Advances in Intelligent Systems and Computing 300

300th Volume of AISC · 300th Volume of AISC · 300th Volume of AISC · 300th Volume of AISC · 300th Vo

Zdzisław S. Hippe
Juliusz L. Kulikowski
Teresa Mroczek
Jerzy Wtorek *Editors*

Human-Computer Systems Interaction: Backgrounds and Applications 3

Springer

Advances in Intelligent Systems and Computing

Volume 300

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

For further volumes:
http://www.springer.com/series/11156

About this Series

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within "Advances in Intelligent Systems and Computing" are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Zdzisław S. Hippe · Juliusz L. Kulikowski
Teresa Mroczek · Jerzy Wtorek
Editors

Human-Computer Systems Interaction: Backgrounds and Applications 3

Springer

Editors
Zdzisław S. Hippe
Department of Expert Systems and Artificial Intelligence
University of Information Technology and Management
Rzeszów
Poland

Juliusz L. Kulikowski
Polish Academy of Sciences
M. Nalecz Institute of Biocybernetics and Biomedical Engineering
Warsaw
Poland

Teresa Mroczek
University of Information Technology and Management
Rzeszów
Poland

Jerzy Wtorek
Gdańsk University of Technology
Gdańsk-Wrzeszcz
Poland

ISSN 2194-5357 ISSN 2194-5365 (electronic)
ISBN 978-3-319-08490-9 ISBN 978-3-319-08491-6 (eBook)
DOI 10.1007/978-3-319-08491-6
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2011936642

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

From the Editors

We are very proud to handle the consecutive book devoted to **H**uman-**C**omputer **S**ystems **I**nteraction (**H-CSI**). The previous monographic volume (**H-CSI**: Backgrounds and Applications **2**, Part I and Part II) received quite good assessment from the scientific community; it also fulfilled our anticipation as a source of up-to-date knowledge in the considered area. This situation encourages us to work out the next volume, giving an insight into the current progress in **H-CSI**. This time however, papers were gathered on our individual invitation of recognized researchers, having significant scientific record in this area. In this way the content of the book contributes the profound description of the actual status of the **H-CSI** field. By chance, it is also a signpost for further development and research.

It is a delightful pleasure to express our gratitude to numerous individual authors, working in: Canada, France, Germany, Indonesia, Ireland, Italy, Japan, Portugal, Malaysia, Mexico, Norway, Slovakia, Slovenia, South Korea, Spain, Syria, Sweden, Tunisia, Turkey, USA and Poland. Many of authors worked together using nets, supplying common articles. In this way, the new volume (**H-CSI: Background and Applications 3**) contains an interesting and state-of the art collection of papers, say reports, on the recent progress in the discussed field.

The contents of the book was divided into the following parts: **I.** General human-system interaction problems; **II.** Health monitoring and disabled people helping systems; and **III.** Various information processing systems.

The general human-system interaction problems (**I**) are presented by papers concerning various application areas, like e.g. brain computers interface systems (A. Materka and P. Poryzała), recognition of emotion (A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch and M.R. Wróbel), recognition of sign language (M. Oszust and M. Wysocki), multimodal human-computer interfaces (A. Czyżewski, P. Dalka, Ł. Kosikowski, B. Kunka and P. Odya), case studies on audience response systems in the computer science course (L. Jackowska-Strumiłło, P. Strumiłło, J. Nowakowski and P. Tomczak), or on global collaboration of students (A.E. Milewski, K. Swigger and F.C. Serce).

Various problems of health monitoring and disabled people helping systems (**II**) are presented in the next group of papers. Many important problems have been touched, for example the detection of sleep apnea by analysis of electrocardiographic signals (P. Przystup, A. Bujnowski, A. Poliński, J. Rumiński and J. Wtorek), the phone recognition of objects as a personal aids for the visually impaired persons (K. Matusiak, P. Skulimowski and P. Strumiłło) or a general research on aiding visually impaired people (S. Yakota, H. Hashimoto, D. Chugo and K. Kawabata**;** M. Yusro, K.M. Hou, E. Pissaloux, K. Ramli, D. Sudiana, L.Z. Zang and H.L. Shi). Besides, in the paper by S. Coradeschi et all, a system for monitoring activities and promoting social interaction for elderly is described.

The group concerning various information processing systems (**III**) consists inter alia of the papers aimed at human life conditions improvement (by A. Astigarraga et all; R. Bianco-Bonzalo et all; P.M. Nauth; S. Suzuki, Y. Fujimoto and T. Yamaguchi).

This book is intended for a wide audience of readers who are not necessarily experts in computer science, machine learning or knowledge engineering, but are interested in Human-Computer Systems Interaction. The level of particular papers and specific spreading-out into particular parts is a reason why this volume makes fascinated reading. This gives the reader a much deeper insight than he/she might glean from research papers or talks at conferences. It touches on all deep issues that currently preoccupy the field of **H-CSI**.

Editors

Zdzisław S. Hippe
Juliusz L. Kulikowski
Teresa Mroczek
Jerzy Wtorek

Contents

Part I: General Human-System Interaction Problems

Part III: Various Information Processing Systems

Part I
General Human-System Interaction Problems

A Robust Asynchronous SSVEP Brain-Computer Interface Based on Cluster Analysis of Canonical Correlation Coefficients

A. Materka and P. Poryzała

Institute of Electronics, Lodz University of Technology, Łódź, Poland,
{andrzej.materka,pawel.poryzala}@p.lodz.pl

Abstract. Brain computer interface (BCI) systems allow a natural interaction with machines, especially needed by people with severe motor disabilities or those whose limbs are occupied with other tasks. As the electrical brain activity (EEG) is measured on the user scalp in those systems, they are noninvasive. However, due to small amplitude of the relevant signal components, poor spatial resolution, diversity within users' anatomy and EEG responses, achieving high speed and accuracy at large number of interface commands is a challenge. It is postulated in this paper that the SSVEP BCI paradigm, combined with multichannel filtering can provide the interface robustness to user diversity and electrode placement. A cluster analysis of the canonical correlation coefficients (computed for multichannel EEG signals evoked by alternate visual half-field LED stimulation) is used to achieve this goal. Experimental results combined with computer simulation are presented to objectively evaluate the method performance.

1 Introduction

The number of "smart" devices and appliances around us grows quickly in the last decades. Not even computers, tablets, cellular phones do comprise a processor with a complex program. Operation and performance of cars, washing machines, microwave ovens, TV sets, etc. strongly depend on the computational power and quality of software of the digital electronic systems embedded in it. Still, the rate of progress in the performance of the computational systems is not accompanied by an equally fast development of the interfaces necessary for information exchange between machines and their users.

In particular, there is a need to develop interfaces that would allow users, who cannot move their limbs, cannot speak, but whose mind operates normally, to enter data into computers without involving the traditional motor pathways of the human nervous system. A solution is a brain-computer interface (BCI) [Wolpaw et al. 2000]. In those interfaces, the intention/will of a user is not expressed by any movement, gesture or command; it is rather "guessed" by the analysis of some measured signals that reflect the brain activity.

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_1, © Springer International Publishing Switzerland 2014

Research projects aimed at development of noninvasive BCI started about 40 years ago. The key factors of focus are speed, number of independent symbols that can be transmitted over the interface and accuracy (lowest error rate). However, due to small amplitude of the signal components, poor spatial resolution, diversity within users' EEG responses, electrode misplacement, and impedance problems its functionality is still far from the expectations. This gives motivation to further research on the interface performance improvement.

In this paper, spatial filtering of the multi-electrode signals is used to make the SSVEP BCI robust to the measurement electrodes displacement and diversity within the operators' EEG responses. The SSVEP paradigm is believed to ensure fastest operation of the interface [Materka and Poryzala 2013]. The asynchronous BCI operation is optimized by identifying best weighted combinations of electrode signals – with the use of cluster analysis of canonical correlation coefficients. Results of experiments with 21 volunteered BCI users are described and discussed to demonstrate the developed method superiority over a number of known alternative techniques.

2 Brain Computer Interfaces

In a brain–computer interface system, users perform mental tasks that invoke specific patterns of brain activity. Those may be invoked by an external stimulation (such as light or sound) or a mental effort of user solely (Fig. 1). The EEG signal is measured, and its relevant features extracted, after necessary preprocessing. A pattern recognition system determines which brain activity pattern a user's brain is producing and thereby infers the user's mental task, thus allowing users to send messages or commands through their intentional brain activity alone. Any particular activity is attributed to a unique symbol transmitted through the interface. The present technological advancement limits applications of the BCIs to a simple cellular phone keyboard with a dozen or so keys or a few-command manipulator for control of a prosthetic or a virtual reality game. The main beneficiaries of the interface are now handicapped persons. It is expected, as BCIs become sufficiently fast, reliable and easy to use, the range of their future applications will encompass many other groups of users.

Most of the phenomena observed in EEG recordings originate in surface layers of the brain cortex, where majority of neurons are positioned perpendicularly to the surface. Due to large number of mutual connections of the cortex neurons, the subsequent waves of depolarization/polarization of their cellular membranes cause synchronization of their activity [Niedermeyer and Silva 2005]. The synchronous activity of a population of nervous cells leads to changes of electric potential on the surface of the cortex, and consequently, on the surface of the skin.

The recording of EEG signal is performed by measuring differences of electric potential between selected points defined on the surface of the human head. Example of standardized locations of the electrodes, defined in 1958 [Oostenveld and Praamstra 2001] to make the measurement points independent of the actual size of the skull is the well-known "ten-twenty" system.

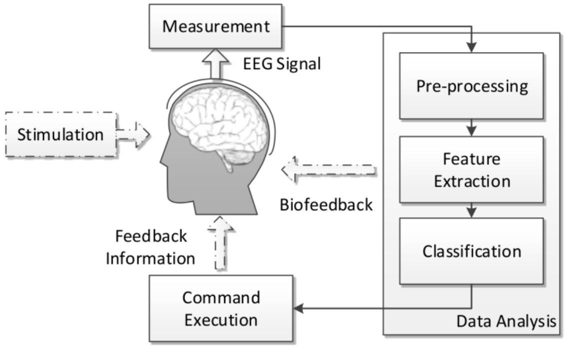


Fig. 1 Basic functional blocks of a brain-computer interface. Optional elements are marked with broken lines

The potential measured on an electrode is a sum of potentials generated by millions of neurons. Thus the measured signals are an average of signals from individual neurons located over some area of the cortex. That is why EEG features poor spatial resolution. Moreover, the potentials of individual neurons have to pass the regions filled with cerebrospinal fluid, bones of the skull and through the skin until finally they reach the electrodes. This causes severe attenuation of the functional waves. The EEG signal that represents electrical brain activity is then very weak, its values are in the range of tens of microvolts. Moreover, the measured signal contains not only the brain activity components of interest. There are other, sometimes many times larger components present (in the order of millivolts), called artifacts. Their sources are of technical or biological origin.

The fact that the EEG signal components that carry the information about the brain activity are weak and are buried in large-amplitude noise makes detection of the BCI user intention difficult. This is the main drawback of the EEG-based brain-computer interfaces. Significant efforts have been taken to design and built EEG measurement devices that would suppress the artifacts and reduce the power of noise relative to the brain signal components of interest [Mason et al. 2007]. One of the latest projects along these lines is described in [Zander et al. 2011]. Advanced signal processing algorithms is another means that leads to reliable detection of the components generated with users intentions.

Four basic categories of noninvasive BCIs have been described in the literature. These categories are related to the brain electrical activity that is invoked, detected and used for sending messages or commands to machine [Wolpaw et al. 2002]. Accordingly, the BCIs use P300 potentials, SSVEP, slow cortical potentials and event-related desynchronization (ERD).

To compare performance of different BCI systems, one should use some standard evaluation criteria [Schlogl et al. 2007]:

- Detection time (a time period between the moment user starts to express their intention to the moment of taking decision by the system).

- Classification accuracy (a ratio of true positive classifications to the sum of true positive, false positive and false negative ones).
- Information transfer rate (bit rate, a parameter used to estimate a theoretical rate of information transfer to the computer) [Kronegg et al. 2005].

The most promising type of the BCI is based on steady-state visual evoked potentials (SSVEP). Relatively large information transfer rate and the number of distinct messages are achieved with the use of the SSVEP-based BCIs [Zhu et al. 2010]. At the same time, high accuracy and speed are obtained at rather small training effort of the user. Thus this type of BCI is the subject of research project discussed in the next Sections.

3 SSVEP BCIs

Most of the SSVEP BCI systems use frequency encoding of the messages. Therefore, detection of potentials generated in result of user's intention is usually based on amplitude or power spectrum analysis (Fig. 2).

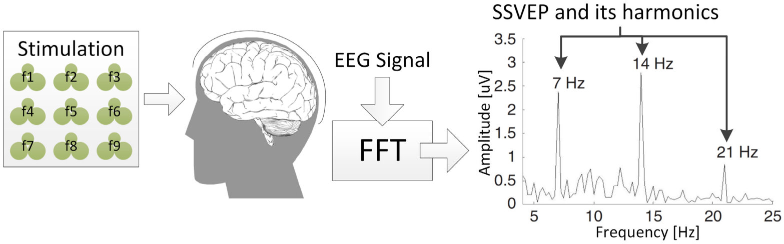


Fig. 2 An SSVEP BCI system with frequency encoding

Referring to Fig. 2, the user concentrates his/her sight on one element (intended to be selected – a target) of the photo-stimulator. Each target is a light source flickering with a unique frequency. There is a message or command attributed to each frequency, so the stimulator plays a role of a virtual keyboard [Materka et al. 2007]. When user focuses his/her attention on a light source of a specified frequency, EEG signals (especially from above primary visual primary cortex) include components of the same frequency and/or its harmonics [Regan 1989]. It is measured over the user's skull and its amplitude spectrum is computed. In the example illustrated by Fig. 2, the user is looking at the stimulator element that is flickering with the frequency of 7 Hz. The SSVEP response is composed of the fundamental frequency, its second and third harmonic, 7 Hz, 14 Hz and 21 Hz, respectively.

In the classical, spectrum analysis based approach, for each stimulation frequency the signal to background ratio (SBR) is computed from the EEG spectrum with the use of Fast Fourier Transform. The background noise could be e.g. the total power of spectrum components in a neighborhood of a given frequency. When the SBR ratio exceed a predefined threshold, a symbol attributed to that frequency of stimulation [Middendorf et al. 2000; Trejo et al. 2006] is decided to be generated at the interface output. In some works, the amplitude of the SBR coefficient is considered a signal feature, which is classified with the use of linear discriminant analysis [Luo and Sullivan 2010]. Other methods include autoregressive spectral analysis [Allison et al. 2008] and wavelet decomposition [Wu and Yao 2008].

The signal-to-background ratio is an essential characteristic of the SSVEP signal. Larger values of SBR lead to shorter time of taking decision and increase the BCI accuracy. Typical stimulators have a form of rectangular fields displayed on an LCD computer screen, each flickering with a different frequency [Cheng et al. 2002]. But it is worthwhile to optimize the visual stimulation to increase the difference between the power of the SSVEP and noise (for e.g. using alternate half-field stimulation method can increase SBR value [Materka and Byczuk 2006]).

Even if the stimulus has been optimized and care has been taken to design measurement equipment as to obtain high signal-to-noise ratio, still the EEG signal is weak and noisy. Then, further signal processing and advanced VEP detection techniques are needed to ensure high accuracy, speed and capacity (i.e. the number of different messages sent over the interface). Taking into account individual anatomical and psycho-physiological differences between users, it is difficult to tell in advance what is the right position for the EEG electrodes to capture most of the information related to BCI users intention. On the other hand, it is impractical to use, say 22 electrodes covering densely the whole skin area on the head. Thus, as a compromise, a limited number of channels (say, 8 electrodes) is considered representative to the problem. The multichannel measurements is a standard now.

It is hypothesized in most research projects that some linear or nonlinear combinations of the channels, individualized for each user, carry the information which is searched for [Cichocki et al. 2008]. An example of obtaining a linear combination (spatial filtering) of the multichannel EEG recordings is shown in Fig. 3.

The optimum linear spatial filter of Fig. 3 should produce new “channels” S for which a ratio of the power of the signal of interest to the noise power is maximum. Among different goals of this procedure, there are Best Bipolar Combination (BBC) of electrodes [Wang et al. 2004], Minimal Energy Combination (of noise), Maximum Contrast Combination (MCC) [Friman et al. 2007] and Canonical Correlation [Bin et al. 2009]. Those multichannel, spatial filtering based detection methods should be (to some extent) immune to small amplitudes of the signal components, poor spatial resolution, diversity within operators’ EEG responses and electrode displacement problems. Optimized, weighted combinations of electrode signals should be identified whenever it is possible.

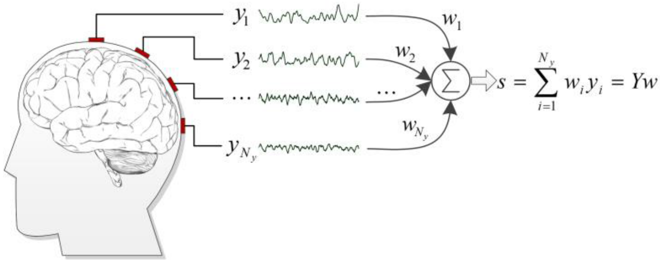


Fig. 3 The concept of spatial filtering of EEG signals

A novel, Cluster Analysis of Canonical Correlation Coefficients (CACC) method for detecting steady-state visual evoked potentials (SSVEP) using multiple channel electroencephalogram (EEG) data has been developed by the authors and described in [Poryzala et al 2012]. Accurate asynchronous detection, high speed and high information transfer rate can be achieved with CACC after a short calibration session. Spatial filtering based on the Canonical Correlation Analysis method proposed in [Bin et al. 2009] was used for identifying optimal combinations of electrode signals that cancel strong interference signals in the EEG. The proposed algorithm, a standard spectrum analysis approach, and two competitive spatial filtering and detection methods were evaluated in a series of experiments with the use of data from 21 subjects [Byczuk et al. 2012]. The obtained results showed a significant improvement in classification accuracy and in an average detection time for a large group of users.

In our recent research we addressed the problem of changing the designed SSVEP-based BCI laboratory demonstration to practically applicable system. Performance of the device evaluated in the carefully controlled lab environment will be decreased in real world conditions, where small amplitudes of the signal components, relatively high power of noise, diversity within users' EEG responses, electrode misplacement, and impedance problems cannot be controlled. Practical device should be convenient and comfortable to use (ideally a limited number of dry, active electrodes should be used) and its performance should be stable and reliable in all possible working conditions [Wang et al. 2008]. Those problems have to be addressed before BCI devices can be put into practical use.

In offline experiments we have evaluated how the misplacement of the measurement electrodes and diversity within users' EEG responses affect the performance of the designed asynchronous Brain-Computer Interface with the CACC detection method (Fig. 4).

As in [Poryzala et al 2012], users were qualified to one of three groups:

Group A (best results, 5 subjects). Subjects who used the device (in our previous studies and tests).

Group B (average results, 10 subjects). Subjects who were not familiar with a BCI device, but actively participated in the experiments.

Group C (poor results, 6 subjects). Subjects with concentration problems or very high unstimulated, spontaneous brain activity.

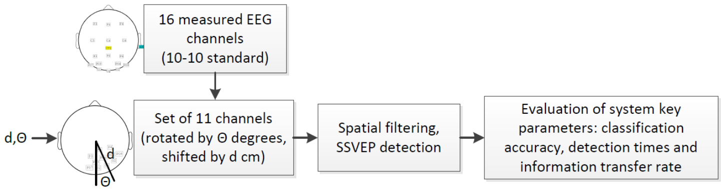


Fig. 4 Effect of displacement of the measurement electrodes on the parameters of the SSVEP based BCI system

For each user from Groups A, B and C, the original, 16-channel EEG data (seven electrodes over the primary visual cortex: PO7, PO3, O1, OZ, O2, PO4 and PO8; nine electrodes evenly distributed over the remaining cerebral cortex: P3, PZ, P4, C3, CZ, C4, F3, FZ and F4) were interpolated (for given displacement defined by shift d and rotation Θ) to the new set of eleven displaced measurement points (Fig. 5a). Data was interpolated both in space and time domain (tessellation-based linear interpolation) in the wide range of rotations (Θ = ±50) and shifts (d = ±4 cm). Rules for d and Θ directions are depicted in Fig. 5b.

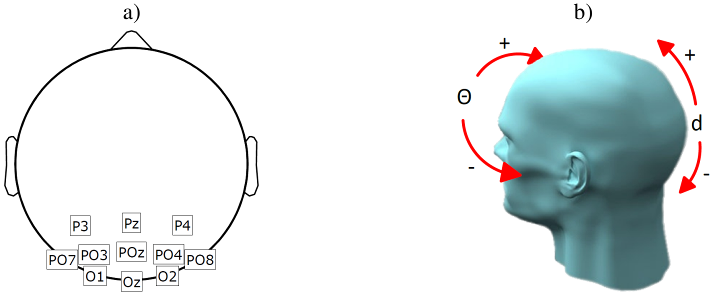


Fig. 5 Set of 11 EEG electrodes for Θ = 0 and d = 0 cm (a). Rules for d and Θ directions (b)

Data for subjects were divided after the interpolation into shorter fragments, containing several stimulation patterns (extracted based on the binary stimulation on and off markers recorded along the original measurement data). The algorithm was evaluated with window length of 2.56 s. Data window step was set to 0.16 s.

Results (classification accuracy, average detection times and information transfer rates) were evaluated in a 5° x 5 cm grid of Θ and d displacement coordinates independently in each group (Fig. 6, Fig. 7 and Fig. 8). Classification accuracy was defined as the number of correctly classified commands relative to the total number of commands classified by the system. Detection time was measured from the moment when the stimulation symbol was switched on to the moment when BCI system detected a command. Information transfer rate (amount of information which can be transferred between the human brain and the BCI system per minute) was defined as:

$$B_t = \frac{60}{T_D}\left(\log_2(N) + P\log_2(P) + (1-P)\log_2\left(\frac{1-P}{N-1}\right)\right), \tag{1}$$

where N denotes the number of commands (5 in case of this particular system), P denotes classification accuracy and T_D denotes average detection time. All system parameters obtained for each user, were averaged in each of the subject groups for every considered misplacement.

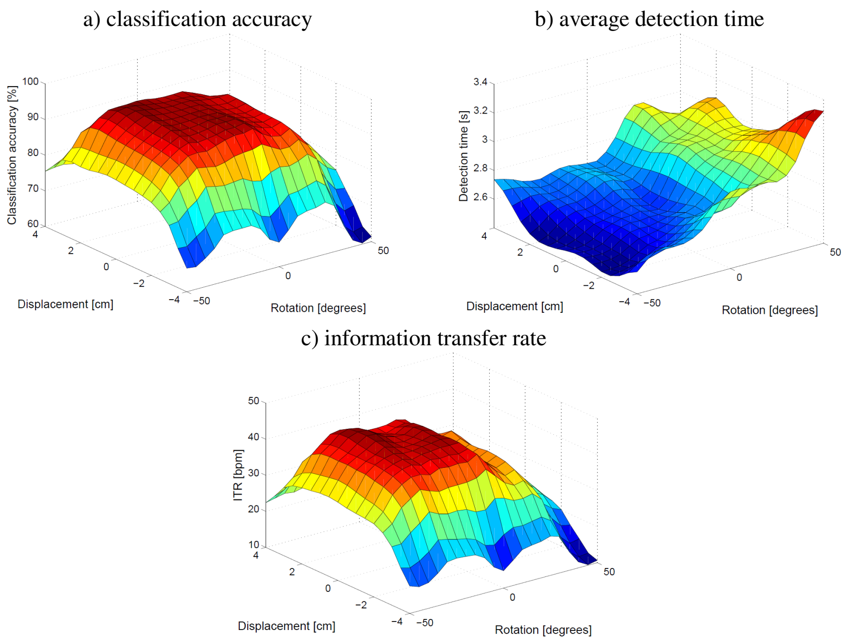


Fig. 6 Parameters of the SSVEP-based BCI system for Group A. Acceptable rotation Θ = -40º - +25º, acceptable shift d = -2.5 - +3.0 cm

Additionally, acceptable rotation and shift values were determined for each case (areas on Θ-d plane, over which classification accuracy does not change by more than ±10% in terms of the value calculated for Θ = 0º and d = 0 cm).

