF ENGINEERING IN MEDICINE, Al. A. Mickiewicza 30, KRAKOW, POLAND;  
{gtoporek@gmail.com, socha@agh.edu.pl, skalski@agh.edu.pl}

**Abstract:** New, increasingly sophisticated and accessible imaging methods, ranging from Computed Tomography (CT), Magnetic Resonance Imaging (MRI) to Positron Emission Tomography (PET) allow scientists and medical doctors to obtain non-invasively potentially life-saving information about patient condition and body. However, the purpose of recent medical imaging is not only to obtain simple visualization and insight into human anatomic structures, but also as a powerful tool for computer-aided diagnosis, radiotherapies, tumor tracking and surgical operation planning and simulation. This is exactly the field where very accurate and reliable image segmentation methods can play a key role. In this work, a cross-platform application software equipped in algorithm for segmentation of anatomical structures from abdomen CT data has been developed. As a segmentation technique balancing good accuracy and efficiency, a semi-automated, region growing segmentation algorithm, based on voxel intensities criteria, was used. Additional features of an application are: ability to visualize large medical data, render segmentation results in 3D as well as support of DICOM standard.

**Keywords:** image segmentation, application software, computed tomography, region growing algorithm

### Introduction

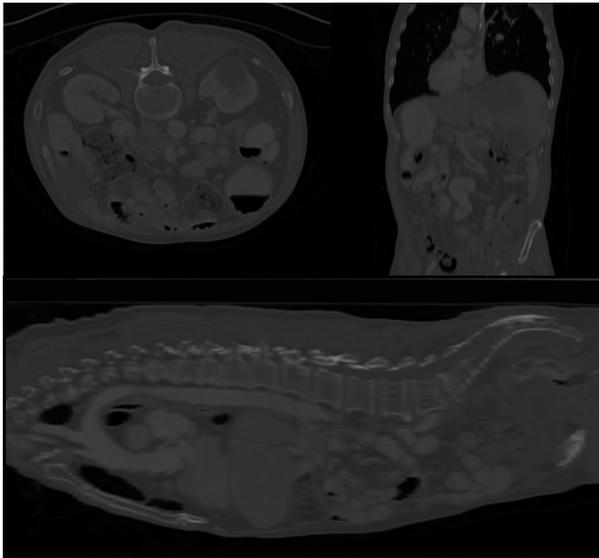
Image segmentation is recently one of the most prominent method in medical image analysis. When dealing with large, multidimensional medical data, the automatic or semi-automatic segmentation techniques aiming to distinguish particular anatomical structures of patient body are crucial [1]. For instance, by evaluating the detailed shape and position of anatomical structures, the surgeons obtain *a priori* knowledge that can be used for optimal surgical procedures planning. Additionally, segmentation of anatomical structures can be applied for better and wiser radiotherapies, while both radiation dose can be targeted with greater precision and the progress of disease or changes in tumor can be tracked during therapy. Thus decreasing necrotic dose of radiation and damage for healthy tissue [2].

Accurate, quantitative, readable and repeatedly data extraction in a reasonable time is not a trivial task, especially when a wide range of biomedical imaging methods and applications (surgery, radiotherapy, diagnosis) is taken into account [3]. There are several reasons that can pose problems for correct and ef-

ficient image segmentation. First of all, as having large amount of 3D data, medical image segmentation is very computationally expensive. In some segmentation methods, for instance region growing, implementing a functional algorithm that allows one to deal comfortably with datasets of a hundred megabytes can be a demanding task. Secondly, as presented in work [4], description and analysis of medical images are highly inhibited due to the complexity and diversity of shapes and sizes of anatomical structures. Having one general algorithm for different organs is almost impossible. Going further, it is sometimes difficult to distinguish the shape and position of the same organ, that belongs to the different patients. It is also the case, that medical images suffer from spatial aliasing and noise, which may cause blurring of structures' boundaries and edges. As a result, traditional *bilevel thresholding* methods [7] and low-level image processing techniques will generate wrong results. Thirdly, one can have difficulties when visualizing and verifying segmented data. Whilst interpretation of 2D segmented images can be an easy task, to analyze a 3D model of a human structure can determine some difficulties. Therefore, it is important to examine segmentation

results not only by verifying single slices, but also to compare them quantitatively with given ground truth data using well-know measurements and verification methods, thus avoiding wrong spatial impressions.

The most popular methods for medical data segmentation nowadays are techniques based on models, including *active contours* [6], *level set* and *atlas-based methods* [5]. However, the main challenges with those techniques are large computational complexity and parameter-based nature, which can pose many problems for inexperienced users, such as medical doctors. On the other hand, methods based on *thresholding* [7], despite of being very simple and robust in implementation, are not applicable to complex, inhomogeneous medical data, as it is for tomographic slices depicted on Fig. 1. Furthermore, reliability and accuracy of medical image segmentation are much more valued than performance and automaticity. That is why *region growing* algorithms, that are a good compromise between complex, parameter-based, computationally expensive *deformable models* and simple but robust *thresholding* techniques, have gained nowadays such a large popularity. This was also the main criterion that encouraged the authors to implement this method in the application software presented in this paper.



**Fig. 1.** Axial, coronal and sagittal view through volumetric data of a human body obtained from a medical computed tomography

## Methods

Segmentation methods based on *region growing* algorithms are in principle quite effective and reproducible, with little cost of computational power and complexity. First of all, one has to choose an initial, starting point of segmentation process known as a seed point, which is a pixel or group of pixels belonging to the object of interest. The seed point can be pointed out manually by an operator (medical doctor or technician), semi-automatically

by using some *a priori* knowledge about segmented structure or purely automatically, without any user interaction (for instance, in the center of an image). Afterwards, one needs to define a size of neighborhood area around seed point in order to perform some statistics or computations within its field. The next step is to determine the main criteria of pixel membership to region in neighborhood of a seed point. In other words, one needs to decide whether a pixel around the initial region should be considered or not as a part of this region – an object of interest, to be exact. Based on selected criteria, pixels in a neighborhood are rejected or added to the initial point, thus resulting in region growing and object segmentation.

While the size of an area in which pixels are evaluated is not very important, the chosen criterion is a critical parameter of an image segmentation via *region growing* algorithms, being responsible for the effectiveness and quality of segmentation. One possible criterion is based on the observation that all pixels intensities from one object are usually placed within the range around a mean value. This approach is used for example in the *confidence connected region growing* algorithm.

The key idea of the algorithm is based on computing region statistics by taking voxels intensity. First of all, an algorithm evaluates the mean  $\mu$  and standard deviation  $\sigma$  values of voxels intensity around the seed point (initial point), that has been chosen by the operator. The size of the neighborhood region in which those values are evaluated needs to be defined before. Then, a proposed by the user parameter  $m$ , so-called multiplier, is used as a weight for standard deviation  $\sigma$ , thus defining an intensity range. In consequence, all voxels values  $I(x, y, z)$  which are both within the evaluated intensity range and within the neighborhood, will be accepted and labeled with an object, as it is in equation 1:

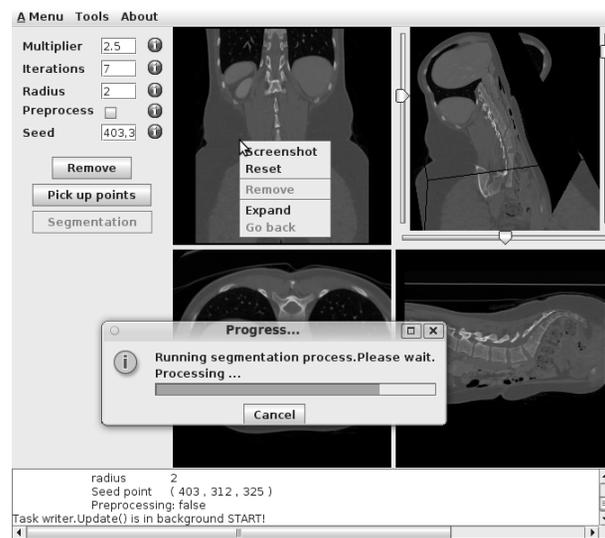
$$I(x, y, z) \in \langle \mu - m\sigma, \mu + m\sigma \rangle \quad (1)$$

The procedure is repeated for every voxels from the neighborhood until any of the voxels will satisfy the criterion. This process is repeated iteratively until any of the voxels can be added to the object. Then, the mean value  $\mu$  and its standard deviation  $\sigma$  for the region are computed again but this time taking into account also all voxels added before. Process is finished if the number of iterations is exceeded, thus causing growing of the segmented object of interest.