It can be observed in Figures 6, 7 and 8, that the proposed CACC method provides a high tolerance for the SSVEP BCI system electrode placement. Allowable, average misplacement of the electrode set (regardless of the subjects' group), within which none or only limited decrease of the device performance is observed can be defined in proposed displacement coordinates as rotation Θ = ±25º and shift d = ±3.0 cm.

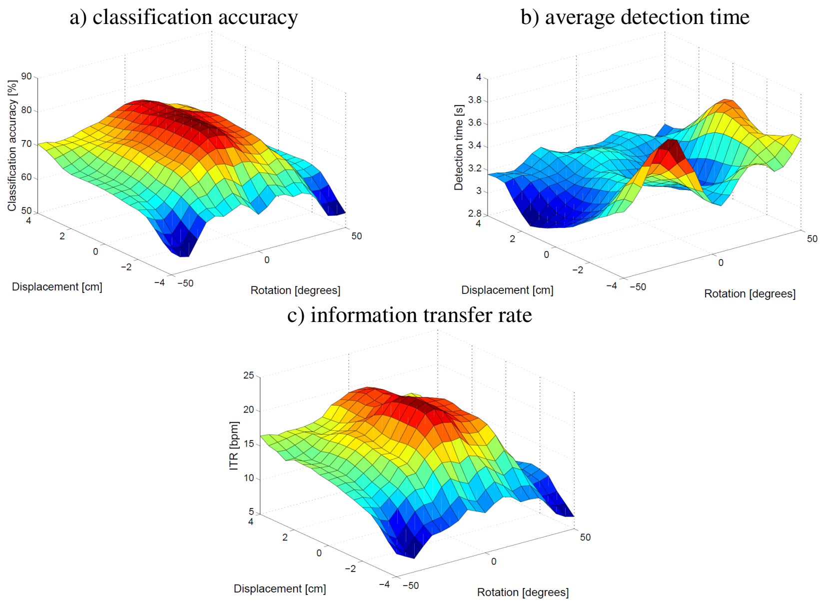


Fig. 7 Parameters of the SSVEP-based BCI system for Group B. Acceptable rotation Θ = -25° - +25°, acceptable shift d = -2.0 - +3.0 cm

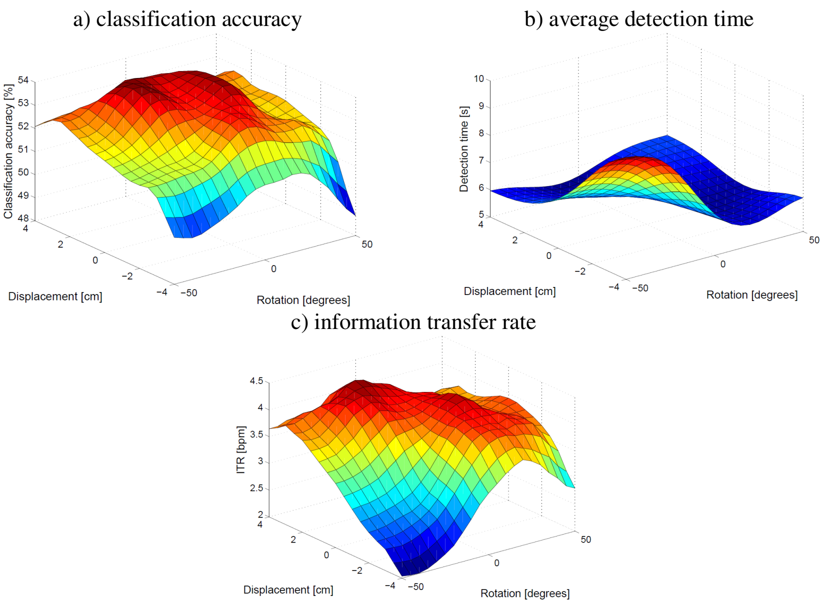


Fig. 8 Parameters of the SSVEP-based BCI system for Group C. Acceptable rotation Θ = -45° - +45°, acceptable shift d = -3.5 - +4.0 cm

4 Summary and Conclusions

It has been shown the SSVEP is a promising paradigm for fast and accurate brain-computer interfaces. The results of our offline experiments demonstrated that the proposed CACC detection method provides stable performance, robustness and reliability in a wide range of measurement electrode misplacements and diversity within users' EEG responses. It is able to identify optimized, weighted combinations of electrode signals and compensates shifts of the electrodes set on top of the subject's head for a large group of users within rotations of ±25° and displacements of up to ±3 cm. This shows its potential to account for individual user anatomical and physiological characteristics. It also proves, that the optimization of SSVEP detection algorithms and their hardware/software implementation for real time SSVEP detection is an important research avenue. But it must be remembered that the BCI research and its various possible applications raise important ethical issues that need to be discussed in different communities to promote acceptance and develop adequate policies [Nijboer et al. 2011].

References

[Allison et al. 2008] Allison, B.Z., McFarland, D.J., Schalk, G., Zheng, S.D., Jackson, M.M., Wolpaw, J.R.: Towards an independent brain-computer interface using steady state visual evoked potentials. Clin. Neurophysiol. 119(2), 399–408 (2008)

[Bin et al. 2009] Bin, G., Gao, X., Yan, Z., Hong, B., Gao, S.: An online multi-channel SSVEP-based brain-computer interface using a canonical correlation analysis. J. Neural Eng. 6(4), 046002 (2009)

[Byczuk et al. 2012] Byczuk, M., Poryzala, P., Materka, A.: On diversity within operators' responses to LED-produced alternate stimulus in SSVEP BCI. Bull. Polish Acad. Sci. 60(3), 7pages (2012)

[Cichocki et al. 2008] Cichocki, A., Washizawa, Y., Rutkowski, T., Bakardjian, H., Phan, A.H.: Noninvasive BCIs: Multiway Signac processing array decompositions. IEEE Computer 10, 34–42 (2008)

[Cheng et al. 2002] Cheng, M., Gao, X., Gao, S., Xu, D.: Design and implementation of a brain-computer interface with high transfer rates. IEEE Trans. Biomed. Eng. 49(10), 1181–1186 (2002)

[Friman et al. 2007] Friman, O., Volosyak, I., Graser, A.: Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces. IEEE Trans. Biomed. Eng. 54(4), 742–750 (2007)

[Kronegg et al. 2005] Kronegg, J., Voloshynovskij, S., Pun, T.: Analysis of bit rate definitions for brain-computer interfaces. In: Int. Conf. Human-Computer Interaction, Las Vegas, NA (2005)

[Luo and Sullivan 2010] Luo, A., Sullivan, T.J.: A user-friendly SSVEP-based brain-computer interface using a time-domain classifier. J. Neural Eng. 7(2), 26010 (2010)

[Mason et al. 2007] Mason, S.G., Bashashati, A., Fatourechi, M., Navarro, K.F., Birch, G.E.: A comprehensive survey of brain interface technology designs. Ann. Biomed. Eng. 35, 137–169 (2007)

[Materka and Byczuk 2006] Materka, A., Byczuk, M.: Alternate half-field stimulation technique for SSVEP-based brain-computer interfaces. Electronics Letters 42(6), 321–322 (2006)

[Materka et al. 2007] Materka, A., Byczuk, M., Poryzala, P.: A virtual keypad based on alternate half-field stimulated visual evoked potentials. In: Proc. IEEE Symp. Information Technology Convergence, Chonbuk, Korea, pp. 297–300 (2007)

[Materka and Poryzala 2013] Materka, A., Poryzala, P.: High-speed noninvasive brain-computer interfaces. In: Proc. 6th Int Conference on Human System Interaction, pp. 7–12 (2013)

[Middendorf et al. 2000] Middendorf, M., McMillan, G., Calhoun, G., Jones, K.S.: Brain–computer interfaces based on the steady-state visual-evoked response. IEEE Trans. Rehabil. Eng. 8, 211–214 (2000)

[Niedermeyer and Silva 2005] Niedermeyer, E., Silva, F.H.L.: Electroencephalography: basic principles, clinical applications, and related fields, pp. 991–1002. Lippincott Williams and Wilkins, Baltimore (2005)

[Nijboer et al. 2011] Nijboer, F., Clausen, J., Allison, B.Z., Haselager, P.: The Asilomar survey: Stakeholders' opinions on ethical issues related to brain-computer interfacing. Neuroethics 6(3), 541–578 (2011)

[Oostenveld and Praamstra 2001] Oostenveld, R., Praamstra, P.: The five percent electrode system for high-resolution EEG and ERP measurements. Clinical Neurophysiology 112, 713–719 (2001)

[Poryzala et al. 2012] Poryzala, P., Materka, A., Byczuk, M.: Cluster analysis of canonical correlation coefficients for the SSVEP based brain-computer interfaces. In: IEEE Conf. NTAV/SPA 2012, pp. 149–154 (2012)

[Regan 1989] Regan, D.: Human brain electrophysiology-evoked potentials and evoked magnetic fields in science and medicine. Elsevier, New York (1989)

[Schlogl et al. 2007] Schlogl, A., Kronegg, J., Huggins, J.E., Mason, S.G.: Evaluation criteria for BCI research. In: Toward Brain-Computer Interfacing, pp. 327–342. MIT Press, Massachusetts (2007)

[Trejo et al. 2006] Trejo, L.J., Rosipal, R., Matthews, B.: Brain-computer interfaces for 1-D and 2-D cursor control: designs using volitional control of the EEG spectrum or steady-state visual evoked potentials. IEEE Trans. Neural Syst. Rehabil. Eng. 14(2), 225–229 (2006)

[Wang et al. 2004] Wang, Y., Zhang, Z., Gao, X., Gao, S.: Lead selection for SSVEP-based brain-computer interface. In: IEMBS 2004, 26th Annual International Conference of the IEEE on Engineering in Medicine and Biology Society, vol. 2, pp. 4507–4510 (2004)

[Wang et al. 2008] Wang, Y., Gao, X., Hong, B., Jia, C., Gao, S.: Brain-computer interfaces based on visual evoked potentials. IEEE Eng. Med. Biol. 27(5), 64–71 (2008)

[Wolpaw et al. 2000] Wolpaw, J., Birbaumer, N., Heetderks, W., McFarland, D.J., Peckham, P.H., Schalk, G., Donchin, E., Quatrano, L.A., Robinson, C.J., Vaughan, T.M.: Brain-computer interface technology: a review of the first international meeting. IEEE Trans. Rehabil. Eng. 8(2), 164–173 (2000)

[Wolpaw et al. 2002] Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain–computer interfaces for communication and control. Clin. Neurophysiol. 113, 767–791 (2002)

[Wu and Yao 2008] Wu, Z., Yao, D.: Frequency detection with stability coefficient for steady-state visual evoked potential (SSVEP)-based BCIs. J. Neural. Eng. 5(1), 36–43 (2008)

[Zander et al. 2011] Zander, T.O., Lehne, M., Ihme, K., Jatzev, S., Correia, J., Kothe, C., Picht, B., Nijboer, F.: A dry EEG system for scientific research and brain-computer interfaces. Front Neurosci. 5, 53 (2011)

[Zhu et al. 2010] Zhu, D., Bieger, J., Molina, G.G., Aarts, R.M.: A survey of stimulation methods used in SSVEP based BCIs. Comput. Intell. Neurosci. (2010); Article ID 702357

Domain Usability, User's Perception

M. Bačíková and J. Porubän

Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics, Technical University of Košice, Slovakia
{michaela.bacikova1,jaroslav.poruban2}@tuke.sk

Abstract. The term 'usability' is generally used today to identify the degree of a user interface[1], application or a device to which it satisfies the user during usage. It is often referred to as "user friendliness" or "software ergonomics". In this paper we argue that usability is formed by two inseparable parts. The first is the *ergonomic usability*, the second aspect we call *domain usability*. During our research we found out, that domain usability is equally important as ergonomic usability, however, it is often neglected by software designers. In this paper we introduce new definitions of understandability and domain usability. Finally the total usability is formed by two aspects – domain and ergonomic. We hope this paper to be a guide or a rule for creating applications that are as close as possible to a domain user. The goal of this paper is to draw attention to domain usability and to stimulate further research in this area.

1 Introduction

> „Current graphical user interfaces are based on metaphors of real world objects and their relations which are well known to anyone from everyday life. Metaphors are presented by the user interface in graphical form as windows, icons and menus".
>
> K. Tilly and Z. Porkoláb [Tilly and Porkoláb 2010]

Usability is often connected with such words as "ergonomics", "human friendliness", user satisfaction with using the application or device, usefulness, effectiveness. Nielsen's definition of usability [Nielsen 1994] is up to this day still used as the cornerstone of defining and evaluating usability and also for placing additional guidelines for creating UIs. But neither the Nielsen's definition nor the known guidelines (such as the ones defined by Badashian [Badashian et al. 2008] or the Java look & feel design guidelines, W3C Web Content Accessibility Guidelines and Android UI Guidelines) explicitly deal with the side of the usability related to the domain content or consistency of UI domain dictionary at all, or they refer to it only in specific boundaries of their context.

[1] It is natural, that the general term *usability* refers also to software applications or devices in general. Our research is however aimed at UIs therefore, when referring to our research, we will refer to the usability of UIs. Our definitions can also be applied generally to any application or device used in a particular domain.

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_2, © Springer International Publishing Switzerland 2014

We think there is a need for defining domain usability as an important attribute of usability. That way at least a basic guide would exist for designing and creating applications, which would correspond to the real world and which would be closer to a domain user. Billman's experiment [Billman et al. 2011] shows the importance of matching the application's terminology with the real world. Billman proved that applications, which have a domain structure better matching the real world, have better usability and thus provide higher performance to their users.

Without the correct terms used in the application's UI, the UI is less usable. Although the UI is really good-looking and ergonomic, if the users do not understand the labels of buttons or menu items, they cannot work with it and hence the whole application is useless. Consequently the domain usability is of a great importance and it can be the decisive point between the application success and failure.

Currently there is a huge amount of applications, which differ not only by their appearance, but also by the terminology used. Even different systems in one specific domain differ in their textual content. During the design and implementation software designers and programmers usually aim to create good-looking and perfectly error-free applications. They arrange the UI components effectively so the end users would not be restrained in their work. The well-known rule of thumb is as follows: "The application should assist the user while performing their work, not getting in the way of it". Programmers however often have a different perspective of how to work with the application in opposite to domain users. Programmers are often more experienced in working with computers and they have their established style of work. But from the domain point of view, they often-times have only a little knowledge about the specific domain, for which the application is developed, because usually they are not domain experts. Thus oftentimes they are not capable of transferring the domain terms, relations and processes correctly into the application.

To summarize our knowledge we identified the main problems as follows:

- There are no clear rules for designing the **term structure** of an application, so it would correspond with the domain.
- There are no official guidelines describing applications, which should match the real world or map domain **terms** and **processes**.
- Even if there were any guidelines and rules, the variety of human thinking, ambiguity and diversity of **natural language** represents a problem when evaluating the correctness of the terminology of applications.

To solve the first two problems, we strive for defining domain usability and introduce examples to illustrate the domain usability definition. We realize that defining the domain usability is not and will never be exact, because of the ambiguity and diversity of the natural language and variety of each person's thinking. It can however serve as a guide or rule for creating applications in a manner, that they would be as close as possible to a domain user.

The goal of this paper is to define and explain domain usability and thus to: i) point out to the problem of the existence of UIs, which are created without respect to their domain; ii) to draw attention to the importance of domain usability; and iii) to stimulate, as much as possible, the research in this area and the creation of domain usability evaluation methods and tools.

This research was not a standalone idea. Our general research area is *automatized domain usability evaluation (ADUE) of UIs*. During the previous three years we conducted an extensive analysis in the area of automatic usability evaluation and semantic UIs and we conducted a research in ADUE. Currently we are preforming experiments in the area of automatized formalization of UIs and automatized domain analysis of UIs which is a presumption for ADUE. Our DEAL extraction tool and its potential for ADUE was described in [Bačíková and Porubän 2013]. During our research we determined that without the proper domain usability definition, heuristics for ADUE cannot be defined. Based on our research and experience in these areas, we argue for this definition.

The contributions of this paper are:

- Identifying the main problems associated with domain application UIs,
- Providing a new definition of domain usability,
- Identifying domain usability in the context of the general usability definition,
- Supporting the creation of applications, which better match the real world,
- Stimulating the research in the area of ADUE.

2 Original Definition of Usability

Usability was first defined by Nielsen [Nielsen 1994] as a whole (but diverse) property of a system, which is related to these five attributes: *Learnability, Efficiency, Memorability, Errors* and *Satisfaction*. Although different usability guidelines have been evolving through time, the usability definition remained unchanged since Nielsen first defined it in 1994 and it still serves as a fundamental guide to create usable software systems and to create new usability guidelines and usability evaluation and testing systems.

3 Ergonomic vs. Domain Usability

The common perception of usability is usually in the terms of user experience, satisfaction with using the application, application quality and effectiveness. Often it is seen from the ergonomic point of view and the domain aspect is neglected or omitted, even if it is included in the definition.

Each software system is developed for a concrete domain therefore its UI should contain *terms*, *relations* and describe *processes* from this specific domain for the user to be able to work with it. If the user does not *understand* the terms in the system's UI, then the whole application is less usable. This application feature can be called understandability. Based on our experience and research and pursuing the existing current work in the area of usability, we will define understandability as follows:

Understandability is the property of a system that affects usability and which relates to the following factors:

- *Domain content*: the UI terms, relations and processes should match the ones from the domain, which the UI is designed for.
- *Adequate level of specificity*: the UI made for a specific domain should not contain terms too general, even if they belong to a parent domain. On the other hand, the terms should not be too specific, if the system is used by people from a more general domain.
- *Consistency*: words used throughout the whole UI should not differ, if they describe the same functionality, the dictionary should be consistent.
- *Language barriers*: there should be no foreign words, the translation should be complete. The language of the UI should be the language of the user.
- *Errors*: a UI should not contain stylistic and grammatical errors.

We can use the term ***domain usability*** to describe the aspect of usability, which is affected by the factor of UI understandability. Although domain usability is affected by understandability, it is not true that understandability = domain usability. Understandability can affect other attributes besides domain usability, for example accessibility.

In the context of usability, understandability can also be perceived as the *relation between the user* (his language) *and the product* (the system).

In the end the *overall usability* can be defined as a connection of two basic aspects: ergonomic usability and domain usability. These two types can be combined together when evaluating usability. Consider a test of the number of steps needed to execute a particular task as an example: a user gets a task which he should execute on two different UIs made for the same domain. Both ergonomic and domain factor affect the completion of the task.

Nielsen's definition may also, in a certain context, involve aspects of domain usability, which can be identified in the following attributes:

- *Learnability*: A system is easier to learn if it contains the proper terms, known to its users. If the correct terms, relations and processes are described by the system, then the users remember the actions better.
  If a UI would not contain any terms, then a user would remember the sequence of steps needed to perform a task as a sequence of clicks on different graphical UI components. The user would remember these sequences as a visualization of these graphical elements and their sequence. However, if the UI also contains the right terms and their sequence is correct (describing a real task in practice), then this sequence is remembered by user not only in the graphical form, but especially as a sequence of terms (e.g. *File* → *Open* → find a file → *OK*) which the user is looking for in the UI when performing the task. This implies that the combination of *both* graphical *and* textual form is more memorable when compared to *only* graphical.

- *Efficiency*: The better the users remember a sequence of steps needed to perform a task, the more efficiently they can perform their work.
- *Memorability*: This aspect was already described in connection to learnability. There are two types of people: people who primary remember things visually and people who remember the actual content. To provide them the combination of both is always better than to give them only one of them. If a system uses the terms known to a domain user and it has the correct positioning of components and good visual properties, the user can choose to remember one of them to be able to find them faster. If the terms in the UI are not known to the user, then they are harder to remember compared to the previous case, because the user has to remember only the positions and appearance of the components without the possibility to choose the other attribute to remember (terms).
- *Errors*: Errors can be both ergonomic and contextual.
- *Satisfaction*: Since UIs encapsulate both textual and graphical aspects, the overall impression is influenced by both aspects. The good looking system could be pleasant to use, but incorrect terms disrupt the user experience. On the contrary, the system could contain the right terms, but if it is ugly or not pleasant to use, the users are less satisfied.

Both ergonomic and domain are two parts affecting the overall usability and Nielsen's definition perfectly covers both of them. However, we argue that domain usability is hidden in the Nielsen's definition and that is the reason why it is often omitted by software designers. Our definition relates directly to domain content of UIs. And based on our definition it is possible to evaluate domain usability - even in an automatized fashion as we have indicated in our feasibility analysis of ADUE in [Bačíková and Porubän 2013].

Since domain usability is a subset of usability, metrics and categorization applicable on general usability can also be applied on domain usability. For example according to Hilbert and Redmiles [Hilbert and Redmiles 2000], domain usability as well as general usability can be divided into *formative* and *summative*.

4 Aspects of Understandability

The individual aspects of understandability will be further discussed in the following subsections along with illustrative examples.

4.1 Domain Content

Imagine a system manufactured for the domain of medicine. Without any explanation or referring to sources for better understanding the issue, the UI of such a system should not contain technical *terms* and *relations* from domains of building construction or traffic. It however should definitely contain terms from medicine. The logical reason is that a medic is usually not familiar with the technical dictionary of a building constructor or a traffic manager. Logically, the UI should also define *processes* from the specific domain of medicine by implementing sequences of events that can be executed on the UI. When performing a task, the user should follow the steps similar to the ones in the real world. For example, one cannot send money into another account without entering the account number and amount. In addition, the domain content should be mapped correctly.

Fig. 1 describes an example of a user's view of a system, which was developed for a different domain. A *motorcycle seller* uses a system made for a different domain of *Car selling*. While the motorcycle seller is trying to find the functionality for selling motorcycles in the system, the system provides only the functionality for selling cars. The user spends time searching for the right term, which reduces the system's usability.

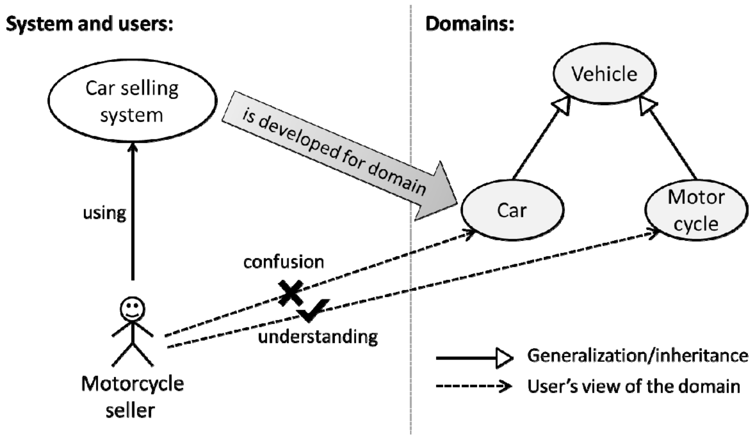


Fig. 1 The perception of a *Car selling system* by a user from the *Motorcycle domain*

While *Car* and *Motorcycle* domains are both subdomains of the *Vehicle* domain, the terms in both domains are not interchangeable. Therefore the *Car selling system* will never be a perfect choice for a *Motorcycle seller* and a *Motorcycle selling system* should be used instead.

Another example is when programmers oftentimes forget about the users and put functionalities and implementation details into the UI, which are very important for programmers, but not important for the users at all. For example logging, icons indicating the state of the system, database ids etc. Such functionalities are unknown to the users and they have no interest to see them in the application.

4.2 Adequate Level of Specificity

It is important to select an adequate level of specificity when creating an application's domain dictionary. Terms that are too generic usually reduce the usability of a UI. If a system is made for a wider range of customers, then it should not use generic terms. A specific term may be preferred, as it provides *more confidence to the users.*

A software company was asked to develop two systems for two different shops, one for a *motorcycle seller* and one for a *car seller*. To spare time and resources, the software company decided to develop one single general system for both of them: a system for selling *Vehicles*. Fig. 2 illustrates such a case. While both users try to find the functionality for selling cars and motorcycles in the system, the system provides only the functionality for selling vehicles. They spend more time searching for the right terms, which reduces the system's usability.

Although the *Vehicle* domain is a parent of *Car* and *Motorcycle* domains, usage of the general terms from the *Vehicle* domain in a system can be confusing for the users from a specific subdomain. Therefore the Vehicle selling system is not a perfect choice for both sellers. Instead, a specific *Car selling system* and *Motorcycle selling system* should be used.

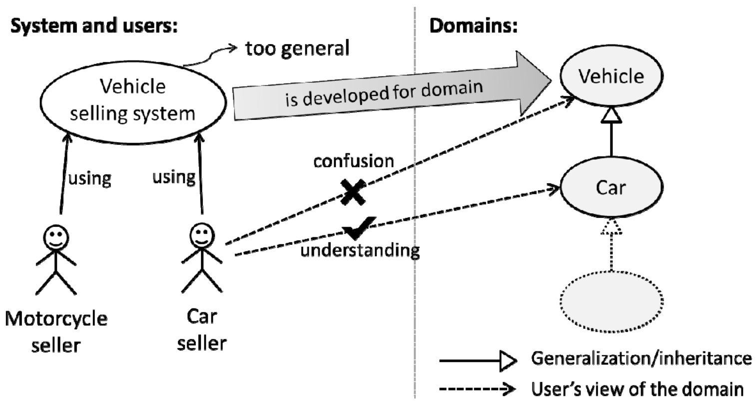


Fig. 2 The perception of a system from a general domain (*Vehicle*) by a user from a specific domain (*Car, Motorcycle*)

On the other hand, the terms in a system should not be too specific, if the system is used by a user who works in a more general domain. Fig. 3 is an example of a system for an *Electro mobile selling system* and its user, a *Car seller*. The user sells cars, but not only electric cars, but also regular cars and trucks. While he/she tries to find the functionality for selling *regular cars* in the system, the system only provides the functionality to sell *electro mobiles*.

While the *Electro mobile* domain is a subdomain of the *Car* domain, the terms in the *Electro mobile* domain are not sufficient for car sellers, because beside the terms from the *Electro mobile* domain, they use also terms from other subdomains of the *Car* domain, such as *Car* and *Truck*. Therefore, a *Car selling system* should be used in this case.

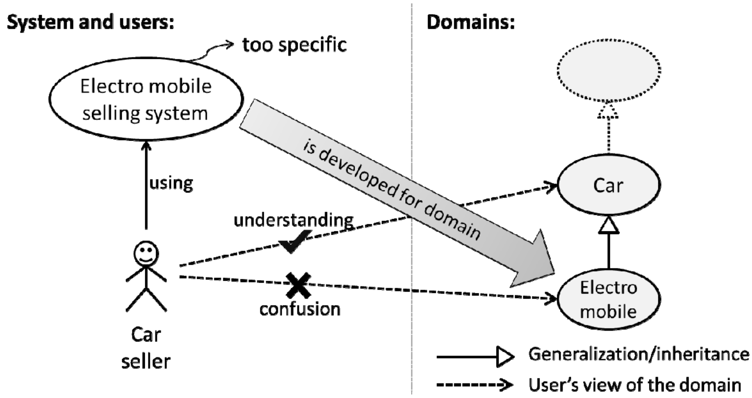


Fig. 3 Example of an *Electro mobile selling system*, which is too specific to be used in the domain of *Cars*

In the development phase of product lines, the diversity of systems used in different specific domains can also be supported by localization. Instead of using different world language in different localization files, different domain-specific languages can be used. For example in a system for selling vehicles, if it is used in the specific sub-domain of car selling, motorcycle selling or bus selling, the localization files could contain alternative terms for these sub-domains.

4.3 Consistency

Clear words and commands should serve as a standard through the whole system, especially if the system consists of several subsystems. For example it is better to use the same wording labels for buttons or menus that have the same functionality throughout the whole system. The notation of such terms is equally important (e.g. „Open“, „open“, „OPEN“). If there are two components of two equal labels doing two different tasks, it is probably not quite right because the user will never know the difference just from looking at their icon and it is laborious to learn the functionality of both by guessing and testing.

Although consistency was not the primary goal of localization, it can be used to support it.

4.4 Language Barriers

If the application is developed for Slovak market, then it should use Slovak language as the primary language. If the application is developed for users from several different countries, it should provide translations to different languages and a possibility to switch between them. The terms in *one single translation* should be *only of the same language*, i.e. the translation should be complete – there should be no foreign words. If there are foreign words in the application, a translation should be present nearby. In addition, if the UI contains any unknown abbreviations, an explanation of their meaning should be present.

Internationalization of UIs is supported by localization files (for example such as the ones provided by the Android platform), where all textual information is stored in a single file in the primary language, and each translation is provided in a separate file, containing a translation of every term from the primary one.

4.5 Errors

When reading any textual information, errors and grammatically or stylistically incorrect words or sentences can reduce readability, distract the readers' attention and extend the time needed to perform their work. Therefore the terminology a UI should be grammatically error-free and stylistics of sentences should be correct.

5 Related Work

An analysis was conducted in the area of domain usability, semantic UIs and ADUE and the most important works were selected which directly refer to domain analysis or at least slightly relate to the topic. The number of works referring to matching the application's content to the real world indicates the importance of domain usability.

Jacob Nielsen in 1994 already underlined the importance of textual content of UIs. He writes only very generally about it, but he stresses the importance of „the system's addressing the user's knowledge of the domain" [Nielsen 1994].

Tilly and Porkoláb [Tilly and Porkoláb 2010] deal with the problem of ambiguity of UI term apparatus design and term semantics in the UI. Their solution is to introduce the so called „semantic user interfaces" (SUIs). The core of a SUI is a general ontology which is a basis for creating all UIs in the specific domain. The UIs can be of a different appearance and arrangement but the domain dictionary must remain the same. Tilly and Porkoláb also stress the importance of application's copying the domain. They identified the problem of absence of standards, which would reduce the designers of intuitive UIs in choosing voluntary attributes (menu labels, buttons, tool-tips, menu topology) for displaying metaphors of UIs. That is why UIs become diverse even inside a specific application domain.

Billman et al. [Billman et al. 2011] refer to the domain content of the application as to the „domain structure". An experiment was performed where a new version of a NASA application was implemented with an emphasis on better copying the real world. The results showed that there is a big difference in the performance of users with respect to usability of the old application and the new, while the new application was better. Billman stresses the need of application's corresponding to the real world and respecting it in the early phases of application design.

Shneiderman [Shneiderman 1984] deals with the complexity of UIs and he stresses that the complexity of the textual content should not be too high because the application will be less usable.

The study of web usability for older adults [Becker 2004] by Becker deals, besides other attributes, with the translation and reading complexity of web pages. The author refers the ARI index or the Kincaid index to evaluate the reading complexity.

Badashian et al. [Badashian et al. 2008] defined the basic guidelines for creating usable UIs. Among other, they stressed the importance of application's term consistency and matching with the real world.

Understandability of web documents is defined in the area of web accessibility [W3C 2008]. Comparing to our definition however, it deals only with some of the attributes: world language of the UI, language barriers and errors. It focuses on web pages specifically, not on UIs in general. Besides understandability it defines also a few other attributes that do not fit into the context of the UI usability. Similar properties were defined by Badashian et al. but the domain usability is not explicitly specified.

The survey paper [Ivory and Hearst 2001] by Ivory and Hearst contains a wide list of automatic usability methods and tools. From over 100 tools, only Sherlock by Mahajan and Shneiderman [Mahajan and Shneiderman 1997] deals with domain content of applications. Sherlock is a family of consistency checking tools. According to Mahajan and Shneiderman, the user studies have shown a 10 to 25% speed-up of work with consistent UIs. Sherlock is a C application and it is able to evaluate applications created in Visual Basic 3.0. It primarily focuses on common action sequences, terms, units, layouts, colors, typography, etc. Sherlock is able to check concordance of terms used in the UI, misleading synonymous computer terms using terminology baskets (such as "Close", "Cancel", "End", "Exit", "Terminate", "Quit"). Among other attributes (such as gridedness, area balances, typefaces, background, foreground colors, margins etc.) the terms are evaluated to be consistent throughout the whole UI. Sherlock however does not evaluate, if different terms are describing the same functionality, or not.

Hilbert and Redmiles in [Hilbert and Redmiles 2000] deal with automated usability information extraction from event sequences, which are the natural outcome of normal operation of each UI. The event sequences should correspond with the real world as much as the domain dictionary.

Ontologies are used nowadays in an ever wider scale in the Semantic Web to denote the semantics of different UI concepts. Ontologies are dictionaries or domain models containing a list of domain terms, their properties and relationships between these terms. Kleshchev and Artemieva [Kleshchev 2011; Artemieva 2011] from the Institute for Automation & Control **Processes** at the Russian Academy of Sciences argue about ontologies leading to a considerable progress in the area of software development. On the other hand, Gribova in [Gribova 2007] presented a method of estimating usability of a UI based on its domain model. These works imply that a domain model (whether in form of a domain dictionary or ontology) is an important part of software development and usability can be estimated based on the domain model. Domain model is an inseparable part of systems and UIs.

6 Conclusion

In this paper we identified the main problems related to the domain content of UIs. We identified understandability of UIs and domain usability as a subset of general usability. We provided examples to illustrate the five attributes of understandability. The definitions could serve as a guide for creating applications better matching the real world and this way they will be as close as possible to domain users. We hope for stimulating research in the area of domain usability and motivating researchers and developers in creating domain usability evaluation methods and tools. We hope for stimulating the research in the area of ADUE. Our future research areas are ADUE methods and heuristics, automatized domain analysis of existing UIs as well as reverse engineering and program comprehension.

Acknowledgment. This work was supported by VEGA Grant No. 1/0305/11 Co-evolution of the Artifacts Written in Domain-specific Languages Driven by Language Evolution. Thanks to Jakub and Sergej for performing a review.

References

[Artemieva 2011] Artemieva, I.L.: Ontology development for domains with complicated structures. In: Proc 1st Intern. Conf. on Knowledge Processing And Data Analysis, pp. 184–202 (2011)

[Bačíková and Porubän 2013] Bačíková, M., Porubän, J.: Ergonomic vs. domain usability of user interfaces. In: HSI 2013: 6th Intern. Conf. on Human System Interaction, pp. 1–8 (2013)

[Badashian et al. 2008] Badashian, A.S., Mahdavi, M., et al.: Fundamental usability guidelines for user interface design. In: Proc. Intern. Conf. on Computational Sciences and Its Applications, Washington, DC, USA, pp. 106–113 (2008)

[Becker 2004] Becker, S.A.: A study of web usability for older adults seeking online health resources. ACM Trans. Comput.-Hum. Interact. 11(4), 387–406 (2004)

[Billman et al. 2011] Billman, D., Arsintescucu, L., et al.: Benefits of matching domain structure for planning software: the right stuff. In: Proc. Annual Conference on Human Factors In Computing Systems, pp. 2521–2530 (2011)

[Gribova 2007] Gribova, V.: A method of estimating usability of a user interface based on its model. Intern. J. Information Theories & Applications 14(1), 43–47 (2007)

[Hilbert and Redmiles 2000] Hilbert, D.M., Redmiles, D.F.: Extracting usability information from user interface events. ACM Comput. Surv. 32(4), 384–421 (2000)

[Ivory and Hearst 2001] Ivory, M.Y., Hearst, M.A.: The state of the art in automating usability evaluation of user interfaces. ACM Comput. Surv. 33(4), 470–516 (2001)

[Kleshchev 2011] Kleshchev, A.S.: How can ontologies contribute to software development? In: Wolff, K.E., Palchunov, D.E., Zagoruiko, N.G., Andelfinger, U. (eds.) KONT 2007 and KPP 2007. LNCS, vol. 6581, pp. 121–135. Springer, Heidelberg (2011)

[Mahajan and Shneiderman 1997] Mahajan, R., Shneiderman, B.: Visual and textual consistency checking tools for graphical user in-terfaces. IEEE Trans. Soft. Eng. 23(11), 722–735 (1997)

[Nielsen 1994] Nielsen, J.: Usability Engineering. Academic Press, Boston (1994)

[Shneiderman 1984] Shneiderman, B.: Response time and display rate in human performance with computers. ACM Comput. Surv. 16(3), 265–285 (1984)

[Shneiderman and Plaisant 2004] Shneiderman, B., Plaisant, C.: Designing the User Interface: Strategies for Effective Human-Computer Interaction, 4th edn. Pearson Addison Wesley (2004)

[Tilly and Porkoláb 2010] Tilly, K., Porkoláb, Z.: Automatic classification of semantic user interface services. Ontology-Driven Software Engineering 6, 1–6 (2010)

Incorporating Usability into Model Transformation

L.B. Ammar[1], A. Trabelsi[2], and A. Mahfoudhi[1,2]

[1] University of Sfax, ENIS, CES Laboratory Soukra km 3,5, B.P:1173-3000, Sfax, Tunisia
[2] University of Valenciennes, LAMIH, UMR CNRS 8530,
BP:311-59304, Valenciennes, France
{lassaad.ben-ammar,adel.mahfoudhi}@ceslab.org,
abdelwaheb@gmail.com

Abstract. Model transformation is a central concept in model-driven development approach, having the capability to generate the system code from a conceptual model that specifies the system abstractly. We notice the lack of model transformation technologies that deal with usability from the first step of the transformation process. Usability features are usually implemented later once the system code is generated. In the present work we propose a strategy to deal with usability in a model transformation technology. We show that the selection of alternative transformation based on usability criteria ensure that the target model will maximize these criteria.

1 Introduction

Nowadays, the Model Driven Engineering (MDE) paradigm is widely applied in the Human-Computer Interaction (HCI) domain. In this context, the Cameleon Reference Framework [Calvary et al., 2001] proved quite appropriate to generate *plastic* User Interfaces (UI). These UIs are able to adapt to the context of use (user, platform, and environment) while preserving usability.

The development process proposed in the Cameleon framework is structured into four levels of abstraction. The first level named *Task and Concepts* (T&C) describes tasks and domain-oriented concepts. The *Abstract User Interface* (AUI) level expresses the UI in terms of Abstract Interaction Objects (AIOs). The *Concrete User Interface* (CUI) level concretizes the AIOs into Concrete Interaction Objects (CIOs). Finally, the *Final User Interface* (FUI) represents the code of the UI in any programming or mark-up language.

Model transformations are a central concept in the Cameleon Framework. They provide a mechanism for automatic creation of target models, based on information contained in existing source models. They are exploited to ensure the adaptation of UIs to their context of use while disregarding the usability aspects. However, usability is crucial for the success of an interactive system. This can explain the ever-increasing of research works dealing with usability in HCI literature.

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_3,

Several methods, techniques and tools have been proposed in order to evaluate the usability of UIs. They involve activities that require a huge amount of resources (expert, end users, usability lab, video recorder …). In addition, the evaluation is usually conducted once the system is implemented. This greatly reduces the ability to go back and makes major changes to the design.

Recent studies have begun to explore the problem of integrating the usability aspects in a UI development process following the MDE principles. Some proposals such as [Abrahao and Insfran 2006; Panach et al. 2011] have demonstrated that evaluating usability early at the development process is an appealing way to ensure the usability of the generated user interface. The evaluation takes as input the conceptual model. It provides a usability report which contains detect problems. Problems are analyzed in order to suggest changes to the design. By means of model transformation and explicit traceability between models, the performed changes directly reflected into the intermediate artifact avoiding usability problems in any future user interface obtained as part of the transformation process.

Other proposals evaluate the usability of a user interface generated with an MDE approach. We quote for example proposals presented in [Abrahao et al., 2008; Fernandez et al. 2009; Aquino et al. 2010]. The usability evaluation is based on the system code and on the generated interface. This makes their application during the model transformations process difficult.

If we focus on the model-driven development paradigm, we notice the lack of proposals that deal with usability in a model transformation process. However, several research works have focused their efforts to drive the selection of alternative transformation by quality attributes [Matinlassi 2005]. They have shown the importance of quality-driven model transformation in the obtaining of a target model which maximizes the desired quality attributes.

An initiative is presented in [Panach et al. 2008] in order to drive the model transformation by usability features. The interest is focused on usability features which have an impact on the system architecture (namely *Functional Usability Features FUF*) while neglecting UI design. However, unusable UIs are probably the main reason which leads to the fail in actual use of interactive system [Seffah et al. 2006].

It becomes clear that incorporating usability features into a model transformation technology is still an immature area and many more research works are needed. In order to covers this need; the present chapter proposes a strategy to support the *usability-driven model transformation.* The objective is to improve the usability of the target model by selecting alternative transformation that maximizes the desired usability criteria.

2 Proposed Method to Ensure Usability

The proposed method focus on the reification process defined in the Cameleon Framework. Particularly, in the model transformation that take the AUI as input to generate the CUI. This transformation associates with each abstract construct coming from the AUI a concrete construct in the CUI model. There may be several concrete constructs for the same abstract one. In this case we talk about alternative transformation. These alternative transformations allow the same functionality but may differ from the usability point of view (see Fig.1).

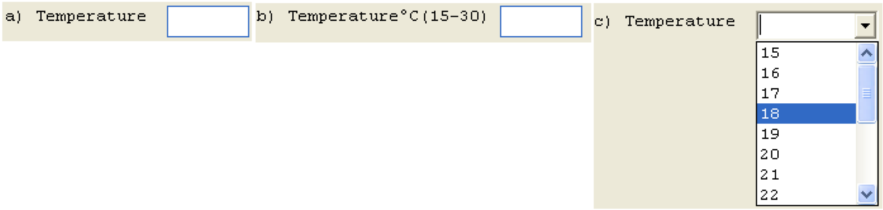


Fig. 1 Three functionally-equivalent user interface that differ from the usability point of view

All the user interfaces of Fig.1 are equivalent from the functional perspective. However, from the non-functional perspective they do not satisfy the same usability criteria. Solution b) allows better user guidance than solution a). It displays the measurement unit of the temperature and the range of accepted values. Hence, the *Prompting* property is well satisfied in solution b). In solution c), user is prevent to make error (i.e. typos) while entering the temperature value. Thus, the *Error Prevention* usability property is well fulfilled in solution c). Hence, selecting the alternative transformation based on the usability criterion can be an appealing way to ensure that such criteria will be meeted in the target model. Starting from this report, we adapt the reification process of the Cameleon framework by adding a set of usability criteria on which the selection of the adequate alternative transformations will be done (see Fig.2).

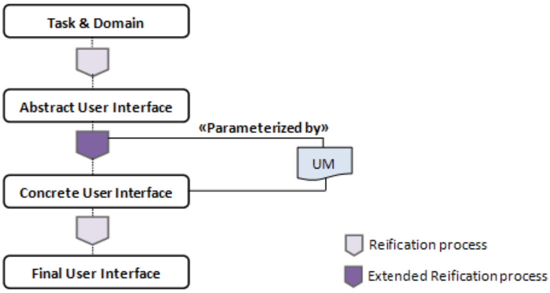


Fig. 2 The extended Cameleon Framework

The main idea of our proposal is to insert a set of practical usability criteria as a parameter to the model transformations engine. To do this, we reformulate the parameterized transformation principles initiated by [Vale and Hammouda 2008].

The applicability of our method is shown through the Cameleon-compliant method presented in [Bouchelligua et al. 2010].

2.1 Usability Criteria Specification

Usability is a difficult concept to quantify. It has several dimensions and several factors seem to impact upon it. Previous studies [Abrahao et al. 2006; Panach et al. 2011; Seffah et al. 2006] have identified a number of factors that contribute to usability of the user interface.

In this chapter, the problem we intend to address is to isolate the most important of these criteria and work out a means of characterizing the impact of each of them to the alternative transformation selection.

To better explain our method, we illustrate the impact of 7 usability criteria (Prompting, Error Prevention, Explicit User Action, User Operation Undoability, User Operation Cancellability, Information Density, and Brevity) on the selection of the adequate alternative transformations. It should be noted that our method can be extended to covers many more usability criteria.

Usability criteria such as *Prompting* and *Error Prevention* are crucial to better guide the user throughout the application.

Interactive system must allow a certain level of control to the human agent. Hence, usability criterion such as *Explicit User Action, User Operation Undoability* and *User Operation Cancellability* are crucial and must be considered.

The screen size of the interactive device may affect some usability criteria such as the *Information Density* and the *Brevity*. Thus, it is essential to investigate the impact of such usability criteria.

2.2 Usability and the Design Decisions Control

We used Kermeta (Kernel meta-modeling) as a transformation language to implement our method. It allows the description of both structure and behavior of models. In what follow, we detail the impact of the selected usability criteria on the selection of the adequate alternative transformations.

Prompting

The Prompting usability property refers to the means available to advise, orient, inform, instruct, and guide the users throughout the application. A simple example of the prompting property is illustrated by the addition of specific information to inform user about the required format while specifying data. The extract of following kermeta code ensures the prompting property.

```
operation UIFieldTreatment (AUIModel:AbstractUserInterface, uic
: CollapsedUIUnit,uiw : UIWindow) is do
var lnk : Link init getAllLinks(AUImodel).detect{c|
c.uicomponent.name == uic.name}
var componentnature : Nature init uic.nature
if (componentnature == Nature.Specify) then
    createFieldIn(uiw,uic,lnk)
    if (PromptingSupport(paramModel) then
        createStaticField(uiw,uic,lnk.name,
lnk.uicomponentannot.dataformat)
    else
        createStaticField(uiw,uic,lnk.name,"")
    end
end
```

Having restored the annotation attached to the abstract component through the link *lnk*, all information about the component (name, nature, required format, etc.) is available. If the input abstract component has the nature *Specify*, we associate with this component an edit field and a label in the concrete UI model. If the prompting property is required, we add the specific information (the required data format in the example) to the label.