## An application software

Previously mentioned segmentation method was integrated into an application software for 3D volumetric medical data segmentation and visualization depicted in Fig. 2. The graphical user interface (GUI) of an application has been implemented in Swing Java widget toolkit – an API that provides graphical components for Java programs. Employing the Java language gave us also portability of an application and system-independence. However, in order to provide both additional image processing routines, segmentation methods and hardware accelerated data visualization pipelines, two low-level, C++ libraries have been linked with Java code. These were: Insight Segmentation and Registration Toolkit (ITK) [9] and Visualization Toolkit (VTK) [10]. By wrapping those two libraries, a dramatic increase in performance has been

noticed due to flexible and efficient memory management (ITK) as well as hardware accelerated 3D rendering (VTK).

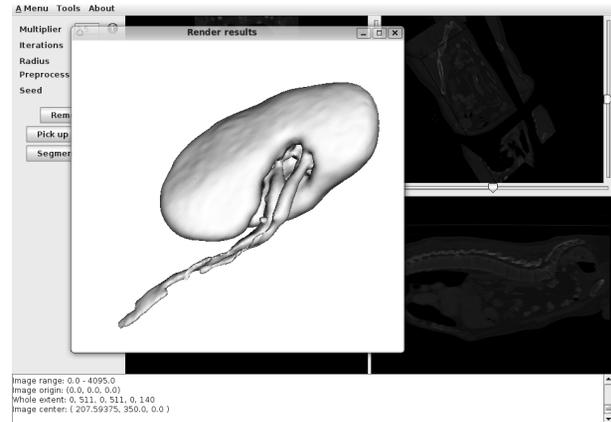


**Fig. 2.** The main window of an application software is equipped in four view panels: three orthogonal planes through the data: sagittal, axial, coronal and three-dimensional view of the volume (top-right corner)

In order to present volumetric medical data in a simple and intuitive way, a main window of an application is equipped in four view panels: three orthogonal planes through the image data (sagittal, axial, coronal) – so-called multi-planar reconstruction planes (MPR) and additional one, placed on the right-corner, that shows three-dimensional view of an image. A three-dimensional view of a data contains all MPR in 3D space, when geometrical relations between them are preserved. A user can change the orthogonal plane by using three sliders. Moreover, each image view is placed on a separate panel, which provides its own events and functionalities, so that a user can interact with it individually. For instance, one can reset the rendering camera or take a screenshot from a chosen plane by simply clicking the right mouse button and selecting the applicable option from pop-up menu.

A developed by authors application software extensively uses multithreading. Performing data processing on the separate processors let us to significantly increase performance of an application. Thus computationally expensive and time consuming 3D medical data segmentation can be easily performed on a personal computer without affecting responsiveness of the user interface. Additional features of an application software that are worth mentioning are: support of wide range of medical data formats, including image series in DICOM format; an interactive rendering module (Fig. 3); user-friendly interface; comprehensive context help and intuitive data range scalability. For more information, interested reader might have a look into a context help of an application.

An additional panel that controls segmentation process is also available. By using it, one can change segmentation parameters, initialize a seed point either by pointing on a view panel or setting it manually and start segmentation algorithm.



**Fig. 3.** The main window of an application software together with additional rendering and visualization module, that can be used to directly show segmentation results

As a result of a segmentation method, one obtains a binary, three-dimensional matrix, containing information whether voxel belongs to the object of interest or not. A binary, volumetric data can be then presented in a separate window of an application software, thanks to implemented rendering module.

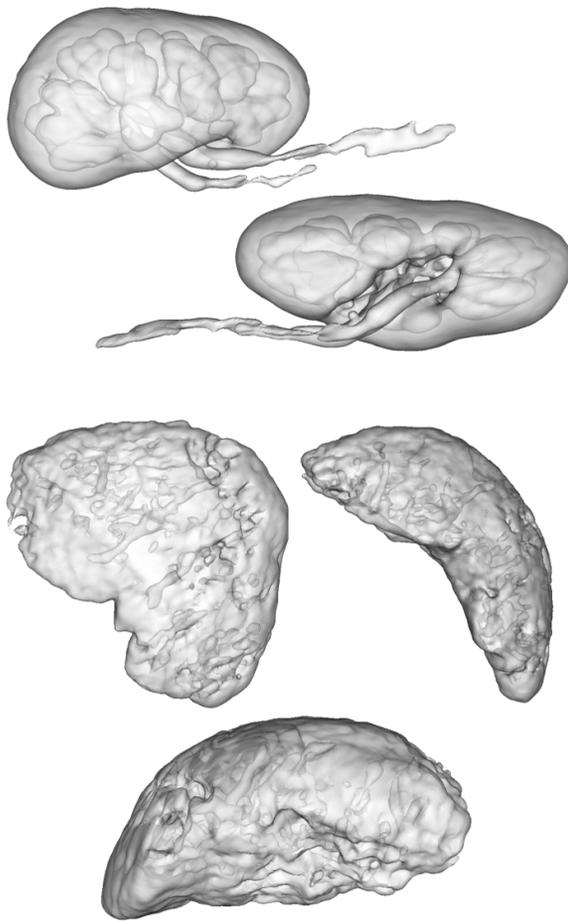
A direct volumetric rendering of the segmentation results can be computed using following methods: volume ray casting (class `vtkVolumeRayCast-CompositeFunction`), maximum intensity projection (MIP) mapping (class `vtkVolumeRayCastMIPFunction`) and iso-surface rendering with a possibility to change color and opacity of the iso-surface (`vtkVolumeRay-CastIsoSurfaceFunction`). Due to the fact of high computational demands, when visualizing volumetric objects, all data is decimated. A user has a possibility of switching between previously mentioned rendering methods, based on own observations and the nature of analyzed data. Add to all, one can easily changes the range of presented data, values of iso-surface, parameters of visualization as well as zooms the view and takes a screen-shots by using the specially implemented for this module events.

Because the segmentation results that are presented as a volumetric rendering are not smooth on boundaries, in order to improve visualization, some additional post-processing has been applied to the data. First of all, the segmentation results are smoothed with a Gaussian filter with a given radius size (class `vtkImageGaussianSmooth`). Then, by applying a marching cubes algorithm one computes the iso-surface represented by a polygonal mesh (class `vtkMarchingCubes`). In order to decrease computational load, we used two filters: `vtkDecimatePro` for triangle reduction without affecting original geometry and `vtkSmoothPolyDataFilter` to improve position of the vertexes on iso-surface. Finally, a visualization algorithm computes the surface normals (class `vtkPolyData-Normals`) in order to provide realistic view of an object (adequate light and shadows).

## Integration with native libraries

In order to ensure system independence of an application software, a Java language has been employed. However, for image processing and visualization, a well-known and widely-used

libraries ITK and VTK has been used. Despite the fact that both libraries are written in C++ there is a possibility to bind them into other languages, including Python, .NET, Tcl and Java. Both libraries use different models of creating objects and memory management schemes. For instance, ITK is highly templated, where on the other hand VTK uses abstract factory design pattern. That is why, it is not a trivial task to transfer data between them. However, a specially designed mechanisms and objects that help developers to forward data from one pipeline to the another one, has been developed by the contributors to ITK and VTK. Nonetheless, it was noticed by the authors of this work that despite the fact that both libraries work well in C++ environment, it is difficult to wrap them into a Java code.



**Fig. 4.** Exemplary segmentation results of kidney and stomach obtained by using integrated into application rendering mode

The main challenge was to transmit data and the segmentation results from ITK's objects to visualization pipeline of VTK. Both filters that are widely used in C++ language for importing and exporting data between mentioned libraries are not working correctly in Java. Due to a highly complex architecture of those filters, a correct wrapping is practically infeasible. The main problem is not only to transfer data but it is also to send signals that control data stream. Referring to the newest declarations of main contributors to ITK and VTK, there will be no future support

of the Java language and no solution to a previously described problem. In consequence, the authors of an application software proposed a solution relying on reading data into both pipelines (data processing pipeline and visualization one) simultaneously. We assumed that a memory for particular objects is initialized only when the object is needed. To do that, a lazy initialization design pattern, that is commonly used in Java programming, has been implied. By applying this programming technique, a consumption of a memory has been reduced to the minimum. However, it is planned in the future to implement our own wrappers around native code to prevent previously mentioned problems.

## Segmentation results

Five segmentation experiments on independent tomographic datasets of abdominal cavity, such as presented on Fig. 1, have been performed to validate segmentation algorithm. The medical data has been obtained from a free, open medical database [8], hence interested reader, equipped with an application software, may repeat those experiments without any obstacles. To distinguish different human organs from abdominal cavity, a segmentation algorithm based on a *region growing confidence connected* method, that was implemented as an integral part of an application software, has been applied to given datasets. Based on chosen seed points, we performed segmentation of the anatomical structures, including stomach, large and small intestine, liver and kidney.

As presented in Fig. 3, obtained results are especially promising for anatomical structures that highly contrast with surrounding tissue, such as kidney or stomach. On the other hand, human organs from abdominal cavity, that are inhomogeneous and present high local variation in voxel intensity values, have blurred edges and are characterized by highly-sparse segments.