Error Prevention

The Error Prevention property refers to the available means to prevent data entry errors. We propose as example to create a dropdown list (or radio button) instead of an edit field when the data to be inserted has a set of enumerated values.

```
if (ErrorPreventionSupport(paramModel) then
     if (lnk.uicomponentannot.conceptNB > 5) then
     createDropDownList(uiw,uic,lnk)
     else
         from var i : Integer init 0 until i == conceptNB loop
         createRadioButton(uiw,uic,lnk)
         end
     end
end
```

The number of the manipulated concept is the main factor that affects the choice of the target element. If the number of manipulated concepts is greater than a threshold (5 in the example) the input element will be realized by a dropdown list. Otherwise it will be realized by a set of radio button.

Information Density

The Information Density refers to the degree in which information will be display to the user in each interface. User interface should not be too dense. Information density can be measured with respect to the total number of interface elements which should not exceed a threshold. In the following kermeta code, we used an example of 20 elements as a threshold [Panach et al. 2011].

```
var Totalelement : Integer init Integer.new
 Totalelement := NBelemnt(AUImodel)
 if (SupportInformationDensity(paramModel)) then
     if (Totalelement>20) then
      uig.uiunitsuit.each{uius|
     var uiw: UIWindow init UIWindow.new
     uiw.name := uig.name
     result.uiwindow.add(uiw)
     uius.collapseduiu.each{cuiu|
      createUIField(inputModel,cuiu,uiw)}
     }
    end
end
```

Having a total number that exceed the threshold, we propose to associate a window to each unit suite which is usually realized by a panel.

Explicit User Action

The Explicit User Action refers to the relationship between the computer processing and the actions of the users.

```
operation createCheckBoxP(uip :UIPanel,uic : CollapsedUIUnit,lnk
: Link) is do
    var ddl : UICheckBox init UICheckBox.new
    ddl.name := lnk.uicomponentannot.data
    ddl.items:= lnk.uicomponentannot.enumValues
    uip.uifieldP.add(ddl)
    if (SupportExplicitUserAction(paramModel)) then
        createButtonP(uip,"Ok")
    end
end
```

The computer must process only actions requested by the users and only when requested to do so. For example, each data entry (edit field, radio button, check box, dropdown list) should be ended by an explicit validation action by the user. The above kermeta code shows the implementation of such property to the check box element.

User Operation Undoability UOU

The UOC refers to the proportion of actions that can be undone without harmful effect to the normal operation. As example, we propose the same strategy as *User Explicit Action* by adding an *Undo* button to each action.

User Operation Cancellability UOC

The UOC refers to the proportion of actions that can be canceling without harmful effect to the normal operation. We propose the same strategy as *User Explicit Action* and *UOU* by adding a *Cancel* button to each action.

Brevity

The Brevity concerns workload with respect to the number of step (keystrokes) necessary to accomplish a task. The reduction of the effort needed to perform a task can be materialized by the elimination of the navigation between windows with respect to the relationship. If the relationship is “simultaneous” both group (source and target) will be concretized by a panel in the same window. If the relationship is “sequential” the target group will be concretized by a panel in the window associated to the source group.

```
if (SupportBrevity(paramModel)) then
    var rsp : UIRelationShip init UIRelationShip.new
    rsp:=       AUImodel.uispatiotemporalrelationship.detect{rs|
rs.source == uig.name}
    if (rsp.type.equals("Simultaneous")) then
        var uiw : UIWindow init UIWindow.new
        uiw.name := "General Window"
        result.uiwindow.add(uiw)
        var srcpanel : UIPanel init UIPanel.new
        srcpanel.name := uig.name
        uiw.uipanel.add(srcpanel)
        var trgpanel : UIPanel init UIPanel.new
        trgpanel.name := rsp.target
        uiw.uipanel.add(trgpanel)
    else
        var uiw : UIWindow init UIWindow.new
        uiw.name := uig.name
        result.uiwindow.add(uiw)
        var trgpanel : UIPanel init UIPanel.new
        trgpanel.name := rsp.target
        uiw.uipanel.add(trgpanel)
    end
end
```

2.3 Discussion

The examples already mentioned are intended to give a clear outlook to the impact of some usability criteria on the selection of the adequate design decisions (alternative transformations). The objective is to ensure that the concrete UI model includes concrete component fulfilling the desired usability properties. In what follow, we illustrate the feasibility of our method with a case study.

3 Case Study

This section investigates a case study in order to illustrate the feasibility of our proposal. The purpose is to allow us to learn more about the potentialities and limitations of our proposal. The research question addressed by this case study is: *Does the proposal ensure the usability of the generated user interface artifact*?

The case study is a "Requesting a credit to buy a car". The scenario is adapted from [Guerrero 2010]. To get information about the credit to buy a car, a bank client does not have to go to the bank branch since the bank portals offer an interactive system which allows resolving such problem. The bank client can perform several tasks using this system: get information about buying tips, simulate a credit to buy a car, request the credit, receive request in line, and communicate with the credit department, etc. For reason of simplicity, we are interested in the "Request the Credit" task.

Fig.3 shows an XML presentation of the abstract user interface (left part) generated for such task. The right part shows a sketch of the abstract user interface presented by an editor developed with the Graphical Modeling framework (GMF) of eclipse.

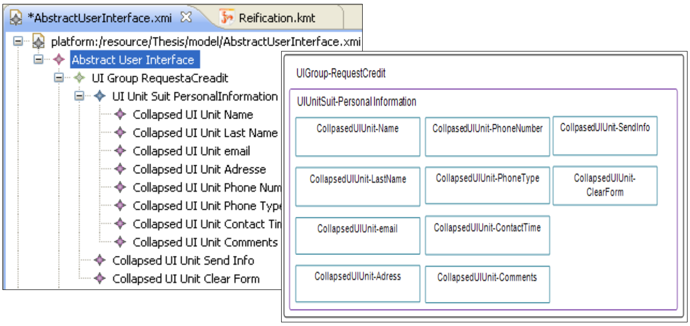


Fig. 3 Abstract User Interface

Fig.4 shows the deduced concrete user interface without taken into account the usability properties for an interactive device with a large screen size (e.g, PC). Although the generated concrete user interface filled their objectives, it does not satisfy some usability properties. Usability properties such as *Prompting* and *Error Prevention* are not supported.

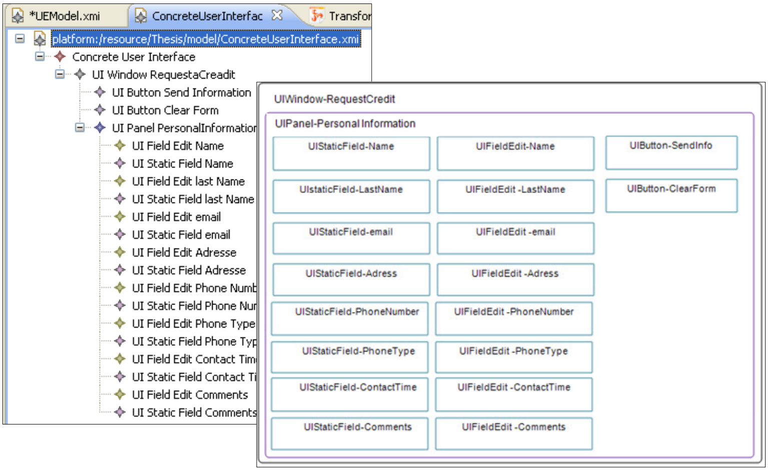


Fig. 4 Concrete User Interface (large screen size)

After inserting the desired usability properties to the transformation process, we can see their impact in the generated UI shown in Fig.5.

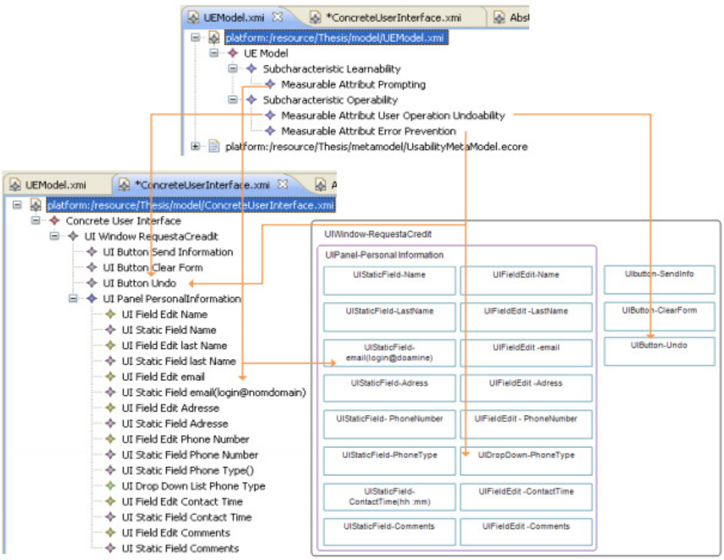


Fig. 5 Usable Concrete User Interface for large screen size

The migration to the «iPAQHx2490PocketPC» platforms raises a new redistribution of the user interface elements. The small screen size (240x320) is not enough to display all information. The information density is the main usability property that is influenced in this case. The number of the concrete component to be grouped is limited to the maximum number of concepts that can be manipulated (5 in the case of «iPAQHx2490PocketPC»). Fig.6 shows the deduced concrete user interface for the underlying platform.

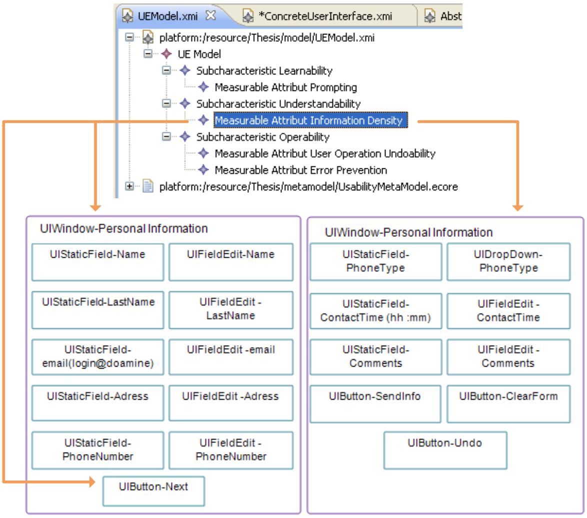


Fig. 6 Usable Concrete User Interface for small screen size

It is highlighted that taking into consideration the usability properties during the transformation process guarantees that these properties will be filled in the generated artifact. The transformation rules will be founded on the basis of the usability properties conveyed throughout the usability model. Hence, the choice of the adequate interface component is constrained by the usability property which is able to meet. Building on the implemented properties, we argue that our proposal contributes to the improvement of the generated user interface.

We note that the take into account of the *Information Density* property has influenced negatively the *Brevity* property. This raises a new issue about the contradictory effect of the usability properties. Therefore, it is recommended to investigate the relationship between usability properties and their contradictory effect.

4 Conclusions and Future Work

In this paper, an extension of the Cameleon reification process is presented. The main motivation of the extension is to keep in mind usability issues during the transformation process. The objective is to ensure that the generated user interface fulfill the desired usability properties. To reach this objective, we build our proposal on the parameterized transformation technique. In such transformation, a parameter model is required to communicate the usability requirement. The specification of the transformation rules is made up following the desired usability properties. The selection of the adequate alternative transformations depends on the usability criteria we want to maximize in the generated artifact.

The usability driven model transformation concept initiated in the present paper is the main advantage of our proposal if comparing with existing one. The model transformations definition is accompanied by a proper level of detail. The execution of the model transformations is illustrated through a practical case study. The case study presented is useful since it highlight the benefits and limitations of our proposal.

Several research studies can be considered as a continuation of this work. As an example, further research works are intended to perform an automatic evaluation process of the intermediate artifacts in order to detect potential usability problems. To do that, we have to propose a consolidated usability model gathering all properties which can influence the usability of a user interface. Usability attributes are intended to be evaluated at the intermediate artifact. Thus, only quantitative measures are needed. The interrelation between usability properties and the contradictory effects of properties can also be targets of attack for future work.

References

[Abrahao and Insfran 2006] Abrahao, S., Insfran, E.: Early usability evaluation in model driven architecture environments. In: Proc. 6th International Conference on Quality Software, pp. 287–294. IEEE Computer Society, Washington DC (2006)

[Abrahao et al. 2008] Abrahão, S., Iborra, E., Vanderdonckt, J.: Usability evaluation of user interfaces generated with a model-driven architecture tool. In: Maturing Usability, pp. 3–32 (2008)

[Aquino et al. 2010] Aquino, N., Vanderdonckt, J., Condori-Fernández, N., Dieste, Ó., Pastor, Ó.: Usability evaluation of multi-device/platform user interfaces generated by model-driven engineering. In: Proc. ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, New York, USA (2010)

[Bouchelligua et al. 2010] Bouchelligua, W., Mahfoudhi, A., Mezhoudi, N., Daassi, O., Abed, M.: User interfaces modelling of workflow information systems. In: Barjis, J. (ed.) EOMAS 2010. LNBIP, vol. 63, pp. 143–163. Springer, Heidelberg (2010)

[Calvary et al. 2001] Calvary, G., Thevenin, D.: A unifying reference framework for the development of plastic user interfaces. In: Nigay, L., Little, M.R. (eds.) EHCI 2001. LNCS, vol. 2254, pp. 173–192. Springer, Heidelberg (2001)

[Coutaz 2010] Coutaz, J.: User interface plasticity: model driven engineering to the limit! In: Proc. 2nd ACM SIGCHI Symposium on Engineering Interactive Computing Systems, New York, USA, pp. 1–8 (2010)

[Fernandez et al. 2009] Fernandez, A., Insfran, E., Abrahão, S.: Integrating a usability model into model-driven web development processes. In: Vossen, G., Long, D.D.E., Yu, J.X. (eds.) WISE 2009. LNCS, vol. 5802, pp. 497–510. Springer, Heidelberg (2009)

[Guerrero-Garcia 2010] Guerrero-Garcia, J.: A methodology for developing user interfaces to workflow information systems. Thèse de Doctorat, Université de catholique de Louvain (2010)

[Matinlassi 2005] Matinlassi, M.: Quality-driven software architecture model transformation. In: Proc. 5th Working IEEE/IFIP Conference on Software Architecture, pp. 199–200. IEEE Computer Society, Washington DC (2005)

[Panach et al. 2008] Panach, J.I., España, S., Moreno, A.M., Pastor, Ó.: Dealing with usability in model transformation technologies. In: Li, Q., Spaccapietra, S., Yu, E., Olivé, A. (eds.) ER 2008. LNCS, vol. 5231, pp. 498–511. Springer, Heidelberg (2008)

[Panach et al. 2011] Panach, J.I., Condori-Fernández, N., Vos, T., Aquino, N., Valverde, F.: Early usability measurement in model-driven development: Definition and empirical evaluation. Int. J. Softw. Eng. Know. 21(3), 339–365 (2011)

[Seffah et al. 2006] Seffah, A., Donyaee, M., Kline, R.B., Padda, H.K.: Usability measurement and metrics: A consoli-dated model. Software Quality Control 14(2), 159–178 (2006)

[Schmidt 2006] Schmidt, D.C.: Guest editor's introduction: Model-driven engineering. Computer 39(2), 25–31 (2006)

[Vale and Hammouda 2008] Vale, S., Hammoudi, S.: Context-aware model driven development by parameterized transformation. In: Proc. 1st International Workshop on Model Driven Interoperability for Sustainable Information Systems held within CAiSE 2008 Conference, pp. 121–133. Springer, Heidelberg (2008)

Enhanced Event Modeling for Human-System Interactions Using IOPT Petri Nets

R. Campos-Rebelo[1,2], A. Costa[1,2], and L. Gomes[1,2]

[1] Universidade Nova de Lisboa, Faculdade de Ciências e Tecnologia, Portugal
[2] UNINOVA, Centro de Tecnologia e Sistemas, Portugal
{rcr,akc,lugo}@uninova.pt

Abstract. This work presents some new ways to analyze a signal that allows the modeling of human-system interactions. In addition to the possibility of creating conditions associated with the current value of the signal, the work focuses mainly on the definition of events and on their use within Input-Output Place-Transition (IOPT) nets models. Currently, non-autonomous events can be defined in two different ways, namely as a threshold detection of a signal, or variation of the first derivative of the signal. Here, this concept is generalized, allowing the analysis of events in higher derivatives of the signal. On the other hand, a new set of types of events are defined for each derivative of the signal, improving existing UP and DOWN events. Finally, an example is used to illustrate discussed concepts.

1 Introduction

Considering the sustained increasing on the complexity of systems, normally associated with new technology developments and needs to make systems more pleasant by creating more appealing and user friendly interfaces, several approaches have been proposed, based on decomposition of the system development flow. An example of these approaches is the Model Based Development approach, which can use specific languages for system specification.

Graphical languages are one type of these languages. State charts [Harel 1987], state-machines variants [Borger and Stark 2003] or Petri nets are some examples of graphical languages, which allow a view of the behavioral model and where implementation code generation can be fully automated, providing a visual representation of the system behavior easily understandable (and analyzed) by a human user.

Since its presentation by Carl Adam Petri in 1962, Petri Nets (PN for short) have benefited from many changes and developments, which have led to the growth of their use in several areas. Engineering is one of those areas.

The introduction of non-autonomous characteristics to Petri nets models (allowing explicit modeling of inputs and outputs, for instance) contributed significantly to the increase of their use. This extension allows the PN model to be capable of modeling, beyond the behavior of the system, interaction with the surrounding environment, making them suitable for modeling controllers and human-system interactions.

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_4, © Springer International Publishing Switzerland 2014

In [David and Alla 1992; David and Alla 2010] some proposals on how to represent the interaction with specific environments are presented. In the literature, several examples of non-autonomous Petri net classes are easily found. For example, the Synchronized Petri nets [Moalla et al. 1978] associates an event occurrence to the firing of a transition. Another class is the Signal Interpreted Petri net, proposed by Frey [Frey and Minas 2000; Frey 2002] to be used as a PLC (programmable logic controller) programming language, and where the concept of logical input and output signal was also used. Another non-autonomous Petri net class targeted for controller modeling is the Input-Output Place-Transition (IOPT) Petri net [Gomes et al. 2007]. IOPT nets are used in this paper to integrate the proposed events modeling and are briefly introduced in section 2.

This work redefines elementary events presented in [Campos-Rebelo et al. 2013a], allowing a better organized, hierarchical and scalable definition. This goal is achieved through the use of the concept of signal differences (equivalent to derivatives in the continuous domain) considering step execution. For each step evolution, a new set of possible events associated to the possible evolutions is presented. Despite the proposal of these new types of events have been developed with human-system interaction modeling in view, their application is not limited to such systems. These new types of events are also of interest in the modeling of other types of controllers, such as embedded controllers, as well as cyber-physical systems. Similarly, the new event type's usage is illustrated with IOPT nets models, but they may be included in other event-based modeling formalism.

The notation used during the paper considers the following representations of signal X: being x_n the value of signal X at execution step n, x_{n-1} the value of signal X at previous execution step, and x_{n-p} the value of signal X occurring p execution steps before. Additionally $\Delta x_n = | x_n - x_{n-1}|$, $\Delta x_{n-1} = | x_{n-1} - x_{n-2}|$ and $\Delta x_{n-p} = | x_n - x_{n-p}|$. This notation is extended to differences of higher order, which can be determined recursively; for example $\Delta^2 x_n = | \Delta x_n - \Delta x_{n-1}| = || x_n - x_{n-1}|-| x_{n-1} - x_{n-2}||$.

In section VI, an application example of a music player controller is presented for better illustrate the use of the various types of events presented.

2 IOPT Nets

An IOPT Petri net, as proposed in [Gomes et al. 2007], is a non-autonomous, low-level Petri net that extends Place Transition Petri net class [Desel and Reisig 1998] with input and output characteristics, such as input/output events and signals to model the environment, among others. These additions allow modeling of interaction with the environment. The IOPT nets support specification of:

- External input signals and external input events in order to constrain transition firing, as well as generation of external output events as a result of transition firing (Mealy style output),
- External output signals associated with marking of places (Moore style output).

As common within discrete event system modeling, the evolution of the system is possible on specific instants, called “tick”. The period between two "ticks" is called step (which is a concept of paramount importance along this paper). IOPT nets have maximal step semantics. This means that in each step, all transitions that are enabled and are ready to fire, will fire. In this sense, the synchronized Petri nets paradigm [David and Alla 1992] is adopted, where a transition is enabled to fire whenever the marking of its input places provide adequate marking, and is ready to fire whenever associated conditions on input signals and events are evaluated as true.

The IOPT nets may also support structuring mechanisms [Gomes and Barros 2005] supported by composition and decomposition techniques [Barros and Gomes 2004; Costa and Gomes 2009].

Currently, IOPT nets accommodate two types of events: autonomous events (event generated by the environment), and non-autonomous events (automatically generated events after processing of external signals). Only the later ones are considered for this paper. A non-autonomous input event, or simply input event, is associated to an input signal and to a threshold level that the signal must cross or reach to generate the event. It is the starting concept from where the proposals of this chapter are built and has been formally presented under the classification of “threshold event”.

3 Signal Analyses

In this work, a signal is considered to have a value at the beginning of each step analysis and an evolution along time, and can be analyzed in two different ways:

- Analysis of the signal value at a given moment in order to verify that this value meets a certain condition
- Definition of events associated with a specific behavioral evolution of the signal.

The main difference between these two types of analysis is related with the duration of the analysis (time window). A condition examines the value of the signal on a given instant of time, while when defining an event, the signal behavior is analyzed in different instants.

Additionally, a signal analysis can also be made taking into account difference of values of the signal between different instants. According with the presented proposal, the value of the signal, its first difference, its second difference or higher ones can be considered for modeling. Given this analysis, one can analyze not only a point or a variation between two points, but also behaviors taking into account three or more time steps of analysis of the signal.

Table 1 presents these two types of signal analysis: defining an event on a signal or on a difference, or associate a condition to the value of the signal or of a difference.

Table 1 Types of Signal analysis

	Events	Conditions
X	$X_{n-1} \leq K \land X_n > K$	$X_n = K$
ΔX	$\Delta X_{n-1} \leq K \land \Delta X_n > K$	$\Delta X_n = K$
$\Delta^2 X$	$\Delta^2 X_{n-1} \leq K \land \Delta^2 X_n > K$	$\Delta^2 X_n = K$
$\Delta^3 X$	$\Delta^3 X_{n-1} \leq K \land \Delta^3 X_n > K$	$\Delta^3 X_n = K$
$\Delta^n X$	•••	•••

Given that $\Delta X_n = X_{n-1} - X_n$ and $\Delta X_{n-1} = X_{n-2} - X_{n-1}$, an event in ΔX analyzes three signal values ((X_{n-2}, X_{n-1} e X_n). Likewise analyzing the event condition $\Delta^2 X$ it has that $\Delta^2 X_n = \Delta X_{n-1} - \Delta X_n = (X_{n-2} - X_{n-1}) - (X_{n-1} - X_n)$ and $\Delta^2 X_{n-1} = \Delta X_{n-2} - \Delta X_{n-1} = (X_{n-3} - X_{n-2}) - (X_{n-2} - X_{n-1})$. Thus $\Delta^2 X$ considers four values of the signal (X_{n-3}, X_{n-2}, X_{n-1} e X_n).

In this work, conditions will not be addressed, being the emphasis given to the analysis of signals events.