## Conclusions

Under this work, a cross-platform, multithread application software for tomographic data segmentation from abdominal cavity, that fulfils all demands concerning medical applications, was successfully developed. By using application it is possible to perform segmentation of 3D medical data from the computed tomography without any special hardware requirements. The segmentation results for most of the anatomical structures, such as kidney or stomach are satisfying, however the authors are aware of some limitations.

## References

1. Wirjadi O.: Survey of 3D image segmentation Methods. Technical Report, ITWM, 2007.
2. Acton S.T., Ray N.: Biomedical Image Analysis: Segmentation. Morgan & Claypool Publishers, 2009.
3. Tadeusiewicz R., Śmietański J.: Acquisition of medical images and their processing, analysis, automatic recognition and diagnostic interpretation. (in Polish) Pozyskiwanie obrazów medycznych oraz ich przetwarzanie, analiza, auto-

- matyczne rozpoznawanie i diagnostyczna interpretacja. Kraków: Wydawnictwo STN, 2011.
4. Tadeusiewicz R., Ogiela M.R.: Medical image understanding technology: artificial intelligence and soft-computing for image understanding. Heidelberg: Springer-Verlag, 2004, pp. 145-149.
  5. Sonka M.: Handbook of Medical Imaging. Vol. 2: Medical Image Processing and Analysis. SPIE The International Society for Optical Engineering, 2000.
  6. McInerney T., Terzopoulos D.: Deformable models in medical image analysis: a survey. Medical image analysis 1996, 1(2): 91-108.
  7. Fenster A., Chiu B.: Evaluation of segmentation algorithms for medical imaging. Engineering in Medicine and Biology Society, 7186-7189, IEEE-EMBS 2005.
  8. <http://public.kitware.com/pub/itk/Data/LiverTumor/>.
  9. <http://www.itk.org>.
  10. <http://www.vtk.org>.



## NEURAL NETWORKS FOR MEDICAL IMAGE PROCESSING

TOMASZ PIECIAK<sup>1</sup>, JOANNA JAWOREK<sup>1</sup>, MAREK GORGON<sup>1</sup>

<sup>1</sup>AGH University of Science and Technology, Faculty of Electrical Engineering, Automatics, Computer Science and Electronics, Department of Automatics, al. Mickiewicza 30, {pieciak,jaworek,mago}@agh.edu.pl

**Abstract:** The proposed article presents the most common types of artificial neural networks used to be performed in the field of medical imaging. The first section describes the use of artificial neural networks in the preprocessing stage, restoration of noisy and distorted images and in conjunction with morphological operations. The second part presents the artificial neural networks in image segmentation problem, particularly in adaptive binarization threshold level selection and as a complement to the active contour method.

**Keywords:** artificial neural networks, medical imaging, image processing, image restoration, morphological operations, segmentation, active contour, pulse-coupled neural networks (PCNN)

### Introduction

Modern methods of processing and analysis of medical images are based on advanced mathematical theories like calculus of variations, stochastic processes (random fields), tensors, differential geometry, graph theory, formal languages theory, and particularly artificial neural networks (ANN). Neural networks could be used for various stages of image processing like: preprocessing, segmentation, detection and recognition of anatomic structures and symptoms of diseases. In the past, various methods of solving specific problems and surveys on described topic have been written [1, 2, 3, 4]. For each of previously mentioned stages of image processing, neural networks could be used as a tool, which directly processes pixels, indirectly processes other data, e.g. pixels are transformed into parameters space, and additionally, ANN might be applied as complementary approach to existing methods of computer vision. One should realize that medical imaging is not limited to image processing and analysis of already acquired images, but can also include neural networks implemented using integrated circuits, which could assist in reconstruction process during X-ray Computed Tomography (CT) or estimate positions of detectors in Positron Emission Tomography (PET).

Relevance of presented issues is confirmed by figure 1, which presents annual increase of amount of new scientific publications in IEEE's database.

### Preprocessing stage

Preprocessing stage could precede actual analysis and pattern recognition applied to medical images, but is not a mandatory step. This stage is intended to remove geometric distortion, perform noise filtration, alteration of contrast and saturation, and increase image sharpness as well [5, 9]. In practical medical applications, the most important aspects are geometric transformations e.g. removing distortion from fiberoscope images acquired in bronchoscopy examination [6] and contrast enhancement, performed e.g. on glial cell images obtained from confocal microscopy [7].

### Noisy and distorted image restoration

Each computer vision system, especially medical one, is prone to adding distortion and noise to acquired image. Need for image restoration techniques has been accentuated during first aerospace flights, in which case photos were heavily distorted by vibrations of camera and constant rotation of spaceships [8]. From the mathematical point of view, model of the distorted image (1), shown in figure 2, can be written as [9]:

$$g(m, n) = f(m, n) \otimes psf(m, n) + \mu(m, n) \quad (1)$$

where:

- $g(m,n)$  – output image,
- $f(m,n)$  – perfect image,

- $psf(m,n)$  – point spread function, responsible for distortion,
- $\mu(m,n)$  – additive noise.

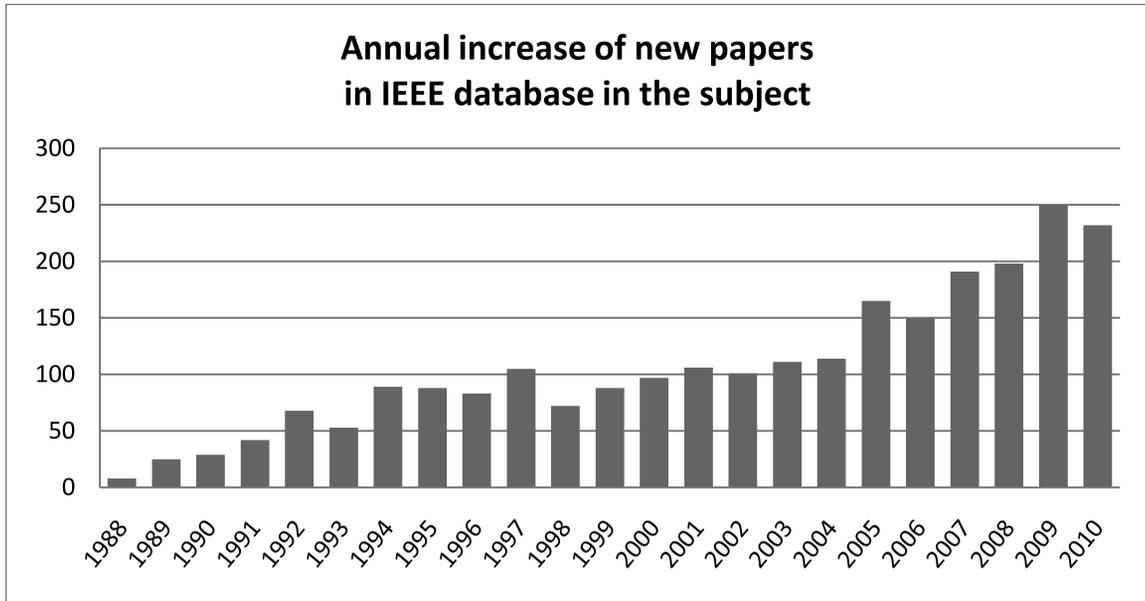


Fig. 1. Quantity of new papers in artificial neural networks in conjunction with medical imaging in IEEE database in each year. Papers include *neural networks* and *medical imaging* tags (Source: Own elaboration based on <http://ieeexplore.ieee.org/>)

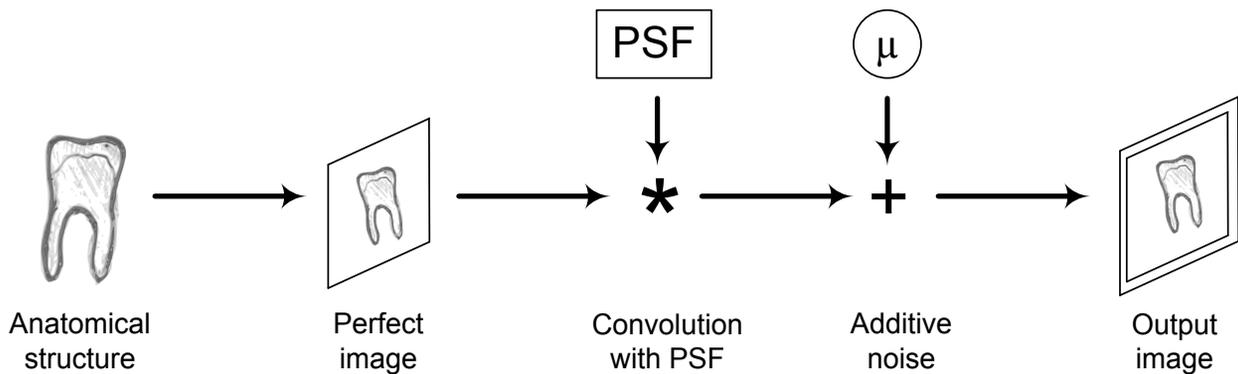


Fig. 2. Data acquisition process in typical computer vision system

Taking into account the randomness of the image, it may be modelled as a multidimensional random process, and therefore as a random field. This means that estimation of the original image becomes a statistical inference problem [10]. Common methods to solve such problem are based on estimation of unknown parameters of probability density function (PDF), and they include Wiener filter and Least Mean Square Error (LMSE). However, these approaches require a priori knowledge of noise level and image distortion. A method allowing to estimate image, PDF function and noise all at once, is the Maximum-Likelihood Estimate (MLE) [10]. In that approach, observed data  $g$  are treated as a