4 Events

In this section, events will be presented in detail. Twelve different types of events are identified and defined. Only two types of signal analysis are presented, namely those dependent on the signal X (called delta zero on X) and those dependent on the first difference of consecutive values of the signal X (called delta one on X). However, these types of events can be applied to higher order of differences (delta N on X).

Table 2 Event types definition

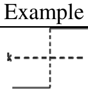
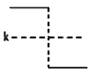
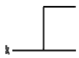
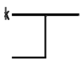
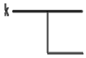
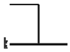
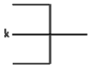
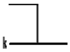
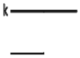
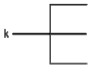
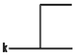
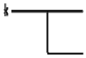
	X		ΔX
Type	Definition	Example	
<>	$Evx^{<>}(k) = (x_n > k)$ $(x_{n-1} < k)$		$Ev\Delta x^{<>}(k) = (\Delta x_n > k)$ $(\Delta x_{n-1} < k)$
><	$Evx^{><}(k) = (x_n < k)$ $(x_{n-1} > k)$		$Ev\Delta x^{><}(k) = (\Delta x_n < k)$ $(\Delta x_{n-1} > k)$
>	$Evx^{<}(k) = (x_n > k)$ $(x_{n-1} \leq k)$		$Ev\Delta x^{<}(k) = (\Delta x_n > k)$ $(\Delta x_{n-1} \leq k)$
≥	$Evx^{\geq}(k) = (x_n \geq k)$ $(x_{n-1} < k)$		$Ev\Delta x^{\geq}(k) = (\Delta x_n \geq k)$ $(\Delta x_{n-1} < k)$

Table 2 *(continued)*

$<$	$Evx^{<}(k) = (x_n < k)$ $(x_{n-1} \geq k)$		$Ev\Delta x^{<}(k) = (\Delta x_n < k)$ $(\Delta x_{n-1} \geq k)$
$\leq$	$Evx^{\leq}(k) = (x_n \leq k)$ $(x_{n-1} > k)$		$Ev\Delta x^{\leq}(k) = (\Delta x_n \leq k)$ $(\Delta x_{n-1} > k)$
$=$	$Evx^{=}(k) = (x_n = k)$ $(x_{n-1} \neq k)$		$Ev\Delta x^{=}(k) = (\Delta x_n = k)$ $(\Delta x_{n-1} \neq k)$
$=+$	$Evx^{=+}(k) = (x_n = k)$ $(x_{n-1} > k)$		$Ev\Delta x^{=+}(k) = (\Delta x_n = k)$ $(\Delta x_{n-1} > k)$
$=^{-}$	$Evx^{=-}(k) = (x_n = k)$ $(x_{n-1} < k)$		$Ev\Delta x^{=-}(k) = \Delta x_n = k)$ $(\Delta x_{n-1} < k)$
$\neq$	$Evx^{\neq}(k) = (x_n \neq k)$ $(x_{n-1} = k)$		$Ev\Delta x^{\neq}(k) = (\Delta x_n \neq k)$ $(\Delta x_{n-1} = k)$
$\neq +$	$Evx^{\neq +}(k) = (x_n > k)$ $(x_{n-1} = k)$		$Ev\Delta x^{\neq +}(k) = (\Delta x_n > k)$ $(\Delta x_{n-1} = k)$
$\neq^{-}$	$Evx^{\neq +}(k) = (x_n < k)$ $(x_{n-1} = k)$		$Ev\Delta x^{\neq +}(k) = (\Delta x_n < k)$ $(\Delta x_{n-1} = k)$

4.1 Events on Delta Zero (X)

The definition of events on Delta Zero is based on the events definition of IOPT in [Gomes et al. 2007], where they were presented as threshold events, further redefined in [Campos-Rebelo et al. 2013a]. This type of events is defined as the occurrence of crossing a specific threshold on the value of the signal. In [Gomes et al. 2007], IOPT nets have defined two types of events, named "UP" and "DOWN", and associated with rising and falling evolution of the signal. A threshold level K is defined for both types of events, and an event is generated based on analysis of the signal value in two consecutive execution step. The "UP" event occurs when the value of the associated signal in the previous execution step was lower or equal to the level K and the current value is higher than the level K. The "DOWN" event, similarly to the "UP" event, occurs when the value of associated signal in the previous execution step was higher than the level K and the value in the current execution step is lower or equal to the level K.

However, considering the analysis of the signal value in two consecutive execution steps, it is possible to obtain several results when comparing with the threshold value K, supporting the possibility of having different modeling attitudes. It his sense, ten new types of event were defined resulting in a new set of twelve types of events, listed in Table 2.

In Table 2, the mathematical definition of the events and a graphical example of each of them are represented. In this new set of events the old UP event is defined as ">" and the old DOWN event is defined as "≤".

Each event represents the conclusion of the combination of two conditions: the comparison of the value of the signal with a certain value K in two consecutive execution steps.

4.2 Events on Delta One (ΔX)

These events can be seen as generated by the derivative of the equations of threshold events, were introduced in [Campos-Rebelo et al. 2013a] and called “momentum event”. They are defined as a variation of the difference of the signal along consecutive execution steps, instead of the variation of the value of the signal. These events can be associated with the concept “velocity” of the signal in contrast to previous events that are associated with the concept of position.

Similarly to the threshold events, two momentum event types were defined considering the difference on the variation of the signal in two consecutive execution steps. They are also called “UP” and “DOWN”, respectively. The momentum “UP” event occurs when the variation of the signal $|x_{n-1} - x_{n-2}|$ in the previous analysis step was lower or equal to the value K and the current variation $|x_n - x_{n-1}|$ is higher than the value K. Similarly, the momentum “DOWN” event occurs when the previous variation of the signal $|x_{n-1} - x_{n-2}|$ was higher to the value K and the current variation $|x_n - x_{n-1}|$ is lower or equal than the value K.

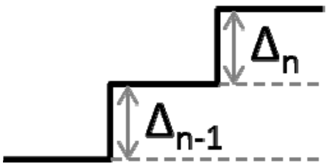


Fig. 1 Analysis of delta one events

Following the same strategy as in events on X, twelve events in ΔX are defined, each one obtained by derivation of the respective last variation event type, as presented in Table 2.

Table 2 presents the mathematical definition of the events and a graphical example on how to analyze each one is represented in Fig. 1, where Δn and Δn-1 are represented. These twelve events represent the elementary events associated in the variation one of the signal.

In this case, each of these events is represented based on two conditions related to a certain value K. A condition is related to the value of the last signal variation and the other one to the current variation of the signal. For these variations it is necessary to take into account values of the signal, in three consecutive steps (n, n-1 and n-2). Respective variations are determined as follows: $\Delta_n = X_n - X_{n-1}$ e $\Delta_{n-1} = X_{n-2} - X_{n-1}$.

4.3 Events on Delta n ($\Delta^n X$)

Delta two variations on the signal X can be defined recursively as differences on differences, and can also be associated with the concept of “acceleration”. Similarly, high delta orders can be defined, as it is possible to define events on the variation of the variations (for each event in each level n of variation, new event is obtained by deriving the event of the same type of variation n-1). However, physical interpretations for higher order delta events could be difficult to find associated to known physical phenomena. Likewise, the number of signal values to be considered for the analysis of the event increases with increasing degree of variation.

5 Delayed Events

In previous sections, all events have been presented considering the analysis of the signal on consecutive execution steps. This is the most used approach in order to detect the event in the shortest time. However, the detection of the event can be done considering different distances between execution steps, instead of two consecutive steps, for example, considering the value of the signal in the current step and in p execution steps before. The analysis of the event with a delay of a number of predefined steps can bring advantages, for example in the filtering of residual values in signal acquisition as when debouncing of the signal is necessary.

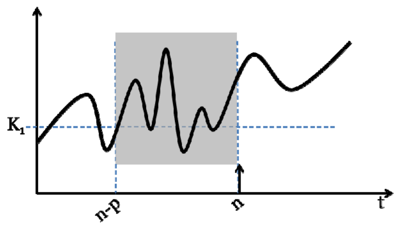


Fig. 2 Analysis of delayed events

Fig. 2 presents the window between two execution steps, where delayed event is a result of the analyses of the signal value in the execution steps n and n-p. The analysis of two consecutive values of the signal can be defined as a particular case of the delayed definitions where p = 1.

6 Application Example

In this section, an example adapted from [Campos-Rebelo et al. 2013b] is used to illustrate the use of some types of the presented events.

Due to its mechanical wear, there has been an increasingly drawn of technology companies to avoid using pushbuttons. In its place many technologies have been created as resistive buttons or the use of auxiliary sensors like light sensor, video camera or gyroscope. This example aims to produce a controller for a simple music player, which has only one on/off capacitive button and a gyroscope. All actions to control the player must be performed using the gyroscope built into the device.

The following play functions are implemented: the “PLAY/PAUSE” (that starts and pauses music playing), the “STOP” (that stops playing), and the “NEXT” (that moves to the next music in the list). It is still necessary to implement the functions of the music selection: the “PREV” (that changes the music that is being played for the previous music) and the “NEXT” (that changes the music that is being played for the next music), and finally the volume controllers: the “VOL.UP” and “VOL.DOWN”, which increases or decreases the output level of the player.

The gyroscope has three axis, as presented in Fig. 3 a), and the controls must be defined by rotation around them. The movements that are associated to each action are presented in Fig. 3 b).

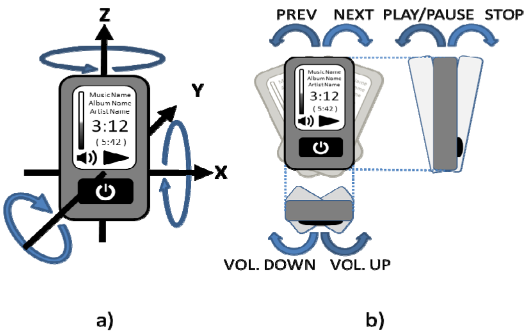


Fig. 3 Music Player informationa) Gyroscope axis representation, b) Representation of the acting movements

Modeling the player controller using IOPT Petri nets is obtained using a model with seven places, seventeen transitions, eight input events associated to three input signals, five output signals and two output events, as presented in Fig. 4. In the places ***vol_up_pl*** and ***vol_down_pl*** the player still is in playing mode like in ***playing*** place. Similarly, in the places ***vol_up_ps*** and ***vol_down_ps*** the player still is in pause mode like in ***pause*** place. Table 3 helps to understand the model, relating each action to its associated motion action.

All movements must have an acceleration higher than a certain value to activate associated action. Initial marking of the model considers place *stopped* marked (the player starts stopped). The user can move to playing mode shaking it with a hard x-axis rotation. In this mode, with the same x-axis movement’s, is possible to switch between playing and pause mode. In each two modes the user can change to next and preview item using a y-axis rotation. It is also possible to change the volume level. These movements are a little bit more complex. The volume starts changing with a hard z-axis rotation and keep changing until the z axis returns to center z-axis value.

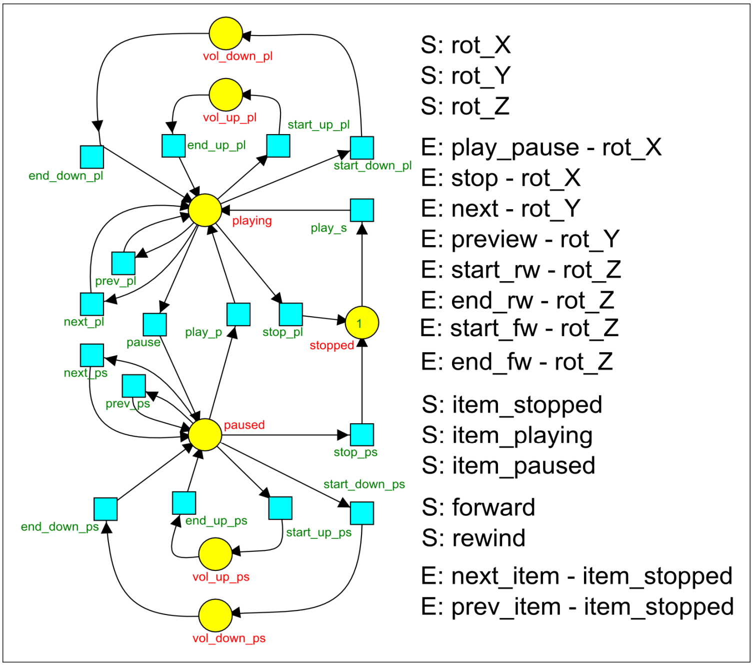


Fig. 4 IOPT Petri net model of the specified music player

Table 3 Association of actions to movements

Action	Movement
Play/Pause	Shake to front
Stop	Shake to back
Next	Shake to right
Prev	Shake to left
Vol. up	rotate to right
Vol. down	rotate to left

In playing or pause mode it is possible to move to stop mode. Finally, the change of volume level starts with the rotation to the specific side and continues increasing or decreasing until the player returns to the center position.

All input events are presented in **Błąd! Nieprawidłowy odsyłacz do zakładki: wskazuje na nią samą.**. For each event, the type of event, the axis where it acts, and the transitions where the event is associated as a fire condition are presented.

Table 4 Model events definition

Event Name	Axis	Type	Transition
Play/Pause	X	Momentum	play_s play_p pause
Stop	X	Momentum	stop_pl stop_ps
Next	Y	Momentum	next_ps next_pl
Preview	Y	Momentum	prev_ps prev_pl
Start_up	Z	Momentum	start_up_pl start_up_ps
End-up	Z	Threshold	end_up_pl end_up_ps
Start_Down	Z	Momentum	start_down_pl start_down_ps
End_Down	Z	Threshold	end_down_pl end_down_ps

7 Discussion

The event types proposed in [Gomes et al. 2007] are very restricted in terms of variation levels and in terms of types of events associated to each level, as well. With those events, it is only possible to detect events that occur by crossing a threshold level. Even among these levels, only two different behaviors can be detected. This leads to large, complex and difficult to understand models, where all other possible behaviors need to be created with sub-models with signal conditions associated.

In the presented example, which aims to control a music player using only movements, is possible to see those restrictions. In this case it is necessary to take into account the speed at which the player is shaken allowing to distinguish between an intended action from the regular movements of the device (due to user movements).

The new types of events presented allow defining events to a higher number of behaviors of the signal, enabling the creation of more compact models. In this example different kinds of events relying on different levels of variation are used. Whenever the player is shaken means that an action is required; only when the player is shaken with more than a certain speed an event is triggered.

It is also possible to create a window on the signal, allowing the analysis of events considering a higher number of execution steps. This analysis may have advantages when trying to analyze a signal with noise or even when it is necessary to receive the same values that are received in another system with a slower variation.

These characteristics leads to models that are easier to produce, understand, and maintain. This allows an easily integration in modeling tools, including with automatic code generation, as the ones available through IOPT-Tools at http://gres.uninova.pt applicable to IOPT nets.

8 Conclusions

Using proposed events in cooperation with IOPT nets models enables the creation of smaller models (for the same system), which consequently are more robust and easier to read and interpret.

The analysis of events with delay provides the possibility to analyze a signal in a time window including several execution steps, allowing the detection of specific signal evolutions.

Supported by proposed characterization of events, it is foreseen to consider composition of events and signal evolution dependencies allowing further improvements in terms of compactness and expressiveness of models.

These event updates are proposed intrinsically matched with IOPT net models, but can also be applied and useful in other classes of Petri nets, both high-level and low-level Petri nets classes or even in other event-based modeling formalisms.

Acknowledgment. The first author was supported by a Grant from project PTDC/EEI-AUT/2641/2012, financed by Portuguese Agency" FCT - Fundação para a Ciência e a Tecnologia".

This work was partially financed by Portuguese Agency" FCT - Fundação para a Ciência e a Tecnologia" in the framework of projects PEst-OE/EEI/UI0066/2011 and PTDC/EEI-AUT/2641/2012.

References

[Barros and Gomes 2004] Barros, J.P., Gomes, L.: Net model composition and modification by net operations: a pragmatic approach. In: Proc. 2nd IEEE International Conference on Industrial Informatics, Berlin, Germany, pp. 309–314 (2004)

[Borger and Stark 2003] Borger, E., Stark, R.F.: Abstract state machines: a method for high-level system design and analysis. Springer, Heidelberg (2003)

[Campos-Rebelo et al. 2013a] Campos-Rebelo, R., Costa, A., Gomes, L.: On Structuring Events for IOPT Net Models. In: Proc. 4th Doctoral Conference on Computing, Electrical and Industrial Systems, Lisbon, Portugal, pp. 229–238 (2013)

[Campos-Rebelo et al. 2013b] Campos-Rebelo, R., Costa, A., Gomes, L.: Events for Human-System Interaction modeling with IOPT Petri nets. In: Proc. 6th International Conference on Human System Interaction, Gdansk, Poland, pp. 56–61 (2013)

[Costa and Gomes 2009] Costa, A., Gomes, L.: Petri net partitioning using net splitting operation. In: Proc. 7th IEEE International Conference on Industrial Informatics, Cardiff, Wales, UK, pp. 204–209 (2009)

[David and Alla 1992] David, R., Alla, H.: Petri Nets and Grafcet: Tools for Modelling Discrete Event Systems. Prentice-Hall (1992)

[David and Alla 2010] David, R., Alla, H.: Discrete, Continuous, and Hybrid Petri Nets. Springer (2010)

[Desel and Reisig 1998] Desel, J., Reisig, W.: Place/transition Petri nets. In: Reisig, W., Rozenberg, G. (eds.) APN 1998. LNCS, vol. 1491, pp. 122–173. Springer, Heidelberg (1998)

[Frey 2002] Frey, G.: Design and formal analysis of Petri net based logic control algorithms. Shaker Verlag (2002)

[Frey and Minas 2000] Frey, G., Minas, M.: Editing, visualizing, and implementing signal interpreted petri nets. In: Proc. 15th German Workshop on Algorithms and Tools for Petri Nets, pp. 57–62 (2000)

[Gomes and Barros 2005] Gomes, L., Barros, J.P.: Structuring and composability issues in Petri nets modeling. In: Proc. 3rd IEEE International Conference on Industrial Informatics, Perth, Australia, vol. 1, pp. 112–123 (2005)

[Gomes et al. 2007] Gomes, L., Barros, J.P., Costa, A., Nunes, R.: The Input-Output Place-Transition Petri Net Class and Associated Tools. In: Proc. 5th IEEE International Conference on Industrial Informatics, Vienna, Austria, pp. 509–514 (2007)

[Harel 1987] Harel, D.: Statecharts: A visual formalism for complex systems. Sci. Comput. Program. 8, 231–274 (1987)

[Moalla et al. 1978] Moalla, M., Pulou, J., Sifakis, J.: Synchronized Petri nets: A model for the description of non-autonomous systems. In: Mathematical Foundations of Computer Science, pp. 374–384 (1978)

Emotion Recognition and Its Applications

A. Kołakowska, A. Landowska, M. Szwoch, W. Szwoch, and M.R. Wróbel

Faculty of Electronics, Telecommunications and Informatics,
Gdansk University of Technology, Gdańsk, Poland
{agatakol,nailie,szwoch,wszwoch,wrobel}@eti.pg.gda.pl

Abstract. This paper aims at illustrating diversity of possible emotion recognition applications. It provides concise review of affect recognition methods based on different inputs such as biometrics, video channel or behavioral data. It proposes a set of research scenarios of emotion recognition applications in the following domains: software engineering, website customization, education, and gaming. The scenarios show complexity and problems of applying affective computing in different domains. Analysis of the scenarios allows drawing some conclusions on challenges of automatic recognition that have to be addressed by further research.

1 Introduction

Inevitably, emotions play an important role not only in our relations with other people but also in the way we use computers. Affective computing is a domain that focuses on user emotions while he or she interacts with computers and applications. As emotional state of a person may affect concentration, task solving and decision making skills, the vision of affective computing is to make systems able to recognize human emotions and influence them in order to enhance productivity and effectiveness of working with computers.

A challenging problem of automatic recognition of human affect has become a research field involving more and more scientists specializing in different areas such as artificial intelligence, computer vision, psychology, physiology etc. Its popularity arises from vast areas of possible applications. In this paper we focus on application of affective computing methods in software engineering, website customization, education and gaming. We propose a number of research scenarios to evaluate the possibility of using emotion recognition methods in these areas.

Section 2 briefly presents the idea and the methods of emotion recognition. Sections 3-6 describe the proposed scenarios to be applied in the four, just mentioned, different domains. Section 7 gives some conclusions on challenges, which have to be solved to implement the scenarios in a real environment.

2 Emotion Recognition

The goal of human emotion recognition is to automatically classify user's temporal emotional state basing on some input data. There are dozens of definitions of

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_5, © Springer International Publishing Switzerland 2014

emotions [Picard 2004]. In this paper we adopt the following distinction based on time: *emotion* is a reaction to stimuli that lasts for seconds or minutes, *mood* is an emotional state that lasts for hours or days and *personality* is an inclination to feel certain emotions. We use the term 'emotional state' to indicate current (temporary) state of a person irrespective of its origin (stimuli, mood or personality).

The stated goal may be achieved by one of many types of classifiers developed in the field of pattern recognition. The approach assumes several stages of classifier's construction which are: data acquisition and feature extraction, creation of a training set containing labeled data and classifier's learning.

In our research we assume that emotion recognition will be based on multimodal inputs: physiological sensors, video, depth sensors and standard input devices. This approach proved to provide more information to the recognition process as different data channels deliver valuable complementary information eliminating potential drawbacks of any individual input [Kapoor and Picard 2005].

Physiological sensors might be used to measure skin conductance, blood volume pulse, muscle impulses, respiratory signal, temperature or heart rate [Szwoch W 2013]. They are non-invasive but sometimes intrusive or not comfortable for users due to special equipment, which is required. Thus it is not possible to apply it in real-life situations when we want to determine emotions of people during usual learning or working processes. But taking specialized measurements of physiological signals can be used in some scenarios as well as for the enhancement or verification of a classifier's accuracy. A great many features may be extracted from physiological signals by calculating their mean, standard deviation, difference, Fourier transform, wavelet transform, high frequency and low frequency power, entropy, thresholding, peak detection etc. [Jerritta et al. 2011].

Video cameras and depth sensors deliver us significant information on facial expression in a non intrusive way. Face expression may be represented by geometric or appearance features, parameters extracted from transformed images such as eigenfaces, dynamic models and 3D models. The difficulty of this approach is the need of image preprocessing and complex pattern recognition algorithms. One of the main problems with facial expression recognition is that it usually works well only in the case of a posed behavior and proper lightning [Szwoch M 2013]. Depth sensors, which usually use non-visual infrared light technology, are generally resistant to insufficient and uneven lightning conditions. Since introducing Microsoft Kinect depth sensors are also used for recognition of human poses, gestures and movements [Ren et al. 2011].

Standard input devices, such as keyboard and mouse, enable a completely unobtrusive way of collecting data, because no special hardware is needed and moreover it may be done during users' usual computer activities. Features extracted from keystrokes may be divided into timing and frequency parameters. Mouse characteristics include both clicking and cursor movement measurements [Kołakowska 2013].

The collected data has to be labeled, i.e. an emotion has to be assigned to each data sample. Most of the emotion recognition algorithms use discrete or dimensional models for affective state representation. The best recognized discrete model is Ekman six basic emotions, including: joy, sadness, fear, anger, disgust and surprise, whereas a three-dimensional PAD model proposes to represent an emotional state as a combination of valence, arousal and dominance values. Usually the labels for the training data are assigned according to specially designed questionnaires given to the users or on the basis of independent observers' evaluations. However, such labeling may not be objective, which in turn may result in poor accuracy of the trained system. To avoid such situation we are going to validate the labels provided by humans with the labels assigned on the basis of physiological measurements.

A great many machine learning algorithms have been already applied in the task of emotion recognition, e.g. SVM, decision trees, linear discriminant analysis, Bayesian networks, naive Bayes, neural networks [Zeng et al. 2009]. An ideal emotion recognition method in the proposed real-life applications would be a combination of adaptive classifiers which could cope with high number of features of different types and would be able to improve its effectiveness with increasing amounts of training data continuously recorded during users' typical activities.

3 Software Engineering

The main purpose of this part of our research is to use emotion recognition based on multimodal inputs to improve some aspects of software engineering process and to overcome the limitations of usability questionnaires [Kołakowska et al. 2013]. We focus on two areas of application in software engineering: usability testing and development process improvement.

3.1 Extended Software Usability Testing

There is a lot of evidence that emotions influence human interactions with software products. There is also a record of investigation on how products can influence human feelings [Hill 2009] which make people buy or not. Therefore investigating emotions induced by products is an object of interest of designers, investors, producers and customers, as well.

Software usability depends on multiple quality factors, such as functionality, reliability, interface design, performance and so on. All of the quality indicators can be improved indefinitely, but there is a point to stop optimizing – it is a customer satisfaction. Measuring this satisfaction with questionnaires may be misleading. We propose to extend usability testing with emotion recognition based on multimodal inputs. We have defined the following test scenarios with required emotional state distinctions depending on the scenario.

Scenario S1. First Impression Test

First impression is a state that is evoked mainly by visual input in human-systems interaction and is created in a very short time (approximately 5 seconds). Research shows, that in web page design first impression is a good predictor of 10-minute usability opinion [Linggaard et al. 2006]. Many of the websites will not have any more time to make an impression than these 5 seconds – the first impression makes the users stay or quit. In first impression testing the most important distinction is to differentiate user's *interest* (*excitement*) from *boredom* or *disgust*. This scenario is especially dedicated to web page usability testing.

Scenario S2. Task-Based Usability Test
The second usability scenario proposed uses cognitive walkthrough method [Blackmon et al. 2002], which is a task-based approach. Software usability evaluation in this method usually involves identification of typical tasks (which may be extracted from use case models) and the optimal processes for performing them (possibly derived from dynamic models, user stories or user instructions). The representative user group performs the tasks in a controlled environment with camera recording, biometric sensors and keystroke analysis tools. Registered channels are then a subject to further analysis of usability and emotional state fluctuation. This scenario is dedicated rather to applications designed to help the user to perform specific tasks and not for entertainment or content access systems. The purpose of emotion recognition in task–based usability testing is to differentiate *frustration* from *empowerment*.

Scenario S3. Free Interaction Test
The third usability scenario proposed is based on free interaction with application, which is supposed to evaluate overall user experience. There are no pre-defined tasks to be performed by a representative user group; instead they are asked to freely interact with application under examination. This scenario is dedicated for entertainment and content access systems, but other applications may also benefit. The objective of emotion recognition in this scenario is the distinction of *engagement* from *discouragement*.

Scenario S4. Comparative Test
Comparative scenario is a selection or combination of methods used in previously defined scenarios performed on two software versions or on the application and the main competitive software product.

3.2 Development Process Improvement

In each segment of the job market the most valuable employees are those who are highly productive and deliver high-quality products or services. A similar situation is with respected software developers [Wróbel 2013]. Employers require high work efficiency and high quality code. Unfortunately, these two requirements are often in conflict, as a computer program developed under time pressure is usually of low quality [MacCormack et al. 2003].

The purpose of this study is to verify the hypothesis that emotions have significant impact on software quality and developers' productivity. The aim is to answer the question on correlation between employee's emotional state and his work efficiency as well as quality of the developed software. The study will also determine the emotional states supporting the work of IT professionals.

We have defined four research scenarios to explore multiple factors of the relationship between programmers' emotional states and their work, including the work environment, personal productivity and quality of the developed code.

Scenario S5. IDE Usability Comparison

This scenario is an adoption of task-based usability test described in scenario 2. Integrated development environments (IDE) are the essential tools used by developers. Their advantages and disadvantages can significantly affect the emotional state of the users. Research will be conducted in a laboratory environment with biometric sensors. User group will be represented by both novice programmers - ICT students as well as ICT stuff with years of experience. The object of the research will be a set of popular IDEs. A developer will perform a series of programming tasks, such as compiling, debugging, refactoring, on three randomly selected environments, excluding those he or she uses the most frequently. Tests will evaluate the quality of those IDEs. However, the collected data will be used to investigate the individually differentiated impact of different problems encountered in an IDE on developers' emotions. The goal of this scenario is to distinguish between *frustration* and *empowerment*.

Scenario S6. Productivity and Emotions

This scenario is designed to answer the question of whether and how an emotional state affects the productivity of the programmer. The research will be conducted in a laboratory environment. Behavior of the maliciously prepared environment will evoke developers' emotions that may affect their productivity, measured for example by the number of lines of code per hour of work. In the first place *stress* associated with time pressure and *boredom* will be induced. The analysis of the collected data will determine the optimal emotional space for developers productivity.

Scenario S7. Code Quality and Emotions

This scenario, despite similarities to the previous one, should not be conducted in a laboratory environment. It is hard to accurately evaluate the quality of the code developed during a short test. Therefore, to provide reliable results, this scenario requires continuous monitoring of the emotional state of the programmer and the collection of incremental versions of the source code. Only the cross-examination of emotional states and source code may lead to the designation of the correlation between quality and emotions. In this scenario, it is essential to detect emotions such as *empowerment*, *frustration* and *stress*.

Scenario S8. Tele- and Office-Working Comparison

The last scenario is designed to detect whether there are differences in emotional states of programmers when working in office or at home. The number of telecommuters is growing rapidly in recent years. This research should be conducted in real work environments. This will be possible only after the development of reliable, non-intrusive methods of user emotional states recognition. The objective of emotion recognition is to detect the whole range of emotions, particularly all those identified in the previous scenarios.

Scenarios S5 and S6 can be conducted in a laboratory environment. In this research, it is possible to use biometric sensors to detect emotions of programmers. This will deliver more accurate recognition of emotions than with the previously developed non-intrusive methods. However, the implementation of the other two scenarios (S7 and S8) will be possible only using non-intrusive methods for detecting the emotions of computer users.

As the computer is the primary working environment of programmers, the implementation of emotion recognition mechanisms in human-computer interface is the natural choice. However, research, as well as the proposed scenarios (except scenario S5), are sufficiently universal to be applied to many professions.

4 Education and e-Education

It has been already reported, that some emotional states support learning processes and other suppress them [Hudlicka 2003, Picard 2003, Sheng et al. 2010]. The distinction of the two groups of emotional states in some cases is not obvious, for example such positive mood as hilarity is not good for learning processes, while slightly negative emotional states foster critical thinking and are appropriate for analytical tasks [Landowska 2013]. Automatic emotion recognition can help to explore these phenomena by making assessments of learner emotional states more objective than typical questionnaire-based investigations.

Scenario E1. Emotional Templates of Educational Tasks
The purpose of this scenario is investigation on emotional states that occur during different types of educational tasks. This investigation aims at identifying emotional templates of educational tasks, which can be defined as distinguishable sets of effective and counter-productive emotional states for solving specific task types. To perform this investigation a representative set of educational tasks should be prepared and both learners' performance in task execution and his or her emotional state should be measured. Analysis of the correlation between performance and emotional states would enable to justify statements on effective and counter-productive emotions for specific task types; however, a significant number of respondents should be engaged in order to make the thesis reliable. Information on effective emotional states can be then used in educational problems diagnosis, educational tool design or in further exploration of educational processes.

Scenario E2. Emotional Stereotypes of Learners
Emotionality is one of the elements of human personality and may differ significantly based on in-born temper, previous experience and socialization process. However, some emotional reactions are common for people living in one culture or having the same experience and similar characteristics is expected also in

educational processes. Learner affective stereotype is a definition of typical emotional states that might be observed in educational settings. It is expected, that novice learners will more frequently show symptoms of frustration, while more experienced ones could feel boredom. To support that thesis with research results, emotional states of different (novice vs. experienced) students will be measured and recognized while they perform the same tasks set of growing difficulty. Learners' stereotypes can be then used in e-educational environments to adapt learning paths or interaction models, when no individual information on a user is available.

Scenario E3. Evaluation of Educational Resources
The goal of this scenario is to evaluate educational resources, especially those prepared for self-learning. In distance and electronic education one of the critical success factors is learner's discipline in following provided learning path. When one fails due to fluctuation of motivation and attention, the learning process is paused or even abandoned. One of the frequently launched causes for course resignation is: "Boring resources". In this scenario observation of student's interaction with resources is combined with monitoring his or her emotional state in order to identify parts of resources that cause boredom. That information may be then used to remove weak points and improve overall resource quality. A set of different types of educational resources including recorded lectures, screencasts and interactive materials will be investigated. This scenario might be also used for quality evaluation of resources provided by virtual universities and other distance learning environments.

Scenario E4. Usability Testing of Educational Tools
In this scenario usability of educational tools is evaluated. Software usability tests are usually based on eye-tracking techniques and we propose to extend it with user emotion recognition, which can be valuable information while evaluating user experience [Kołakowska et al. 2013]. Typical tasks performed with educational tools include: accessing the tool, searching resources, launching resource, performing interactive tasks, viewing results or feedback information, communicating with teachers or class mates and more. More specific task description for this scenario will be performed using cognitive walkthrough method [Blackmon et al. 2002]. Then representative group of students will be asked to perform some tasks in controlled environment that will additionally record and recognize their emotional states. Information on affect and its fluctuations (especially identification of frustration) can help to improve software products that are designed to support learning processes.

5 Enhanced Websites Customization

With the grow of the Internet, service providers collect more and more information about their users. Based on these data, content, layout and ads are displayed according to a user's profile created. Adding information about the emotions of users could provide more accurate personality models of the users.

We have defined two scenarios, the first to explore how emotions of web surfers influence their behavior on websites, the second to determine what emotions are triggered by different types of on-line ads.

Scenario W1. Affective customization
The purpose of this scenario is to determine how emotions affect the way users consume information on websites. The main expected outcome of this investigation is a list of factors that, in conjunction with a specific mood, increase the time spent on the website. The study will examine the impact of the following factors: page layout, content and photo sets. Users are intended to review a set of prepared web pages (with different values of factors). Based on biometric sensors and cameras their mood will be recognized. These data will be aggregated with information about the time spent on each site. Analysis of the results will help to determine, for example, which website layout would interest a bored person and which is the best for an angry one.

Scenario W2. Advertisement Reaction Model
The revenue model for a significant number of websites is based on the on-line advertising [Dubosson-Torbay et al. 2002]. However, users describe them as uninformative, ineffective and very intrusive [McCoy et al. 2007].

The aim of this scenario is to find the most eye-catching and interesting advertisement types for different groups of Internet users. In the laboratory environment, the emotional reaction will be measured in response to various formats of on-line ads. Additionally, using eye-tracking tool, information about the user focus on adverts will be collected. A set of information about time elapsed before noticing the ads, user emotional response and duration of focused attention will allow choosing the appropriate type of advertising, depending on the target audience.

6 Video Games

Video games belong to the wide area of entertainment applications. Thus, assuming the existence of human emotions, and in fact basing on them, they attempt to make a player become emotionally attached with them. As the primary goal of a video game is to entertain the player [Adams 2009], each video game tries to allow the player to fulfill his or her “dream”. Standard video games try to do it in different ways depending on their genre and involving such elements as good gameplay, immersing storytelling, novelty, graphics and so on [Adams 2009].

Although video games belong to applications in which emotions naturally play an important role, only few of them try to incorporate their players’ affective state into the gameplay. Such games can be referred as affective or more properly affect-aware games. The importance of affect in delivering engaging experiences in entertainment and educational games is well recognized. Potential profits for affect-aware video games are not to be underestimated. Unfortunately, this “affect-awareness” is usually “statically” built-in the game at its development stage basing on the assumed model of so called representative player [Adams 2009]. There are two problems with such attitude. Firstly, each player differs in some way from that averaged model. Secondly, and more important, player's emotional state can change even radically from session to session making it almost impossible to predict current user emotions at the development stage.

There are a lot of reasons that can influence human's behavior during playing the game. They could be divided into factors connected with the game, such as increasing monotony or becoming an accustomed player, and to game independent factors which are connected with current physical and mental conditions of the player. The first group of reasons may be in some extent predicted or estimated by the game designer but that is impossible to the reasons belonging to the other group. That is why the real-time recognition of player's affect may become so important for video games industry in the nearest future. Video games that are able to dynamically react to the recognized current player's emotions may be called truly affect aware video games.

Scenario G1. Emotional Templates in Video Games

Though generally people are able to express wide spectrum of emotions, not all of them are observed while playing video games. Expressed emotions depend on many factors such as game's genre, player's genre and experience, and on many other predictable or even unpredictable factors [Adams 2009]. Moreover, emotional reaction in the situation highly depends on individual personality and even current mood or unpredictable external factors.

The goal of this scenario is to verify which emotions are common while playing video games and which are rare. The scenario will also allow verifying the hypothesis that for specific game genres and specific group of game players some expressions are more common than others.

A typical test in this scenario will cover recording of players' emotions while playing several different video games with specially prepared scenes. Additional questionnaire will provide further information on the player's age, gender and gaming experience.

Scenario G2. Stimuli Adaptation

Games try to attract the player's attention to tie him or her as long as possible. Game developers use an interesting story, high quality graphics and an intense arousal to keep the attention of the player.

The goal of this scenario is to verify, whether long, frequently repeated stimulation causes weakening of player's reaction due to stimuli adaptation. After a specified time, the player can stop to react on stimuli, and it is possible that he or she will be expecting more and more powerful experiences. The question is whether it is necessary to keep the user's attention at the highest possible level, or whether we should interlace periods of intense emotional arousal with periods of silence, which will allow the player's senses to rest.

The scenario assumes that a study group of players will be exposed to intense stimuli during the game. The time after which the player's reaction to stimuli disappears and also the moment of weariness of the game will be tested. During the experiment, the observer and the player will determine the moment of time or time period in which there was a loss of response to stimuli.

Scenario G3. Player Reaction to External Signals on Different Levels of Immersion
Sometimes game players are so strongly immersed into the virtual world, that they stop to notice the real world around them. It is important to monitor the depth of players' immersion to detect when they stop responding to external stimuli. This can help for example to protect players against addiction. When the player is too absorbed in the play, affect-aware game can reduce its attractiveness, causing gentle return to the real world.

During the experiment player's reaction to external stimuli will be investigated. The problem may be to get the appropriate involvement in the game. Player reaction (rapid vs. slow) will determine the degree of his engagement. During the experiment, various types of disturbing events are planned to be used, such as, acoustic, haptic and other.

Scenario G4. The Influence of Affect Feedback on Player's Satisfaction
Sometimes a video game can become boring or too stressing for the player. Detecting such situations would allow giving a proper feedback to the player, changing the current gameplay character. On one hand such capability could improve satisfaction of experienced players and on the other it could protect novice or young players from excessive violence. Adapting feedback to stimulate the player can also be used in therapy in suppressing negative emotions by proper stimulation.

The purpose of the scenario is to check whether the affect-aware games can increase players' satisfaction from playing. For this scenario a specially developed affect aware video game will be used. The scenario assumes the use of a questionnaire, in which the player will determine the level of satisfaction within the game when increasing and decreasing difficulty according to the detected emotions.

7 Conclusions

The paper has presented a number of possible applications of emotion recognition in different areas such as software engineering, website customization, education and gaming. Although some of the presented research scenarios are ready to be applied, in the case of most of them a few challenging problems still have to be solved. First of all the limitations of emotion representation models not always let us precisely describe the actual feelings of a user. We often do not realize the real set of possible emotional states, which should be considered in an application. Even if we are able to define this set of emotions, the fuzzy nature of emotional states and their instability along time entail subsequent difficulties. Moreover the quality of the training data and the way emotion labels are assigned strongly influence the results of the training algorithm. Finally the accuracy of the recognition process often does not fulfill the requirements of a system working in a real environment. All this does not let us ignore high uncertainty of emotion recognition methods, especially when combining separate models. This is an open research problem requiring further investigation.

Acknowledgments. The research leading to these results has received funding from the Polish-Norwegian Research Programme operated by the National Centre for Research and Development under the Norwegian Financial Mechanism 2009-2014 in the frame of Project Contract No Pol-Nor/210629/51/2013

References

[Adams 2009] Adams, E.: Fundamentals of Game Design. New Riders Publishing (2009)

[Blackmon et al. 2002] Blackmon, M., Polson, P., Kitajima, M., Lewis, C.: Cognitive Walkthrough for the Web. In: Human Factors in Computing Systems: Proceedings of CHI 2002, pp. 463–470. ACM, New York (2002)

[Dubosson-Torbay et al. 2002] Dubosson-Torbay, M., Osterwalder, A., Pigneur, Y.: E-business model design, classification, and measurements. Thunderbird International Business Review 44(1), 5–23 (2002)

[Hill 2009] Hill, D.: Emotionomics: Winning Hearts and Minds. Kogan Page Publishers, USA (2009)

[Hudlicka 2003] Hudlicka, E.: To feel or not to feel: The role of affect in human-computer interaction. Int. J. Hum.-Comput. St. 59(1), 1–32 (2003)

[Jerritta et al. 2011] Jerritta, S., Murugappan, M., Nagarajan, R., Wan, K.: Physiological signals based human emotion recognition: a review. In: Proc. IEEE 7th International Colloquium on Signal Processing and its Applications (2011)

[Kapoor and Picard 2005] Kapoor, A., Picard, R.W.: Multimodal affect recognition in learning environments. In: Proceedings of the 13th Annual ACM International Conference on Multimedia, pp. 677–682 (2005)

[Kołakowska et al. 2013] Kołakowska, A., Landowska, A., Szwoch, M., Szwoch, W., Wróbel, M.: Emotion Recognition and its Application in Software Engineering. In: Proc. 6th International Conference on Human Systems Interaction, pp. 532–539 (2013)

[Kołakowska 2013] Kołakowska, A.: A review of emotion recognition methods based on keystroke dynamics and mouse movements. In: Proc. 6th International Conference on Human Systems Interaction, pp. 548–555 (2013)

[Landowska 2013] Landowska, A.: Affective computing and affective learning – methods, tools and prospects, EduAction. Electronic Education Magazine (2013) (in printing)

[Linggaard et al. 2006] Linggaard, G., Fernandes, G., Dudek, C., Brown, J.: Attention web designers: You have 50 miliseconds to make a good first impression. Behaviour and Information Technology 25(2), 115–126 (2006)

[MacCormack et al. 2003] MacCormack, A., Kemerer, C.F., Cusumano, M., Crandall, B.: Trade-offs between productivity and quality in selecting software development practices. IEEE Software 20(5), 78–85 (2003)

[McCoy et al. 2007] McCoy, S., Everard, A., Polak, P., Galletta, D.F.: The effects of online advertising. Communications of the ACM 50(3), 84–88 (2007)

[Picard 2003] Picard, R.: Affective computing: challenges. Int. J. Hum.-Comput. St. 59(1), 55–64 (2003)

[Picard et al. 2004] Picard, R., Papert, S., Bender, W., Blumberg, B., Breazeal, C., Cavallo, D., Strohecker, C.: Affective learning – a manifesto. BT Technology Journal 22(4), 253–269 (2004)

[Ren et al. 2011] Ren, Z., Meng, J., Yuan, J., Zhang, Z.: Robust hand gesture recognition with kinect sensor. In: Proc. 19th ACM International Conference on Multimedia, pp. 759–760 (2011)

[Sheng et al. 2010] Sheng, Z., Zhu-Ying, L., Wan-Xin, D.: The model of e-learning based on affective computing. In: Proc. 3rd IEEE International Conference on Advanced Computer Theory and Engineering, p. V3-269 (2010)

[Szwoch M 2013] Szwoch, M.: FEEDB: a Multimodal Database of Facial Expressions and Emotions. In: Proc. 6th International Conference on Human Systems Interaction, pp. 524–531 (2013)

[Szwoch W 2013] Szwoch, W.: Using Physiological Signals for Emotion. In: Proc. 6th International Conference on Human Systems Interaction, pp. 556–561 (2013)

[Wróbel 2013] Wróbel, M.R.: Emotions in the software development. In: Proc. 6th International Conference on Human Systems Interaction, pp. 518–523 (2013)

[Zeng et al. 2009] Zeng, Z., Pantic, M., Roisman, G., Huang, T.: A survey of affect recognition methods: Audio, visual, and spontaneous expressions. IEEE T. Pattern Anal. 31(1), 39–58 (2009)

Analysis and Comparison of Classification Methods in a Brain Machine Interface

E. López-Arce[1], R.Q. Fuentes-Aguilar[1,2],
I. Figueroa-García[3], and A. García-González[1,2]

[1] Biomedical Engineering Department, Tecnológico de Monterrey,
Campus Guadalajara Jalisco, México
eduardo.lopezarce@gmail.com
[2] Chair of Bioscience, Tecnológico de Monterrey, Campus Guadalajara Jalisco, México
alexgargo@itesm.mx
[3] Biomedical Signal and Image Computing Laboratory,
University of British Columbia, Canada
ivanf@ece.ubc.ca

Abstract. In this chapter, the analysis and comparison of four classification methods for a simplified Brain-Machine Interface (BMI) are presented. The BMI involves the use of only one electroencephalography (EEG) bipolar connection: O1-P3. The EEG signal is processed to extract features of the alpha wave (7.19~14.4Hz) in order to determine whether the subjects have their eyes closed or open. The signal processing technique is based on a Discrete Wavelet Transform (DWT) algorithm that allows processing the data on-line while the EEG signal is being recorded. The features that are extracted using DWT are then classified in order to estimate the subjects' eyes state. The four classifiers here compared are: Naive Bayes (NB), Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), and Adaptive Neuro-Fuzzy Inference System (ANFIS). A set of 2029 input vectors were considered for the comparison of the classifiers, they were divided as follows: 1771 for training (60%), 609 for testing (30%) and 203 for validation (10%). The best classification accuracy of the validation data was obtained by the NB classifier (93.59%) and the worst classification accuracy by the ANFIS classifier (88.17%). However, using a contingency table it was shown that the SVM obtained the best performance to classify eyes closed, which was consider the most important feature to classify in the BMI application used. The final output of the classifiers is the input of a microcontroller that generates a pulse-width modulation signal that opens or closes a robotic hand if the subjects' eyes are open or closed, respectively.