set of random variables defined over the square lattice. Next, the estimator  $\hat{f}_M$  is calculated as an argument  $\hat{f}$ , for which likelihood function  $p(\cdot)$  reaches its maximum value [11, 20]:

$$\hat{f}_M = \arg \max_f p(g | \hat{f})$$

In general, image recovery is a process minimizing the mean square error (MSE) of the  $\hat{f}$  estimator of unobservable  $f$  parameter (a perfect image). The functional may be written as [11]:

$$E = \frac{1}{2} \|g - (psf) \hat{f}\|^2 + \frac{1}{2} \lambda \|D\hat{f}\|^2 \quad (2)$$

where:

- $\hat{f}$  – estimate of a perfect image,
- $\mu$  – constant,
- $D$  – high-pass filter.

Second part of (2) functional is called the regularization term, and it is related to the noise contained in the image. As processed image contains more noise, value of  $\lambda$  parameter becomes greater, but overestimation of it distorts the  $\hat{f}$  [11]. The process of image restoration using artificial neural networks leads to solution of a quadratic programming (QP) problem, defined as quadratic form [11]:

$$E = -\frac{1}{2} \hat{f}^T W \hat{f} - b^T \hat{f} + c \quad (3)$$

where:

- $W$  – matrix, in which each  $(i, j)$  element is a weight between  $i$ -th and  $j$ -th neuron,
- $b$  – representing the bias of each neuron,

-  $c$  – constant.

Comparing quadratic form (3) with functional (2), parameters  $W$ ,  $b$  and  $c$  are functions of  $psf$ ,  $D$  and  $\lambda$ ,  $\mu$  variables, respectively [11].

In paper [21], authors have presented the method using an artificial neural network, in which, for an image containing  $S$  greyness levels, each pixel is represented by  $S$  neurons. The aggregation of input data is realized by the following formula:

$$u_i = \sum_{k=1}^S \sum_{j=1}^L w_{i,j} v_{i,k} + b_i$$

where  $v_{i,k}$  is state of the  $k$ -th neuron for  $i$ -th image point.

State of the  $i$ -th neuron may be altered to  $\Delta f_i$ , which, in result, updates the quadratic functional (3) [11, 21]:

$$\Delta \hat{f}_i = \text{sgn}(u_i)$$

$$\Delta E = -\frac{1}{2} w_i (\Delta \hat{f}_i)^2 - u_i \Delta \hat{f}_i$$

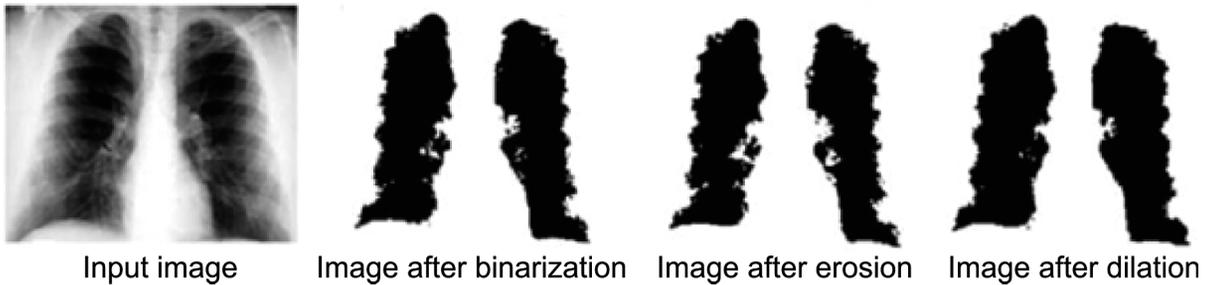


Fig. 4. The result of erosion and dilation process (Input image: <http://www.mp.pl/artykuly/?aid=18226>)

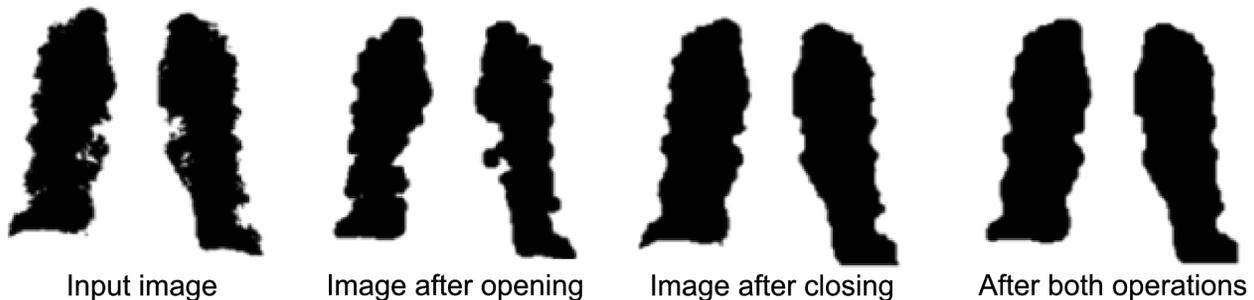


Fig. 5. The result of opening, closing and both operations (opening and closing)

In case of  $\Delta E < 0$ , state of neuron  $v_{i,k}$  is updated. If none of the pixels is altered in comparison to the previous iteration, the algorithm stops.

In fact, distortion adjustment problem solved by artificial neural networks basically means minimization of energy (3) in each direction of  $L$ -dimensional space (as number of dimension of

the space is also a number of points the image contains). In case of negative-defined  $W$  matrix ( $\hat{f}^T W \hat{f} > 0$ , therefore matrix  $-W$  is positive defined), an algorithm achieves the global minimum (local minimas do not exist), as noted in [11]:

$$f = -W^{-1} b$$

**Morphological operations**

Morphological operations are one of the most important transformations in the preprocessing of the digital images. As a result of these transformations both the form, structure and shape of the represented objects changes. The basic idea in mathematical morphology is a structuring element (Fig. 3), which defines the pattern that is compared with a fragment of the processed image [5, 12].

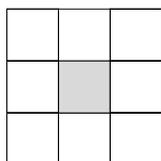


Fig. 3. Structuring element with the highlighted central point

The implementation of morphological transformations for binary images is based on the movement of the structural element throughout the analyzed image and comparing it with the original image for each of pixels. The basic morphological transformations are: erosion, dilation, opening and closing [5, 9].

Erosion, which is a structural element mesh filled with ones, involves the removal of all the points of an image with a value of 1, which have at least one neighbour with a value of 0 [5]. The purpose of erosion is the removal of small, isolated elements and smoothing of disparities at the edges. Figure 4 presents the erosion operation to remove unnecessary elements on the medical images.

Dilation is an operation that is opposite to erosion. Pixels in the output image have value 1 if not all pixels in the neighbourhood are equal to 0. The main task of the dilatation is to close small holes, connect closely located objects, and fusion of larger objects (Fig. 4) [5].

The main disadvantage of both erosion and dilation is the change of the size of the analyzed image surface, which negatively affects especially the analysis of medical images. Erosion reduces the area, while dilation increases it [5]. To get rid of these disadvantages two transformations are introduced, which are combination of the above operations, defined as:

$$\begin{aligned} \text{opening} &= \text{erosion} + \text{dilation} \\ \text{closing} &= \text{dilation} + \text{erosion} \end{aligned}$$

Figure 5 presents the result of opening, closing, and the result of using both operations.