1 Introduction

A Brain Machine Interface (BMI) is a communication path between the brain and an external device [Becedas 2012]. Their development has increased particularly since last two decades; however, it is possible to identify a permanent basic

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_6, © Springer International Publishing Switzerland 2014

sequence to build them, involving the following stages: i) Brain signals acquisition, ii) Preprocessing (mainly filtering process), iii) Feature extraction and iv) Classification, which is essentially a pattern recognition system [Jain et. al. 2000]. Classification is the final part of the translation process between the brain´s command and its execution by a control interface [Bi et al. 2013]. In this sense, a classifier may be viewed as just another input/output device where the input is a set of measurement data and the output is the most likely class associated with that data set [Semmlow 2008]. In this chapter we present a comparative analysis concerning four supervised classifications methods in a Brain Machine Interface (BMI), supervision implies that the input vector is identified as a member of a predefined class. We have previous experience, implementing a basic BMI supported by online wavelet decomposition [López-Arce et al. 2013], in order to solving the problem of detection the condition of open and closed eyes by measuring surface EEG activity in the occipital zone. This simplified BMI allow us to obtain a platform of control easily implemented in an embedded device. However, we needed to improve the correct classification of extracted features from the filtered signal; it is known that this is one of the current challenges in the BMI area [Bi et al. 2013]. Classification algorithms have to keep a reduced complexity and a high generalization capacity in order to satisfy an appropriated accuracy and an efficient practical implementation (real time), at the same time they have to deal with the curse of dimensionality, which means a large training set is not required for acceptable results, even with very high dimensional feature vectors [Jain et. al. 2000]. In this chapter, the four analyzed algorithms are: Naive Bayes classifier, Linear Discriminant Analysis, Support Vector Machine and Artificial Neural Fuzzy Inference System. All these methods have been reported in literature as part of different BMI structures [Jahnkhani et al. 2008; Hsu 2012; Nicholas-Alonso and Gómez-Gil 2012]. We focused our attention in their ability to assess correctly each command, considering the percentage of specificity and sensibility in each case, that means, their faculty to detect in a correct way the corresponding class, and without the presence of false positive (eyes closed when they are open) or false negative (eyes open when they are closed). Traditionally only the average error is analyzed [Becedas 2012]. As will be discussed, this traditional paradigm could mask important errors.

The chapter is distributed in a presentation of basic BMI concepts, emphasizing the classification process, a methodology and an experimental set-up explanation, the classification results and a comparative discussion. Finally some remarks and conclusions about the behavior of the proposed structures are presented.

1.1 Some Basic BMI Concepts

Signals in BMI, and Their Preprocessing. The signal acquisition of the brain can be obtained by invasive or noninvasive procedures. Our attention is focused in a noninvasive technique, the electroencephalography (EEG). EEG shows the highest temporal resolution for a noninvasive technique, which makes it attractive for real time implementations. More over EEG is the cheapest one, and highly portable [Becedas, 2012]. The raw EEG preprocessing consists in a filtering process over the raw signal or a digital filtering applied after digitalization, the typical bandwidth after this process is of 0.1 to 100 Hz, plus removing the breathing interference and the direct current component [Majumdar 2011].

Feature Extraction. The feature extraction stage identifies discriminative information in the brain signals. The signal is mapped onto a vector containing effective and discriminant features from the observed signals [Nicholas-Alonso and Gómez-Gil 2012]. It is possible to divide feature extraction methods into three categories: dimensional reduction methods, time and/or frequency methods and parametric modeling. In this chapter, a time-frequency method is implemented, the wavelet transform, a power useful tool for non-stationary signals, which decompose the signal in both the time and the frequency domain at multiple resolutions. As a matter of fact, we use the Discrete Wavelet Transform (DWT), which was introduced to reduce this redundancy and complexity. The DWT translates and dilates the mother wavelet in certain discrete values only [Gandhi et al. 2011].

Classification. Classification is a form of machine learning, and the complexity of a classification algorithm is often referred to as *machine capacity* [Semmlow 2008]. Increasing machine capacity leads to a more complex decision boundary between classes. If the classifier has more capacity than the appropriate for the data, it will be overtrained for the training data, performing well on the training set, but not generalizing to classify a different data set. Usually, the number of free parameters increases if the machine capacity is higher. As is reported in [Bi et al. 2013; Nicholas-Alonso and Gómez-Gil. 2012] a variety of classifiers have been used to translate extracted features from EEG signals into an output command in BMI´s, four of them are described below.

Naive Bayes Classifier (NB). A naive Bayes classifier is a probabilistic classifier based on applying Bayes' theorem, is a classifier which aims to assign, with the highest probability, an observed feature vector $\boldsymbol{x}$ from its class $\boldsymbol{c}$. It assumes that the predictive attributes are conditionally independent given the class and secondly, that the values of numeric attributes are normally distributed within each class [Jain et. al. 2000;Nicholas-Alonso and Gómez-Gil 2012]. Let **C** be the random variable denoting the class of an instance and **X** be a vector of random variables denoting the observed attribute values. Let $\boldsymbol{c}$ be a particular class label and $\boldsymbol{x}$ represent a particular observed attribute value. If we have a test case $\boldsymbol{x}$ to classify, the probability of each class given the vector of observed values for the predictive attributes may be obtained using the Bayes' theorem, and then predicting the most probable class, as follows.

$$p(C = c \mid X = x) = \frac{p(C = c)\, p(X = x \mid C = c)}{p(X = x)} \tag{1}$$

Linear Discriminant Analysis (LDA) [Semmlow 2008]. It is a simple classifier with acceptable accuracy and without high computation requirements. Linear classifier assumes that classes are linearly separable. The class predicted by a linear discriminator is given by the output of a linear equation:

$$y = \sum_{i=1}^{m} x_i w_i + b \tag{2}$$

$\mathbf{x_i}$ are the input variables. The classifiers free parameters include $\mathbf{w_i}$ (weights) and the threshold **b,** there exist many methods to compute $\mathbf{w_i}$ [Semmlow 2008; Nicholas-Alonso and Gómez-Gil. 2012].

Support Vector Machine. It is primarily a two-class classifier, similar to LDA, but in contrast to LDA, SVM is based on an optimization criterion, here is the width of the margin between the classes, i.e., the empty area around the decision boundary defined by the distance to the nearest training patterns [Nicholas-Alonso and Gómez-Gil. 2012]. These patterns, called support vectors, finally define the classification function. Their number is minimized by maximizing the margin. In SVM analyses, the classes are assumed to be identified as +1 or -1, the decision boundary is then at *y=0*, this is: $y = \sum_{i=1}^{m} x_i w_i + b = x_i w + b = 0$. The difference between lines boundaries is given by:

$$M = \frac{(1-b)}{\|w\|} - \frac{(-1-b)}{\|w\|} = \frac{2}{\|w\|} \tag{3}$$

So the maximum margin to separate two linear classes is obtaining by minimizing the norm of weights $\|w\|$.

Adaptive Neuro-Fuzzy Inference System (ANFIS): It is a class of adaptive networks that is functionally equivalent to fuzzy interface systems [Jahnkhani et al. 2008]. For classification and data analysis, it is assume the fuzzy inference system under consideration has two inputs *x* and *y*, an output *z*, and the rule base contains rules if-then, the Sugeno type are where the classification criteria is established, its architecture is formed by five layers (see Fig. 1). The inputs are the data to be classified and represent the grade for the fuzzy memberships. At the output, an internal layer is associated within the fuzzy region specified by the fuzzy rules [López-Arce et al. 2013]. Layer five O_i^5 then, sums all incoming signals as

$$O_i^5 = \sum_i w_i \times f_i = \frac{\sum_i w_i \times f_i}{\sum_i w_i} \tag{4}$$

where w_i are the weights of the incoming layers, and f_i are the function output fuzzy rules, this is the ANFIS output. The output is corrected with the training, where the parameters in the consequence of the rule are adapted to fit with the data classification criteria established previously, in this sense, it is possible to evaluate the ANFIS performance in different training epochs. The learning algorithm is selected as a hybrid method that uses back propagation for parameter associated with an input membership function and a least square estimation for parameters associated with output membership.

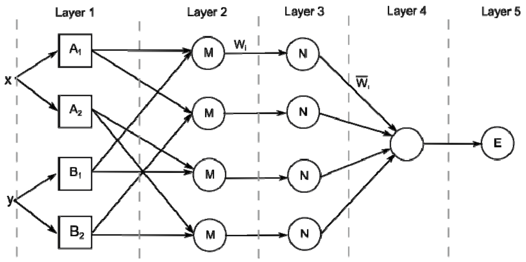


Fig. 1 ANFIS model structure used in the presented BMI: two inputs, one output, and five layers [López-Arce et al. 2013]

2 Methodology

We have reported in [López-Arce et al. 2013], a Brain-Machine Interface (BMI) based on EEG in order to detect two events: closed eyes and open eyes. After a comparison process of feature extraction methods [López-Arce et al. 2012] the wavelet transform method was implemented. This fundamental structure: EEG acquisition joint with Wavelet transform is the base to evaluate four classification methods: NB, LDA, SVM and ANFIS classifier. The objective is to keep the fairest conditions during the test, delivering the same feature vector for classifier input.

2.1 Experimental Paradigm

The experiment was carried out to a total of 7 male subjects in the age group of 21 to 53 years (28.3 ± 11.3) with no prior history of nervous system disorder. The experimental paradigm used to determine if the subjects had their eyes closed or open is the same as the one presented in [López-Arce et al. 2013]. The EEG electrodes were placed on the scalp according to the international 10-20 system considering the bipolar connection O1-P3. The register was carried out with a three channel EEG- acquisition device manufactured by the authors [López-Arce et al. 2012] The device has a Common-Mode Rejection Ratio (CMRR) of 100dB, and variable gain schedule from 50 up to 5000, the digitalization sampling frequency is 230 Hz. It is worth to comment that it exists a delay $\mathbf{t_d}$ between the occurrence of the event eyes closed and its perturbation in the alpha band, $\mathbf{t_d}$ must be characterized for each subject under test, it is commonly of around 600ms [López-Arce et al. 2012], this effect is illustrated in Fig 2, and was considered for the results included in this chapter.

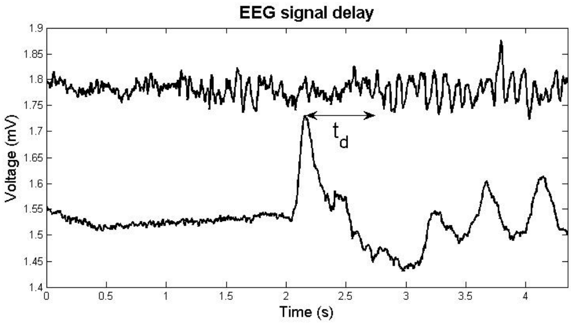


Fig. 2 Delay in the EEG signal changes after subject closes its eyes. Above signal is the EEG register, and below signal is the electrooculographic signal

2.2 Feature Extraction

The proposed algorithm is based on DWT and enables the BMI to process the data on-line, while recording the EEG signal [López-Arce et al. 2013]. The decomposition was performed in five levels, obtaining the wavelets decomposition tree shown in Fig. 3. *Daubichies* wavelet was selected as mother wavelet. Defining the normalized detailed coefficient at level 4 as cD_4, the BMI uses the mean and the maximum value of the last two calculated cD_4 wavelet components as the inputs of the four classifiers due to the fact such coefficient cD_4 represents the alpha waves with a bandwidth of 7.19~14.4Hz. This band is used to identify eyes close in a conscious person [Becedas 2012].

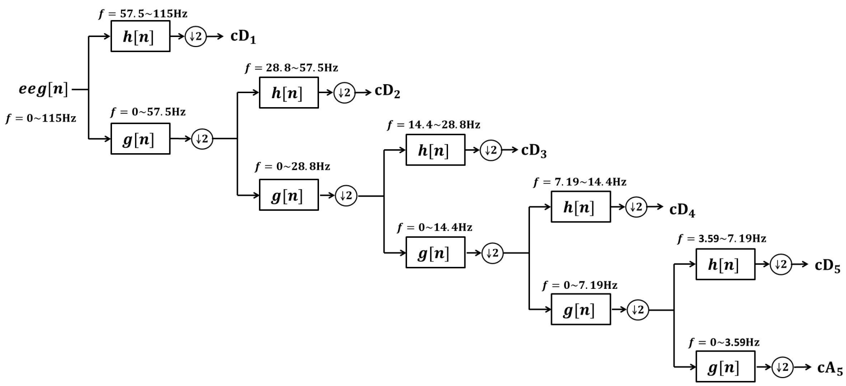


Fig. 3 Wavelet decomposition tree over 5 levels: 230 samples per second

2.3 *Classification Stage*

It is possible to consider a general structure for any of the proposed classifiers, as shown in Fig. 4.The input vector **X**, **X= [x_1 x_2]** with its components: **x_1** is the mean of the last two calculated ***cD_4*** and **x_2** is the maximum value of the last two calculated ***cD_4*.** The output vector **y** is the associated class, i.e. **y= C_1** if eyes were closed and **y= C_2** if eyes were open.

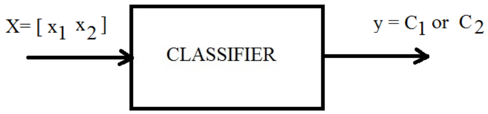


Fig. 4 General structure for any of the proposed classifiers, the input vector **X** has two components as explained in the text and the output vector **y** can take two possible values

All proposed classification methods are supervised type, for that reason each input vector **X** is previously associated with the corresponding output, obtaining a set of vectors: **data=[x_1 x_2 y]. A** total of 2029 vectors were considered for the comparison of the classifiers, they were divided as follows: 1771 for training (60%), 609 for testing (30%) and 203 for validation (10%). The validation data were acquired during a 14.5 seconds on-line session in which 45.32% of the total of validation data corresponds to eyes closed and 54.68% to eyes open. The same training, testing, and validation data were used for the four classifiers. Classifiers were implemented by the LDA toolbox, SVM toolbox and the ANFIS Toolbox of MATLAB™.

3 Results and Discussion

In this section the comparison between the performances of the four classifiers is presented, we based our discussion on the graph of boundaries decision and the contingency tables.

3.1 *Naive Bayesian Classifier*

During the NB adjustment, the training data and the testing data were considered obtaining a classification accuracy of 93.20% and 92.45%, respectively. The tuned NB classified the validation data correctly in a percentage of **93.59%**. It is important to remark that the output of the NBC classifier has always a probability associated with its decision making. The average probability of the training, testing and validation data associated with the classification accuracy of its decision making is shown in Table 1.

Table 1 NB probability associated with its decision making

Data Set	Classification Accuracy (%)	Average Probability (%)	
		Eyes closed	Eyes open
Training	93.20	96.29	96.32
Testing	92.45	93.35	96.69
Validation	93.59	91.03	93.80

As it can be appreciated, even if the output of the NB has classification accuracy higher than **90%** for all data sets, its average probability is not equal to 1 for any of them. This means that there exists always the probability that the classifier decision is incorrect and belongs to the opposite class. According with Table 1, it is more probably a misclassification occurs for the case of eyes closed.

3.2 Linear Discriminant Analysis

During the calculation of the vector of weights $\mathbf{w_i}$ the training data were correctly classified with an accuracy of 93.11% and the testing data with 90.97%. The final evaluation with the validation data set once the LDA is tuned was **90.15%** of accuracy. Opposite to NB there is not a probability associated with this result.

3.3 Support Vector Machine

After obtaining the SVM structure (the base of support vectors), the training data and the testing data were classified obtaining an accuracy of 92.54% and 89.49%, respectively. The validation data set was obtained on-line, and under this condition the classification accuracy was **90.64%**.

3.4 Adaptive Neuro-Fuzzy Inference System

The training of the ANFIS was carried out considering an increment of 25 epochs at a time, using the training data set. After each increment the testing data set is classified. This procedure, allowed us to get some control regarding the overtraining, whenever the classification accuracy begins to decrease. The results obtained in training the classifier from 0 to 250 epochs is shown in Table 2.

It is possible to observe that even if the ANFIS performance increase with the training data set, it does not apply to affirm a generalization result. That means that the classifier becomes too specific for the training data when the training epochs are excessive, which is consider one of the main problems of this type of classifiers [Chang and Yu 2005; Semmlow 2008]. Therefore, it does not have better classification accuracy for a different set of data, such as the testing data. In this case, the highest ANFIS classification accuracy with the testing data was achieved with 50 training epochs. The classification accuracy was 90.97%. The online validation data, the trained ANFIS presented an accuracy of **88.17%**.

Table 2 ANFIS performance evaluation

Training epochs	Accuracy (%)	
	Training data	Testing data
25	93.32	89.98
50	**93.89**	**90.97**
75	93.07	90.15
…	…	…
225	93.16	90.64
250	93.32	90.31

3.5 Classifiers Comparison

In Fig. 5 are shown the calculated classification boundaries for each algorithm. The area where the eyes are classified as *eyes closed* or *eyes open* is indicated. It is possible to affirm the simplicity of the LDA and its variant the SVM compare to NB and the highly nonlinear ANFIS structure to assign each class. This situation leads to a deficient generalization for NB and ANFIS. In Fig. 5, it is also shown the common set of data that was misclassified for all of the classifiers, taking into account exclusively the validation data. The data marked as eyes open (*) means that the output of the classifiers was *eyes open*; however, the actual state of the subject eyes was *eyes closed*. On the other hand, the data marked as eyes closed (o) indicate when the classifiers' output was *eyes closed* and the actual state of the subject eyes was *eyes open*.

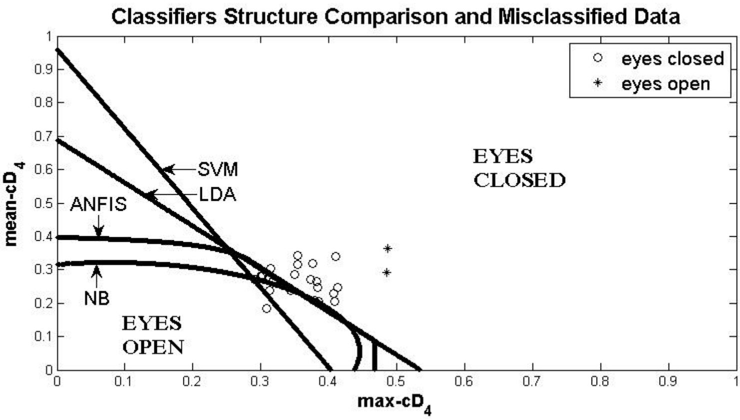


Fig. 5 Comparison of boundaries decision: marks in the graph represent the common misclassified data

It is worth to notice that most of the misclassified data are near of the common boundary of the classifiers and they mostly belong to the *eyes open* class, but they are misclassified as *eyes closed*. However, Fig. 5 does not show if this misclassification pattern is the same for every classifier. Therefore, the contingency tables were computed for each classifier and they are presented in Table 3. This table shows the percentage obtained by comparing the actual state of the subjects' eyes and the output of the classifiers for the validation data. The validation data were distributed as **45.32%** of the cases *eyes open,* and **54.68%** of the cases *eyes closed,* in a total sample of **203 cases**. In Table 3, it can be appreciated that general accuracy is in fact the summation of True Positives (*eyes closed* classified as *eyes closed*) and True Negatives (*eyes open* classified as *eyes open*) [Semmlow 2008]. ANFIS classifier has the worst general performance and lowest sensibility to detect the condition of *eyes closed*, and the higher specificity to detect *eyes open*; however, *eyes closed* is in fact the most important feature to be detected in our BMI. NB classifier shows the best accuracy and an acceptable sensitivity and specificity, but it is important to keep in mind that these results are associated to a probability factor (see Table 1) and these percentages must be weighted consequently. On the other hand, LDA and SVM show a similar performance and particularly SVM presents the best sensitivity to detect *eyes closed* (**44.83%** of **45.32%**). Additionally SVM has one of the lowest machine complexity, these reasons allow us to confirm that SVM is the best choice for the BMI presented.

We do not dismiss the fact that it has been assumed that *eyes closed* and *eyes open* are linearly separable classes, which is not completely true. Although there are techniques to transform this nonlinearly separable problem into a linearly separable [Semmlow 2008], our motivation is to keep the BMI structure as simple as possible.

Table 3 Contingency tables for classifiers, data are presented in relative percentages to de total of cases distributed in validation data set

NBC Contingency Table

Classifier Output	*Actual State*: Eyes closed	*Actual State*: Eyes open	
Eyes closed	44.33	5.419	
Eyes open	0.9852	49.26	
Total	45.32	54.68	93.59

LDA Contingency Table

Classifier Output	*Actual State*: Eyes closed	*Actual State*: Eyes open	
Eyes closed	41.38	5.911	
Eyes open	3.941	48.77	
Total	45.32	54.68	90.15

SVM Contingency Table

Classifier Output	*Actual State*: Eyes closed	*Actual State*: Eyes open	
Eyes closed	44.83	8.867	
Eyes open	0.4926	45.81	
Total	45.32	54.68	90.64

ANFIS Contingency Table

Classifier Output	*Actual State*: Eyes closed	*Actual State*: Eyes open	
Eyes closed	34.48	0.9852	
Eyes open	10.84	53.69	
Total	45.32	54.68	88.17

4 Conclusions

The analysis and comparison of four classification methods for a simplified BMI from EEG signals in order to determine if the subject's eyes were open or close is presented. The best classification accuracy of the validation data was obtained by the NB classifier (93.59%) and the worst classification accuracy by the ANFIS classifier (88.17%). However, using a contingency table it was shown that the SVM obtained the best performance to classify eyes closed, which was consider the most important feature to classify in the BMI application used.

References

[Becedas 2012] Becedas, J.: Brain–machine interfaces: basis and advances. IEEE Trans. Syst., Man, Cyber., Syst., Part C: Applications and Reviews 42(6), 825–836 (2012)

[Bi et al. 2013] Bi, L., Fan, X.A., Liu, Y.: EEG-based brain-controlled mobile robots: a survey. IEEE Trans. Human–Mach. Syst. 43(2), 161–176 (2013)

[Chang and Yu 2005] Chang, S., Yu, H.: Novel probability neural network. IEEE Proceedings – Vision, Image and Signal Processing 152(5), 137–144 (2005)

[Gandhi et al. 2011] Gandhi, T., Panigrahi, B.K., Anand, S.: A comparative study of wavelet families for eeg signal classification. Neurocomputing 74(17), 3051–3057 (2011)

[Hsu 2012] Hsu, W.Y.: Neuro-fuzzy prediction for brain-computer interface applications. In: Fuzzy Inference System-Theory and Applications, pp. 299–312 (2012)

[Jahnkhani et al. 2008] Jahankhani, P., Revett, K., Kodogiannis, V.: A rule based approach to classification of eeg datasets: a comparison between anfis and rough sets. In: 9th Symposium on Neural Network Applications in Electrical Engineering, pp. 157–160 (2008)

[Jain et. al. 2000] Jain, A.K., Duin, R.P.W., Mao, J.: Statistical pattern recognition. IEEE Trans. Pattern Anal. Mach. Intell. 22(1), 4–37 (2000)

[López-Arce et al. 2012] Lopez-Arce Vivas, E., Garcia-Gonzalez, A., Galindo, F., Sanchez-Gonzalez, V.: Algorithm to detect six basic commands by the analysis of electroencephalographic and electrooculographic signals. In: Proc. IEEE Conference on Biomedical Robotics and Biomechatronics, Rome, Italy, pp. 189–194 (2012)

[López-Arce et al. 2013] López-Arce, E., García-González, A., Figueroa, I., Fuentes, R.: Discrete wavelet transform and anfis classifier for brain-machine interface based on EEG. In: Proc. IEEE Conference on Human System Interaction, Sopot, Poland, pp. 137–144 (2013)

[Majumdar 2011] Majumdar, K.: Human scalp eeg processing: various soft computing approaches. Applied Soft Computing 11(8), 4433–4447 (2011)

[Nicholas-Alonso and Gómez-Gil. 2012] Nicholas-Alonso, L.F., Gómez-Gil, J.: Brain computer interfaces, a review. Sensors 12, 1211–1279 (2012)

[Semmlow 2008] Semmlow, J.: Biosignal and medical image processing. CRC Press, EE. UU (2008)

Some Approaches to Recognition of Sign Language Dynamic Expressions with Kinect

M. Oszust and M. Wysocki

Department of Computer and Control Engineering,
Rzeszow University of Technology, W. Pola 2, 35-959 Rzeszów, Poland
{marosz,mwysocki}@kia.prz.edu.pl

Abstract. The paper considers recognition of isolated Polish Sign Language words observed by Kinect. A whole word model approach with nearest neighbour classifier applying dynamic time warping (DTW) technique is compared with an approach using models of subunits, i.e. some elements smaller than words, resembling phonemes in spoken expressions. Such smaller models are obtained using data-driven procedure involving division of time series representing words from a training set into subsequences which form homogeneous groups. Symbols are assigned to these groups and then gestures receive symbolic representations (transcriptions). Such transcriptions are classified using nearest neighbour approach based on edit distance. Two sets of features have been used: one based on Kinect's skeletal images and the other using simplified description of hands extracted as skin coloured regions. Ten-fold cross-validation tests of classifiers using data representing signed Polish words were performed. Subunit based approach proved to be superior, in particular in the case when only one learning example was available. Features taking into account description of hands led to better recognition results in comparison with those based on skeleton.