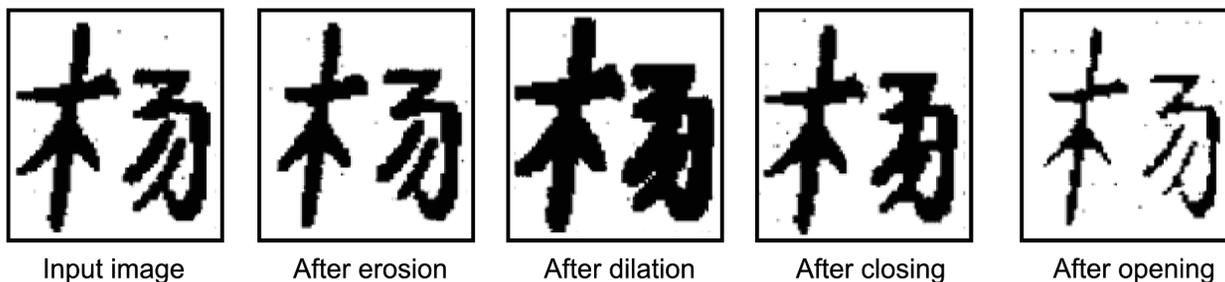


Fig. 6. Example of the DTCNN network performing morphological operations [16]

The most commonly used neural networks that perform the morphological transformations are cellular neural networks CNN. CNN consist of elementary, identical processors called cells. These cells, combined with adjacent units, allow us to implement distributed information processing [15]. To perform the morphological operations different type cellular neural networks such as: fuzzy cellular neural networks FCNN, pulsing cellular neural network PCNN, and discrete time cellular neural networks DTCNN are used [14, 15, 16].

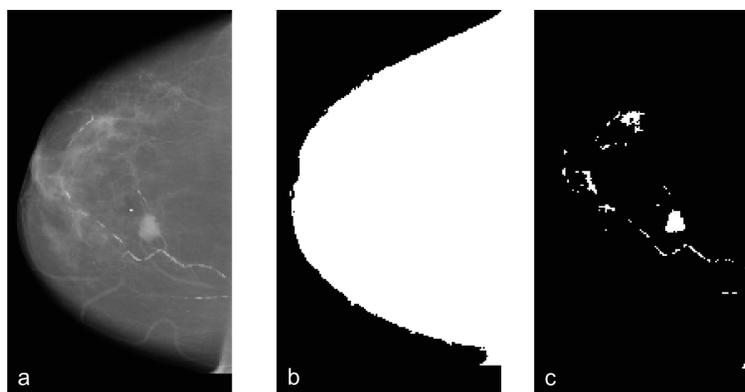
Table 1 presents the most common types of neural networks used to perform morphological operations.

FCNN are the most commonly used neural networks, due to higher effectiveness they mostly replaced the CNN networks [15]. Very interesting and innovative solution has been presented in [16] which uses discrete time cellular neural networks that perform basic morphological transformations. Erosion and dilation are performed by a single layer network, adopting the following parameters for each cell:  $A = 0$ ,  $B = S$  (structural component),  $I = 0$ . The input data is a network of binary images. Opening and

closing operations are performed by a two-layer networks that are build from single layer networks of erosion and dilation. Figure 6 presents the results for the DTCNN network.

Tab. 1. Overview of types of neural networks used for the morphological transformation

#	Neural network	Morphological operations	Applications	Bibliography
1	CNN	Erosion, dilatation	Stereo images	(Mizutani E. et al., 1998) [17], (Roska T., Kek L. 1995) [18]
2	FCNN	Erosion, dilatation, opening, closing	Medical images	(Yang T. et al., 1996) [15]
3	PCNN	Erosion, dilatation	Medical images	(Ranganath H. S. et al., 1997) [19]
4	DTCNN	Erosion, dilatation, opening, closing	Medical images, stereo images	(Yang L. B., 1996) [16]



**Fig. 7.** Example of medical image with changing threshold. The aim of this operation is to extract mammographic microcalcifications on images, which can proceed from tumours in the breast tissue (Input image: Department of Bioinformatics and Telemedicine UJ CM)

## Image segmentation

Segmentation of objects present on digital images is fundamental and inseparable part of every medical computer vision system for analysis and interpreting images. Extraction of areas containing anatomic structures or results of physiological processes allows quantitative or qualitative evaluation and classification of disease entities. Applied approach to segmentation process is dependent on specific problem to be solved, medical presumptions, experts' knowledge and type of medical images, which are at disposal. On the other hand, inherent character of the approach could be based on an grouping object with regard to their intensity, texture, shape and geometric relations to each other [5, 9]. One of the simplest approach to image segmentation is a binarization process, but it is extremely difficult to find a category of medical images, that this method could apply to on its own. Nevertheless, binarization methods based on the adaptive threshold determination, using a moving window, may constitute the part of the system for separation of objects. Recent approaches to the problem of image segmentation possess firm theoretical mathematic background, including fields like: probabilistic models, calculus of variations, fuzzy logic, graph theory and artificial neural networks [22, 23].

### Adaptive threshold binarization

Binarization (thresholding) is one of the basic and most important methods of the image processing. It causes the transformation of the values of each pixel into binary values (zero and one), which defines the object and background. As a result of this operation, the details are removed, and the amount of information about each point of the image is reduced (only two possible values). This allows us to focus on crucial objects. The purpose of binarization is the classification of each pixel in case of their belonging to the objects or to the background, and therefore plays an important role in the analysis and pattern recognition. The main problem is selecting a suitable binarization threshold that would enable distinguishing objects from the background. The

most common technique is the choice of binarization threshold while undertaking the histogram analysis [5, 12]. For images with sharp contrast between the elements, the threshold lies in the valley between the peak intensity of the background, and the peak intensity of the object [5].

The use of a variable threshold different objects in the image can be marked (Fig. 7 a-c).

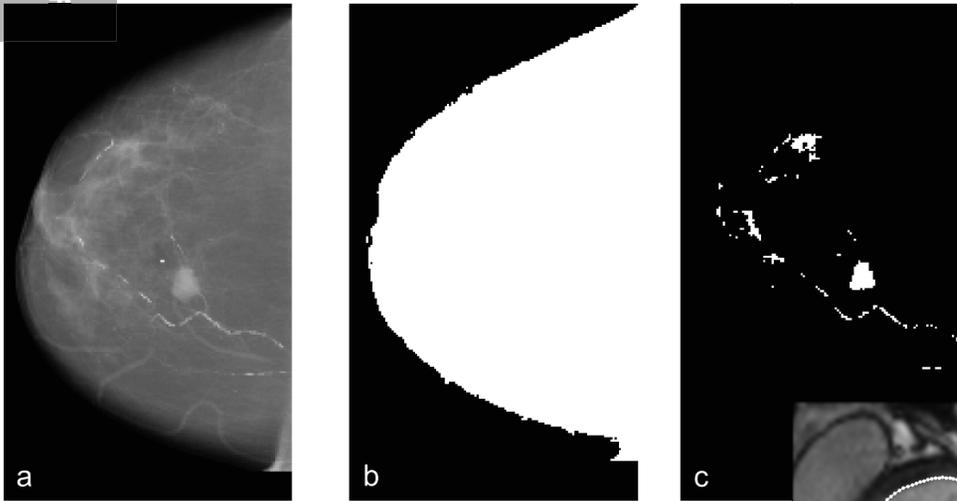
The neural networks that are used for image binarization can be divided into two groups. The first of group includes algorithms that assign pixel values or parts of the image to pre-defined classes of patterns (in the case of binarization we have two classes 0 and 1). The classification is done using the neural network that takes as input a set of calculated parameters. The second group consists of neural networks that determine the local threshold. Important task of this approach is not the automatic choice of the threshold, but the determination of the size and shape of the analyzed part of the image. The most commonly used neural networks for image binarization are self-organizing Kohonen networks, Hopfield networks and recursive MLP networks with backpropagation [25]. Table 2 presents the neural networks that are used in the image binarization.

**Tab. 2. Overview of types of neural networks used for the image binarization**

#	Neural network	Applications	Bibliography
1	Kohonen	Medical image	(Papamarkos N., 2001) [26], (Ntogos N., Veintzas D., 2008) [27]
2	Hopfield	Medical images, holograms	(Just D., Ling D. T., IBM Research Center, 1991) [28], (Ke Z., Gui-Zhong L., 1998) [29]
3	MLP (BP)	Medical images	(Chen T., Takagi M., 2006) [30]
4	CNN	Stereo images	(Cyganek B., Korohoda P., 1999) [31]

The outlined neural network applications have shown great effectiveness in the image binarization. The most significant advantage is the automatic selection of threshold values, resistance to noise, no need to select and limit areas to be analyzed.

Deformable models



There are many methods and algorithms defined with respect to deformable models, e.g. active contour models, deformable surface and templates models [32, 33]. One of the most widely used ones is the active contour method proposed in [34]. An active contour method, along with its variations, has been successfully used in segmentation of organs like: cerebral cortex on MRI images [35], heart valve on echocardiography sequences [36], lungs on X-Rays [37], retinal vessels on ultrasonography images [38], vascular vessel trees on computer tomography images [39] and the left ventricle of the heart on MRI images [40, 41]. The greatest advantages of the active contour algorithm is low sensitivity to distortion, additive noise and acquisition artifacts. Such well-founded benefits make the active contour method a very good approach to the problem of image segmentation, in particular, when used on ultrasonographic images lacking clear boundaries of anatomic structures.

In general, an active contour segmentation is based on deformable curve, which deformations are defined by limitations, internal and external forces (Fig. 8). The process of fitting the curve to the object's boundary is iterative in nature. The method is derived from the definition of the continuous curve, which represents the contour of an arbitrary object. The curve (4) is defined by parametric equation with varying time:

$$\psi(s, t) = (x(s, t), y(s, t)), s \in \langle 0; 1 \rangle \tag{4}$$

The total energy of the model, which represents the curve  $\psi(s, t)$ , is defined as  $E_{snake}$  functional [32, 34]:

$$E_{snake} = \int_0^1 (E_{int}(\psi(s, t)) + E_{ext}(\psi(s, t))) ds \tag{5}$$

where:

- $E_{snake}$  – refers to total energy of the  $\psi(s, t)$  curve,
- $E_{int}$  – denotes internal energy, which bends  $\psi(s, t)$  curve,
- $E_{ext}$  – denotes external energy, which attracts  $\psi(s, t)$  curve to the desired boundary.

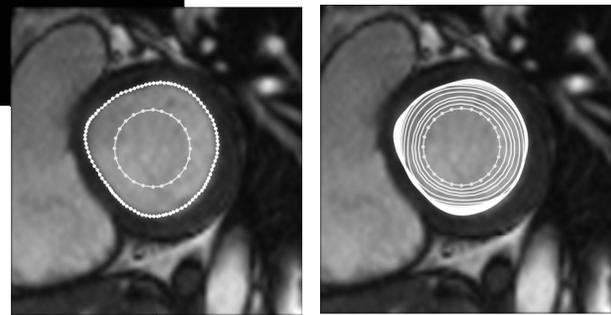
The internal energy  $E_{int}$  of the  $\psi(s, t)$  curve is expressed by weighted sum of partial derivatives of the curve with respect to the  $s$  [32, 34]:

$$E_{int} = \frac{1}{2} \begin{pmatrix} \alpha & \delta\psi^2 & \delta^2\psi^2 \\ \delta s & & \delta s^2 \end{pmatrix} + \beta$$

where:

- $\alpha$  – coefficient of tension,
- $\beta$  – coefficient of rigidity.

In most cases, the external energy is expressed as transformed gradient field of



the image, defined as [42]:

$$E_{ext} = -|\nabla I(x, y)|^2$$

Minimization of functional (5) can be achieved via iterative approach, which changes coordinates of curve's points accordingly to simultaneous linear equations [43]:

$$X^\tau = (I + \gamma A)^{-1} (X^{\tau-1} + \gamma f_x(X^{\tau-1}, Y^{\tau-1}))$$

$$Y^\tau = (I + \gamma A)^{-1} (Y^{\tau-1} + \gamma f_y(X^{\tau-1}, Y^{\tau-1}))$$

where:

- $f_x, f_y$  – denote, respectively, horizontal and vertical components of vector representing external forces.

The task can basically be viewed as an optimization problem. Successive iterations of the algorithm lead to the minimization of curve's energy  $E_{snake}$ , which is a crucial aspect of curve's movement in an image space. The curve deformation is dependent on tuning of model parameters ( $\alpha, \beta, \gamma$ ), initial conditions and selected method of calculating external energy, which brings curve to the contour.

**Fig. 8.** Results of matching curve to the left ventricle contours on the short axis MRI images: the initial conditions and final curve are fitted to the endocardium (left); successive iterations show convergence to the endocardium boundary (right)

Over the years, an active contour method was evolving and subsequent researchers introduced elements of fields like: calculus of variations [42], stochastic processes [44], wavelet analysis [45], fuzzy logic [46], and artificial neural networks. One of the first improvements related to neural networks, was

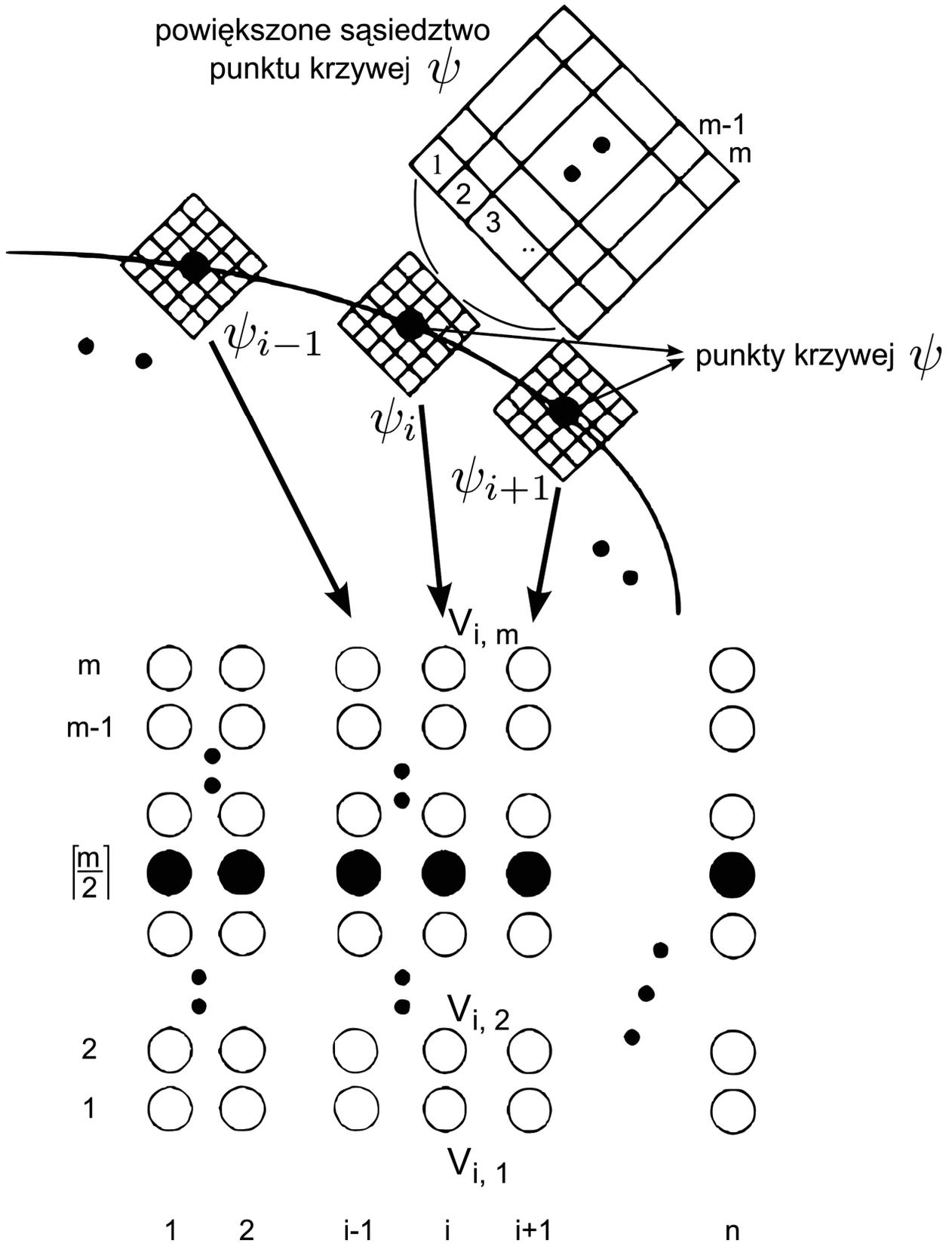


Fig. 9. The mapping scheme of curve's points and neighbourhoods [47]

applying Hopfield network to achieve the goal of minimization of the functional (5) [47]. The curve's points and their neighbourhoods are mapped onto two-dimensional Hopfield network with  $\times$  mutually connected neurons ( $n$  – denotes number of points,

$m$  – size of their neighbourhoods) (Fig. 9). Let  $v_{i,k}$  mean binary state of  $(i, k)$  neuron,  $T_{i,k,j,l}$  as  $(i, k)$  and  $(j, l)$  weight of the connection, and  $I_{i,k}$  as a constant (bias). The aggregation of input data of  $(i, k)$  neuron is given by formula:

$$u_{i,k} = \sum_{j=1}^n \sum_{l=1}^m T_{i,k,j,l} v_{j,l} + I_{i,k} \quad (6)$$

Output signal  $v_{i,k}$  for  $(i, k)$  neuron is calculated via rough thresholding:

$$v_{i,k} = g(u_{i,k}) \quad (7)$$

$$g(x) = \begin{cases} 1 & \text{dla } x \geq 0 \\ 0 & \text{dla } x < 0 \end{cases}$$

An iterative approach to energy minimization (5) and, by extension, convergence of the curve to desirable anatomic structure contour, is basically described as randomized selection of neuron, data aggregation accordingly to the formula (6) and evaluation of output function based on rough thresholding (7). The procedure is repeated, until number of neurons, which change in the set of random draws, exceeds specified threshold. Otherwise, the curve converges to the boundary of segmented object.

Method of energy minimization (5) is numerically stable (in the classical approach, when parameters are in certain range, the curve might oscillate), additionally minimization of  $E_{snake}$  energy can be parallelized.

### Image segmentation based on Pulse-Coupled Neural Networks

The inspiration for model of Pulse-Coupled Neural Networks (PCNN) comes from cat's visual cortex [48]. PCNN used for image processing task is two-dimensional network, which has mutually connected neurons. Each neuron corresponds to single pixel of the image and consist of three parts: the dendritic tree, modulation field and pulse generator. The dendritic tree (input layer) includes two types of inputs: feeding and linking inputs. While linking input accepts values of output signals from other neurons in previous iteration, the feeding input accepts signals from other neurons as well as external stimulus. Then the feeding and linking outputs are non-linearly combined together in modulation field as an internal neuron activation. At last, the pulse generator controls the dynamic threshold level [4, 52].

The Pulse-Coupled Neural Network has been successfully applied to many medical image segmentation processes e.g. extraction of rodent brains [49], blood cell segmentation [50] as well as human brains, abdominal regions and lungs [51].

There is also proposed hybrid image segmentation model, which conjoin PCNN network and Expectation-Maximization (EM) algorithm, and is called EM-PCNN method [52]. The Expectation-Maximization algorithm estimates the distribution parameters for a white matter, gray matter and cerebrospinal fluid of the image intensity histogram. After that, these parameters are used as a fitness function to be evaluated a number of times for each tissue.

Apart from segmentation process, there are many other applications of PCNN networks, including image filtering and enhancement, data compression, image fusion and feature extraction [4, 53].

An interested reader can look for special issue of *IEEE Transactions on Neural Networks*, Vol. 10, Issue 3, 1999 journal which focused on PCNN networks.

### Conclusions

When making decision to use artificial neural networks to support processing and analysis of medical images, one should be aware, what their capabilities, scope of use and possible consequences of application are. Firstly, it is to be decided, what kind of data are meant to be processed. Depending on whether these are going to be vectors of characteristics, or points acquired directly from

the image (e.g. image recovery), or neural networks are meant to play a complementary role (like in active contour method), it is necessary to properly define structure of the network and target function. One also has to take into consideration potential size of the neural network structure and, in particular, amount of data needed in the learning process. In case of segmentation of an actual organ, number of markers representing the curve in active contour method might reach several hundreds. Additionally, it is advised to carefully pick the size of neighbourhood, because corresponding increase in number of network nodes may be of several orders of magnitude. In spite of aforementioned drawbacks, interest in using artificial neural networks does not fade away, in fact, as shown on figure 1, throughout the last decade it has a slightly upward trend.

## Acknowledgment

This work is funded by the AGH University of Science and Technology as a research project No. 11.11.120.612.

## References

- Mikrut Z., Tadeusiewicz R.: Sieci neuronowe w przetwarzaniu i rozpoznawaniu obrazów. In: Biocybernetyka i Inżynieria Biomedyczna 2000, t. 6: Sieci neuronowe. Warszawa: EXIT, 2000, pp. 459-493 (in Polish).
- Shi Z., He L.: Application of Neural Networks in Medical Image Processing. Proceedings of Second International Symposium on Networking and Network Security, 2010, pp. 23-26.
- Egmont-Petersena M., de Ridder D., Handels H.: Image processing with neural networks – a review. *Pattern Recognition* 2002, 35: 2279-2301.
- Lindblad T., Kinser J.M.: Image Processing Using Pulse-Coupled Neural Networks. Heidelberg: Springer-Verlag, 2005.
- Tadeusiewicz R., Korohoda P.: Komputerowa analiza i przetwarzanie obrazów. Kraków: Wydawnictwo Fundacji Postępu Telekomunikacji, 1997 (in Polish).
- Socha M., Duda K., Zieliński T.P., Duplaga M.: Algorytmiczna korekcja zniekształceń geometrycznych kamery bronchoskopu. XV Sympozjum Modelowanie i Symulacja Systemów Pomiarowych, 2005, pp. 219-228 (in Polish).
- Pawliczek P., Romanowska-Pawliczek A., Soltys Z.: Parallel Deconvolution of Large 3D Images Obtained by Confocal Laser Scanning Microscopy. *Microscopy Research and Techniques* 2010, 73: 183-194.
- Banham M.R., Katsaggelos A.K.: Digital Image Restoration. *IEEE Signal Processing Magazine* 1997, 14, 2: 24-41.
- Gonzalez R.C., Woods R.E.: Digital Image Processing. Upper Saddle River, NJ: Pearson Prentice Hall, 2008.
- Zhang J., Katsaggelos A.K.: Image Recovery Using EM Algorithm. In: Madiseti V.K., Williams D.B. (eds), Digital Signal Processing Handbook. Boca Raton, FL: CRC Press, 1999, pp. 29.1-29.26.
- Guan L., Perry S.W., Wong H.S.: Adaptive Image Processing: A Computational Intelligence Perspective. Boca Raton, FL: CRC Press, 2001.
- Tadeusiewicz R., Śmietański J.: Pozyskiwanie obrazów medycznych oraz ich przetwarzanie, analiza, automatyczne rozpoznawanie i diagnostyczna interpretacja. Kraków: WSTN, 2011 (in Polish).
- Osipowicz K.: Budowa komórkowych sieci neuronowych i ich zastosowania do rozpoznawania obrazów. *Software 2.0* 2002, 2 (in Polish).
- Duch W., Korbicz J., Rutkowski L., Tadeusiewicz R. (red.): Biocybernetyka i Inżynieria Biomedyczna 2000, t. 6: Sieci Neuronowe. Warszawa: EXIT, 2000 (in Polish).
- Yang T., Yang L.-B., Wu C.W., Chua L.O.: Fuzzy cellular neural networks: applications. 4th International Workshop on Cellular Neural Networks and Their Applications, Seville, Spain, 1996, pp. 225-230.
- Yang L.B.: Implementation of binary mathematical morphology using Discrete-Time Cellular Neural Networks. 4th International Workshop on Cellular Neural Networks and Their Applications, Seville, Spain 1996, pp. 7-12.
- Mizutani E., Kozek T., Chua L.: Roadway lane marker extraction by motion detection CNNs. Proc. IJCNN, Alaska, 1998, Vol. 1: 503-508.
- Roska T., Kek L.: CSL-CNN Software Library: Templates and Algorithms (Version 6.4), DNS-CADET-15, MTA Sz-TAKI, Hungarian Academy of Sciences, Budapest, 1995.
- Ranganath H.S., Kuntimad G., Johnson J.L.: A neural network for image understanding. Handbook of Neural Computation. Bristol: IOP Publishing, 1997.
- Winkler G.: Image Analysis, Random Fields and Markov Chain Monte Carlo Methods. A Mathematical Introduction, Berlin, Heidelberg: Springer-Verlag, 2006.
- Zhou Y.T., Chellappa R.W., Vaid A., Jenkins B.K.: Image Restoration Using a Neural Network. *IEEE Transactions on Acoustics, Speech and Signal Processing* 1998, 36, 7: 1141-1151.
- Suri J.S., Setarehdan S.K., Singh S.: Advanced Algorithmic Approaches to Medical Image Segmentation. State-of-the-Art Applications in Cardiology, Neurology, Mammography and Pathology. London: Springer, 2002.
- Wismueller A.: Segmentation with Neural Networks. In: Bankman I.N. (ed.), Handbook of Medical Image Processing and Analysis. Amsterdam: Elsevier, 2009, pp. 113-143.
- Chaira T., Ray A.K.: Fuzzy Image Processing and Applications with MATLAB. Boca Raton, FL: CRC Press, 2009.
- Ogiela M., Tadeusiewicz R.: Modern Computational Intelligence Method for the Interpretation of Medical Images. Berlin, Heidelberg: Springer-Verlag, 2008.
- Papamarkos N.: A Technique for Fuzzy Document Binarization. 10th International Conference on Information and Knowledge Management, ACM Symposium on Document Engineering, Atlanta, 2001, pp. 152-156.
- Ntogos N., Veintzas D.: A binarization algorithm for historical manuscripts. 12th WSEAS International Conference on COMMUNICATIONS, Greece, 2008, pp. 43-51.

28. Just D., Ling D.T.: Neural networks for binarizing computer-generated holograms. *Optics Communications* 1991, 81, 1-2: 1-5.
29. Ke Z., Guizhong L.: Binarization processing for blurring edge image with Hopfield network. *Journal of Electronics & Information Technology* 1998, 20, 1: 38-43.
30. Chen T., Takagi M.: Image Binarization by Back Propagation Algorithm, ISPRS, 1991, pp. 345-349.
31. Cyganek B., Korohoda P.: Improved neural network for fast extraction of information from stereo-images. Proc. 4th Conf. Neural Networks and their Applications, Zakopane, 1999.
32. McInerney T., Terzopoulos D.: Deformable Models. In: Bankman I.N. (ed.), *Handbook of Medical Image Processing and Analysis*. Amsterdam: Elsevier, 2009, pp. 145-166.
33. Szczypiński P.M.: Modele deformowalne do ilościowej analizy i rozpoznawania obiektów w obrazach cyfrowych, Wydział Elektrotechniki i Elektroniki Politechniki Łódzkiej, Łódź, 2000, [http://www.eleletel.p.lodz.pl/pms/Doktorat\\_2000.pdf](http://www.eleletel.p.lodz.pl/pms/Doktorat_2000.pdf) (in Polish).
34. Kass M., Witkin A., Terzopoulos D.: Snakes: Active contour models. *International Journal of Computer Vision* 1988, 1: 321-331.
35. Xu C.: Deformable Models with Application to Human Cerebral Cortex Reconstruction from Magnetic Resonance Images. PhD thesis, The Johns Hopkins University, 1999, [http://vismod.media.mit.edu/pub/elwin/this/cheyang\\_xu\\_thesis.pdf](http://vismod.media.mit.edu/pub/elwin/this/cheyang_xu_thesis.pdf).
36. Shang Y., Yang X., Zhu M., Hin B., Liu M.: Prior Based Cardiac Valve Segmentation in Echocardiographic Sequences: Geodesic Active Contour Guided by Region and Shape Prior, *Pattern Recognition and Image Analysis. Lecture Notes in Computer Science* 2005, 3523: 447-454.
37. Silveira M., Marques J.: Automatic segmentation of the lungs using multiple active contours and outlier model. 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2006, pp. 3122-3125.
38. Al-Diri B., Hunter A., Steel D.: An Active Contour Model for Segmenting and Measuring Retinal Vessels. *IEEE Transactions on Medical Imaging* 2009, 28, 9: 1488-1497.
39. Deklerck R., Nyssen E., Markova A., de Mey J., Yang X., Sun K.: Vascular Active Contour for Vessel Tree Segmentation. *IEEE Transactions on Biomedical Engineering* 2011, 58, 4: 1023-1032.
40. Pieciak T.: Myocardial Segmentation Based on Magnetic Resonance Sequences. *Bio-Algorithms and Med-Systems* 2010, 6, 12: 85-90.
41. Lee H.-Y., Codella N.C.F., Cham M.D., Weinsaft J.W., Wang Y.: Automatic Left Ventricle Segmentation Using Iterative Thresholding and an Active Contour Model with Adaptation on Short-Axis Cardiac MRI. *IEEE Transactions on Biomedical Engineering* 2010, 57, 4: 905-913.
42. Xu Ch., Prince J.L.: Snakes, Shapes and Gradient Vector Flow. *IEEE Transactions on Image Processing* 1998, 7, 3: 359-369.
43. Acton S.T., Ray N.: *Biomedical Image Analysis. Segmentation, Synthesis Lectures on Image, Video & Multimedia Processing*. San Rafael, CA: Morgan & Claypool, 2009.
44. Pluempitiwiriwaj Ch., Moura J.M.F., Lin Wu Y.-J., Ho Ch.: STACS: New Active Contour Scheme for Cardiac MR Image Segmentation. *IEEE Transactions on Medical Imaging* 2005, 24, 5: 593-603.
45. Wu H.-H., Liu J.-Ch., Chui Ch.: A Wavelet-Frame Based Image Force Model for Active Contouring Algorithms. *IEEE Transactions on Image Processing* 2000, 9, 11: 1983-1988.
46. Krinidis S., Chatzis V.: Fuzzy Energy-Based Active Contours. *IEEE Transactions on Image Processing* 2009, 18, 12: 2747-2755.
47. Tsai C.-T., Sun Y.-N., Chung P.-C.: Minimising the energy of active contour model using a Hopfield network. *IEE Pro-*

# INSTRUCTION FOR AUTHORS

## The scope

The edition of the journal *Bio-Algorithms and Med-Systems* was initiated to open the discussion in the interdisciplinary scientific activity focused on medicine in both: experimental and theoretical research including the computer science application to diagnostics, therapy and education. The molecular level oriented problems (*Bio-Algorithms*) as well as technical applications to the hospital use (*Med-Systems*) are in the focus of our interest. The journal is open to cover variety of interests.

The journal is edited quarterly by *Jagiellonian University, Medical College, Kraków, Poland*.

Each corresponding Author is also asked to give information about potential reviewers .

Manuscript should be prepared in English. Usage of correct language is the responsibility of the Author.

Journal *Bio-Algorithms and Med-Systems* is registered in SCOPUS database monitoring citations of publications.

The procedures of registration in other databases have been undertaken.

Each submitted material (manuscript) undergoes the peer review by two reviewers. Author is asked to answer to the reviewers comments and/or follow the advice correcting the paper according to reviewers' suggestion.

Polish Ministry of Science and Education journal rating: 6.000.

Submission of manuscript to *BAMS* implies that the copyright of the entire material except those already copyrighted by other publications like tables, figures etc. and those permission the Author(s) have already obtained, belongs to the Publisher.

The text shall be prepared in \*.docx (preferable) or \*.doc format with good quality of graphic of figures.

The article organization:

1. Title, Authors names, affiliation, key words (max 6)
2. Abstract – no more than 250 words
3. Introduction – describing the Interests and the main idea of the paper in relation to other publications introducing the Reader to the current Known How
4. Materials and Methods – the description of data sources, tools applied for the work
5. Results – presentation of the new original achievements
6. Conclusions/Discussion – presentation of the further plans, the relation to previous works to make possible the comparison of the results with others. The critical comments are also expected.
7. Acknowledgements – the source of founding and information as to the personal collaboration and discussion with specialists not included in Authors' list
8. References: the last section of the paper presenting the references in order of their appearance in text in form [1,2].

Examples:

[ 1 ] Pyrczak W, Sarapata K: Instructions ..., *Bio-Algorithms and Med-Systems* 1, 386-387, 2005.

[ 2 ] Tadeusiewicz R, Ogiela M. R. *Medical Image ...*, Springer Verlag, Heidelberg, 2004

9. Tables and Figures should be numbered consecutively using Arabic numbers.
10. Color Figures – Cost of printing color figures will be charged to the Authors at the rate of U.S. \$ 125 per plate of four color figures.
11. Reprints – in order for reprints should be placed at the time of submission of the article itself, so that it may be out for printing without loss of time, if found suitable for order for not less than 100 reprints is to be made.

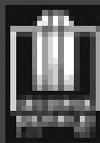
## MANUSCRIPT SUBMISSION

**The only way to submit the paper possible is through the e-mail address [bams@cm-uj.krakow.pl](mailto:bams@cm-uj.krakow.pl).**

The electronic service *submission\_office* (in preparation) will be available soon on the webpage [www.bams.cm-uj.krakow.pl](http://www.bams.cm-uj.krakow.pl)

The corresponding Author will/may be asked to submit:

1. cover letter presenting the information usable for the Editorial Board explaining the accordance of the paper with the specialization field represented by the Journal.
2. list of potential reviewers (together with the email addresses)
3. the main body paper
4. figures (tables shall be put into the main body)
5. ethic and copy right transfer questions will be asked



www.abokbedu.com.pl

ISSN 1898-9661

# 3<sup>rd</sup> Polish Conference and Working Meeting Biomedical Engineering – Education

www.abokbedu.com.pl

24-25 May 2012

Polish Society of  
Biomedical  
Engineering in  
Cooperation with  
PABE and  
PABE

Cracow



## OKIBEDU

2012

Polish Society of  
Biomedical  
Engineering in  
Cooperation with  
PABE and  
PABE

## invitation

The conference is organized by the Polish Society of Biomedical Engineering in cooperation with the Polish Association of Biomedical Engineers (PABE) and the Polish Society of Biomedical Engineering (PABE).

### Workshops

- Biomedical Engineering in Education
- Biomedical Engineering in Industry
- Biomedical Engineering in Research

### Workshops

- Biomedical Engineering in Education
- Biomedical Engineering in Industry
- Biomedical Engineering in Research

### Workshops

- Biomedical Engineering in Education
- Biomedical Engineering in Industry
- Biomedical Engineering in Research

The conference is organized by the Polish Society of Biomedical Engineering in cooperation with the Polish Association of Biomedical Engineers (PABE) and the Polish Society of Biomedical Engineering (PABE).

The Polish Society of Biomedical Engineering (PABE) is a non-profit organization that promotes the development of biomedical engineering in Poland. It is a member of the International Federation of Medical and Biological Engineers (IFMBE) and the European Association of Biomedical Engineers (EABE). The Polish Society of Biomedical Engineering (PABE) is a non-profit organization that promotes the development of biomedical engineering in Poland. It is a member of the International Federation of Medical and Biological Engineers (IFMBE) and the European Association of Biomedical Engineers (EABE).

### Let's meet together in Cracow!

Cracow, Poland, 24-25 May 2012