1 Introduction

Aiding hearing impaired people with automatic translation tools not only will improve quality of their daily communication, but it also fits in and benefits the area of natural human-computer interfaces. Since in sign language recognition hands observation plays important role, computer vision seems to be attractive and promising approach. In order to facilitate hand segmentation from the background coloured hand gloves, markers [Awad et al. 2009] or detection of skin coloured objects [Ong and Ranganath 2005] are applied. Then hands are tracked and values of attributes describing gestures are computed. Vision based, colour segmentation algorithms are constrained by lighting conditions, and therefore many systems impose some restrictions for the background or clothing of the person performing gestures. Active depth cameras are insensitive to some of such limitations providing 3D information of observed scene, but most of them do not provide colour information making hand modelling and tracking very demanding. Kinect sensor

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_7, © Springer International Publishing Switzerland 2014

[Shotton et al. 2011] offers a colour and depth camera images. Furthermore, its build-in software can interpret coarse body gestures and produces a skeletal image, which consists of 20 connected 3D points in the shape of the body.

Most published works with Kinect consider simple (static or dynamic) hand gestures or body movements with good recognition results. For example in [Ren et al. 2011] both the depth and colour information from Kinect sensor to recognise 14 static hand shapes useful for arithmetic operations and 3 shapes for *Rock-paper-scissors* game were used. Static hand shapes representing American Sign Language digits were recognised in [Keskin et al. 2011] with 99% recognition rate. Dynamic sign language gestures were recognised on the basis of skeletal image [Lang et al. 2012] or using skin colour based hand segmentation [Ong et al. 2012].

Gestures are mostly classified using Hidden Markov Models (HMMs) [Kraiss 2006], neural networks [Zahedi and Manashty 2011] or applying the nearest neighbour approach with DTW [Oszust and Wysocki 2012]. These solutions use whole word models approach, i.e. one word model represents one sign. In order to provide better performance, especially with large sign vocabulary, signs are modelled using smaller units than words (called *subunits*), which resembles modelling speech by means of phonemes. Subunit models are concatenated to form sign models [Agris et al. 2008]. In [Kraiss 2006] HMMs and an iterative process of data-driven extraction of subunits were applied. Two state HMMs representing subunits were concatenated to model single signs. The boundaries of subunits result from the alignment of appropriate feature vector sequence to the states by the Viterbi algorithm. Different approach can be found in [Han et al. 2009]. Here authors define subunits' boundaries using hand motion discontinuity and adapting temporal clustering by DTW to merge similar time series segments and finally selecting a representative subunit from each cluster. Another direction is designing a new, subunit based classifier choosing features responsible for discrimination of a given gesture class [Ong et al. 2012].

The main contribution of our paper is that it presents experimental comparison of two approaches to recognition: one using whole word model and the other based on subunits obtained by our data-driven procedure involving division of time series representing words from a training set into subsequences which form homogeneous groups. Both methods are applied for two different sets of features: one based on Kinect's skeletal images and the other using a simplified description of hands extracted as skin coloured regions. All experiments are performed on real Polish sign language (PSL) words data.

The rest of the paper is organised as follows. Section 2 gives preliminary information concerning PSL and our gesture dataset. The Kinect sensor description and characterisation of features used for recognition are shown in Section 3.

Section 4 is focused on clustering of time series representing the gestures to show whether automatic procedure is able to reveal the natural division of the data. Section 5 presents a data-driven subunits extraction method. Recognition results are shown and compared in Section 6. Section 7 concludes the paper.

2 Gesture Corpus

Sign language is the language of deaf people communicated by their deaf parents. PSL signs are static or dynamic and mostly two-handed [Hendzel 1986]. During gesture performance hands often touch each other or appear against the background of the face. In cooperation with PSL interpreter we selected 30 common words that can be used at the doctors'. This topic is particularly important due to need of ensuring patients' privacy. Chosen words are as follows: *obtain*, *write*, *bed*, *must*, *to rest*, *operation, meninges*, *analysis*, *examine*, *be*, *want*, *cotton*, *tooth*, *inflammation*, *healthy*, *ill*, *lie down*, *translucence*, *come*, *rescue*, *family*, *hearing*, *tablet*, *where*, *and*, *how much*, *other*, *cost*, *blood*, and *drop*. Each word was performed 10 times by a PSL interpreter. The data have been registered with the rate of 30 frames/s using Kinect TCP software [Kinecttcp 2011].

3 Features Used for Recognition

Kinect sensor offers a colour camera image and the depth video stream with the resolution of 640x480. A developer also gets a skeletal tracking - skeletal images of two tracked people consisting of a collection of 20 3D points in the joints of arms and legs, the position of the head, hands and spine. For gesture recognition we used 6 points representing right and left hand, wrist and elbow (in total 18 features).

Results of classification based on skeletal images have been compared with classification using description of hands detected as skin coloured regions. In order to determine whether a pixel is in the skin colour red, green and blue components in RGB colour space are analysed [Vezhnevets et al. 2003]. For demanding lightning conditions one can use more robust solution with a 2D Gaussian distribution of skin colour pixels in the normalised RGB space [Oszust and Wysocki 2012].

The segmentation algorithm searches for sets of contiguous pixels in the skin colour forming objects taking into consideration the depth image. Adjacent pixels are assigned to the same object only if the depth difference between them is smaller than a threshold. In the next step hands and face are identified using their known positions in previous frame and then hand features are calculated. Feature selection was motivated by linguistic directions, i.e. every sign language gesture can be analysed by specifying at least three components [Hendzel 1986]: (i) the place of the body against which the sign is made, (ii) the shape of a hand or hands, (iii) the movement of a hand or hands. For each hand we used the following set of 7 features: the gravity centre with respect to the gravity centre of the face and depth difference between mean depth of the hand and the face (as 3D information about hand placement), as well as: area, compactness, eccentricity, and orientation of the hand (as four features representing hand shape description). Feature values were normalised to the range from 0 to 1. Time series of the horizontal placement of the gravity centre of the right hand with respect to the gravity centre of the face for executions of the word *analysis* are shown on Fig. 1.

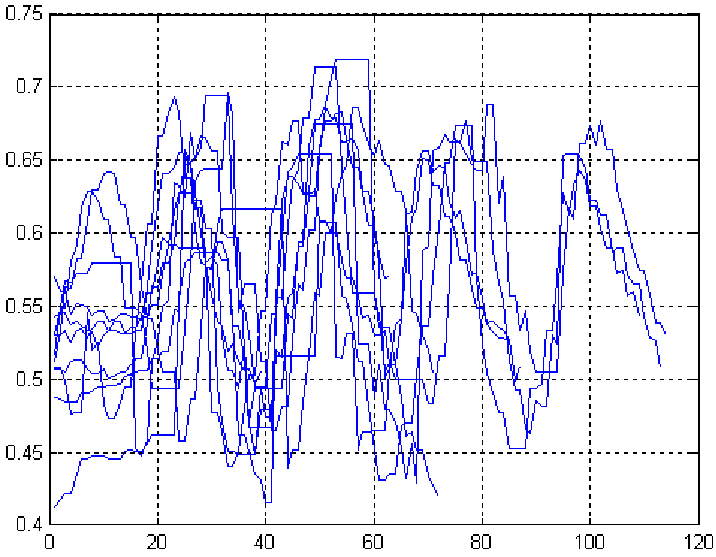
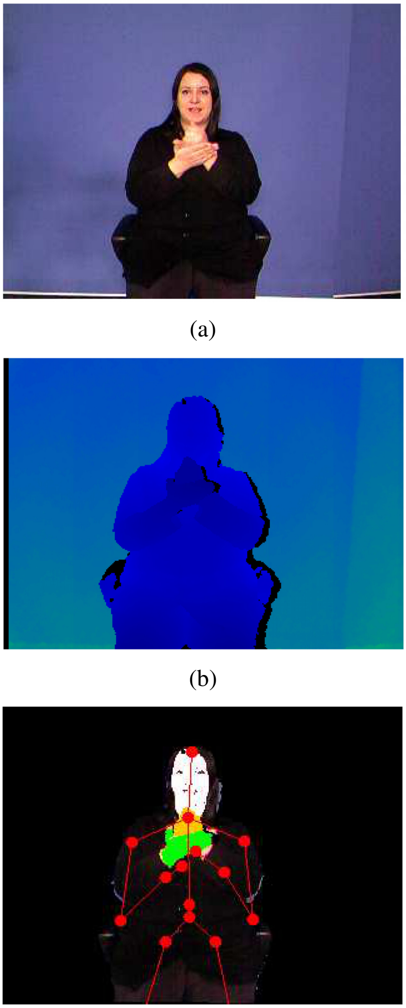


Fig. 1 Ten executions of word *analysis*: the horizontal placement of the gravity centre of the right hand with respect to the gravity centre of the face

Fig. 2 presents sample images obtained from the sensor: the colour image, the depth image and the processed image showing skeleton and hands identified with help of depth information.



(a)

(b)

(c)

Fig. 2 Sample images: from the colour camera (a), from the depth camera (b), after processing (c) with the face coloured white, the right hand – green, the left hand – orange, skeleton – red

4 Gesture Data Analysis

At first we present experiments concerning clustering of data based on skin colour approach in order to show whether automatic procedure is able to reveal the natural division of the data. We consider *hubness* phenomenon [Marussy and Buza 2013; Radovanovic et al. 2010] to find which data samples tend to be representatives of other classes and likely to be the cause of recognition problems.

In the clustering vectors are put into groups based of their resemblance [Theodoridis and Koutroumbas 2008; Xu and Wunsch 2009]. In general, there is no available information about how vectors should be grouped and how many groups should be formed. In our case we know that gestures belong to 30 classes and we want to examine how well gesture data is separated.

Time series representing PSL gesture is a vector containing chronologically ordered observations [Fu 2011], and it has to be considered as a whole. In order to cluster such time series we have to take into account their continuous character and chose appropriate technique for their comparison. Most clustering algorithms calculate similarity between vectors assuming equal dimensions of data being clustered. In our approach we created a proximity matrix containing pairwise proximities (distances) between time series representing gestures (see matrix W in (1) for n time series X). To compute distances between time series we applied DTW [Xu and Wunsch 2009] due to its ability to expand or compress the time when comparing sequences that are similar but locally out of phase.

$$W = \begin{bmatrix} d_{DTW}(X_1, X_1) & \cdots & d_{DTW}(X_1, X_n) \\ d_{DTW}(X_2, X_1) & \cdots & d_{DTW}(X_2, X_n) \\ \vdots & \ddots & \vdots \\ d_{DTW}(X_n, X_1) & \cdots & d_{DTW}(X_n, X_n) \end{bmatrix} \tag{1}$$

In our experiments we used K-means clustering [Xu and Wunsch 2009]. This technique starts from an initial assignment of vectors into a given number of clusters and iteratively improves the assignments calculating a mean vector at each iteration. If no further changes in assignments appear, the algorithm stops.

To assess quality of clustering we applied RAND index [Maulik and Bandyopadhyay 2002], which measures the number of pairwise agreements between two decompositions D_1 and D_2. This index is useful when we know how data should be clustered, say decomposition D_1. It is calculated as follows: $RAND(D_1, D_2) = (a+b)/(a+b+c+d)$, where a denotes the number of pairs of elements that are in the same cluster in both decompositions, b denotes the number of pairs of elements that are in different clusters in both decompositions, c denotes the number of pairs of elements that are in the same cluster in the first decomposition D_1, but in different clusters in D_2, and d denotes the number of pairs of elements that are in different clusters in D_1 but in the same cluster in D_2. RAND values are considered better if they are closer to 1.

Because K-means produces locally optimal results we have run it 1000 times and obtained mean RAND value equal 0.9948 with standard deviation equal 0.0006. The number of agreements ($a+b$) turned out to be much larger than the number of disagreements ($c+d$).

Hubness is a phenomenon, in which *hubs* are defined as data points that appear unusually often among the k-nearest neighbours of other data points [Marussy and Buza 2013; Radovanovic et al. 2010]. It seems that *bad hubs*, i.e. data points with most neighbours from other classes are seen as data examples which could be difficult to recognise due to resemblance to other data classes. In experiments we used W

matrix and prepared lists of *k*-nearest neighbours of each gesture realisation. Then we found realisations which were often chosen as neighbours of realisations out of different classes. We changed *k* in the range from 1 to 9 and found that some realisations of words *drop* and *cost* were found on many lists of gestures from different classes. It means that this data samples could be wrongly recognised.

Analysing clustering results we observed that some PSL words were performed without preservation of ideal representation of the gesture. Therefore some realisations of words: *must*, *examine*, *lay down*, *family* or *drop* were assigned to other word clusters. Among these realisations we found *bad hubs*, additionally confirming that it is worth paying attention to them in case of recognition problems.

Gesture data analysis revealed that there are some words' realisations, which could be misclassified using the nearest neighbour approach. Data itself also contain many divisions not connected with known assignment of gesture realisations. It can make the classification task more difficult.

5 Data-Driven Subunits

The first problem that has to be solved is how to break down time series of feature vectors into subunits. Due to the lack of linguistic knowledge and non-existence of sign language video corpus with annotation on subunit level for Polish Sign Language we applied a data-driven procedure to find subunits. Time series of a given feature which belong to all learning gesture examples are divided into subsequences constrained by gestures' lengths and minimal and maximal length of subsequences. Subsequences do not overlap and each of them begins at the beginning of a gesture or at a time sampling point where previous subsequence ends. Sign gestures are both sequential and simultaneous [Kraiss 2006]. Sequentiality indicates that the order of shown hand shapes and places of execution is important. Simultaneity means that during gesture performance the features (e.g. hand position and shape) can change in parallel, although not necessarily in synchrony. Hence, we processed each feature (see section 3) independently.

In the next step subsequences are clustered using K-means algorithm [Xu and Wunsch 2009]. The algorithm requires operations on elements of equal length. Therefore, we represented each subsequence by mean and standard deviation of its elements.

Finally the decision problem consisted in choosing a number of clusters and determining cut points of time series representing gestures to obtain clusters of good quality. After some experiments we decided to set the number of clusters equal to 10. Quality of clusters was characterised by the mean distance among subsequences and their cluster centres. To optimise this criterion we adapted the immune algorithm CLONALG, often used in a wide variety of optimisation tasks [De Castro and Von Zuben 2002]. The main loop of the CLONALG algorithm [Trojanowski and Wierzchon 2009], repeated predefined number of generations, consists of evaluation of antibodies, clonal selection, hipermutation and apoptosis. The antibody represents a solution of the time series partitioning problem, i.e. it contains cut points. In the clonal selection step the algorithm selects a subset of best ranked antibodies and hipermutates them. The hipermutation means a random operation changing antibodies. During the apoptosis worst antibodies are replaced by randomly generated new antibodies.

After obtaining clusters of similar subsequences a symbol is assigned to each cluster and gesture representations are transformed from time series to sequences of symbols. For gestures from the set used to determine subunits (training set) such a transcription is obtained automatically. For a gesture to be recognised appropriate partitioning of the related time series into elements which are closest (in the DTW sense) to known subsequences has to be determined first. Then the recognition is performed by nearest neighbour using *edit distance* [Xu and Wunsch 2009] as a measure of the difference between two transcriptions. The method resembles DTW. It uses dynamic programming to find a minimum number of operations (insert, delete, replace) required to transform one string of characters into the other. Our classifier took a decision based on sum of edit distances calculated for all features characterising the gesture (see section 3), i.e. for all strings of symbols obtained from appropriate time series.

6 Recognition Results

To compare the described approaches to recognition of PSL words we performed ten-fold cross-validation tests using the nearest neighbour classifier. Respective data

Table 1 Recognition rates (in %) in cross-validation tests of the nearest neighbour classifier

Skin colour based features					Kinect's skeleton			
Approach	Whole word		Subunits		Whole word		Subunits	
Test\Variant	A	B	A	B	A	B	A	B
1	100	73.3	95.1	77.7	90.0	54.4	81.6	54.3
2	100	69.3	97.3	83.8	86.7	60.7	89.7	67.2
3	100	61.8	96.6	80.1	83.3	48.1	83.5	63.0
4	100	73.3	100	89.4	93.3	61.8	96.4	70.0
5	100	76.3	99.9	90.7	96.7	64.4	96.9	75.0
6	93.3	70.0	100	90.4	93.3	58.1	95.2	76.8
7	100	76.3	99.9	92.6	93.3	65.2	93.2	79.7
8	96.7	73.7	99.7	90.2	90.0	70.0	93.2	78.7
9	100	70.7	99.9	91.4	86.7	60.4	92.0	74.7
10	93.3	66.67	99.9	90.4	80.0	51.1	97.0	73.7
Mean	98.3	71.1	98.8	87.7	89.3	59.4	91.9	71.3
Stdev.	2.7	4.2	1.8	5.2	4.9	6.3	5.5	7.9
Minimum	93.3	61.8	95.1	77.7	80.0	48.1	81.6	54.3
Maximum	100	76.3	100	92.6	96.7	70.0	97.0	79.7

sets were divided into ten separable subsets, each subset with data representing one realization of each of the 30 words. Two variants were considered. In the first variant (A) nine subsets were used as training set and the remaining, tenth subset as the test set. In the second variant (B) the training and the test set were swapped to show how the classifier behaves when only one learning example is available. Whole word and subunit based recognition results for two groups of features are compared in Table 1. Subunit based classifier was run 100 times due to stochastic nature of optimisation. Hence the mean values of recognition rate are given. In CLONALG we used 100 generations of 100 antibodies.

Results obtained by subunit based classifier are generally better, especially in variant B, when only one learning example is available (17% better for skin colour based features, 12% better for Kinect's skeleton features). The results confirm usability of subunit based approach, especially when the training set is small.

Worse recognition results with Kinect's skeleton were caused by hand tracking errors [Obdrzalek et al. 2013] and inability to differentiate hand shapes using one 3D point as hand representation.

More detailed analysis showed that some miss-classified word's realisations were also indicated during dataset evaluation by clustering which additionally proved usability of such prior examination.

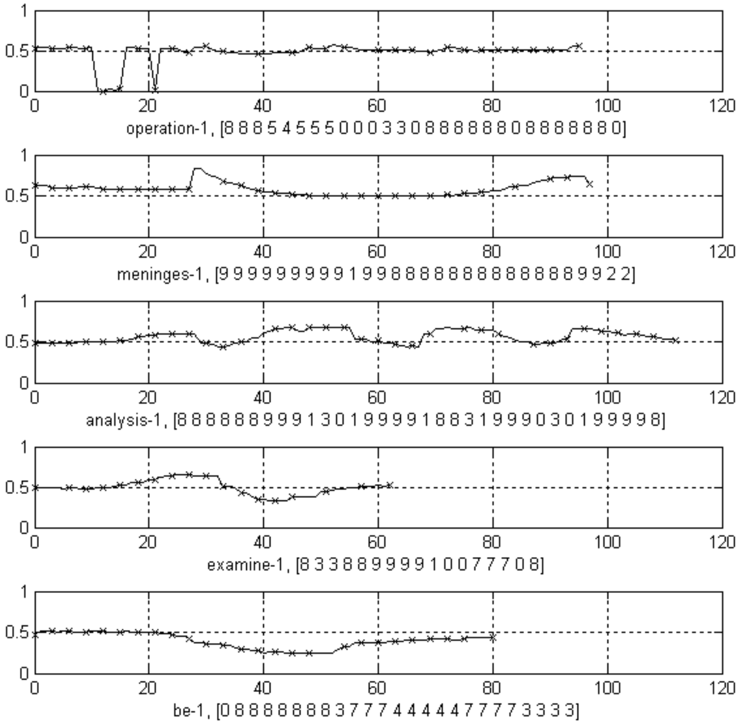


Fig. 3 Time series of the horizontal placement of the right hand centre representing 5 exemplary words realisations from the training set. Subunit's borders are marked with crosses; words' transcriptions are given in brackets

Exemplary subsequences, obtained for the horizontal placement of the right hand centre, extracted from training set, with related symbolic transcriptions are shown in Fig. 3.

7 Conclusions and Future Works

The paper presents comparison of classification of Polish sign language words based on whole word models with classification using units smaller than words. The second approach proved to be superior, in particular in the case where only one learning example was available. Two sets of features obtained with Kinect sensor were considered: (i) features based on Kinect's skeletal image and (ii) features describing hands as skin coloured regions. In the second case Kinect's ability to extract person body from the background and to distinguish partially occluded, skin coloured objects on the basis of depth information has been used. Due to better hand description proposed features led to better recognition results. Obtained feature vectors representing PSL words were clustered and bad hubs were found which provided additional information about processed data.

Future works will concern experiments with larger vocabulary including words and sentences.

References

[Agris et al. 2008] Agris, U., Zieren, J., Canzler, U., et al.: Recent developments in visual sign language recognition. Universal Access in the Information Society 6(4), 323–362 (2008)

[Awad et al. 2009] Awad, G., Han, J., Sutherland, A.: Novel boosting framework for subunit-based sign language recognition. In: Proc. IEEE International Conference on Image Processing, Piscataway, NJ, pp. 2693–2696 (2009)

[Cooper 2010] Cooper, H.: Sign language recognition: generalising to more complex corpora. Centre for Vision Speech and Signal Processing, PhD thesis, University of Surrey (2010)

[De Castro and Von Zuben 2002] De Castro, L.N., Von Zuben, F.J.: Learning and optimization using the clonal selection principle. IEEE Transaction on Evolutionary Computation 6, 239–251 (2002)

[Fu 2011] Fu, T.C.: A review on time series data mining. Engineering Applications of Artificial Intelligence 24(1), 164–181 (2011)

[Han et al. 2009] Han, J., Awad, G., Sutherland, A.: Modelling and segmenting subunits for sign language recognition based on hand motion analysis. Pattern Recognition Letters 30(6), 623–633 (2009)

[Hendzel 1986] Hendzel, J.: Polish sign language dictionary. Publishing House Pojezierze (1986) (in Polish)

[Keskin et al. 2011] Keskin, C., Kirac, F., Kara, Y.E., et al.: Real time hand pose estimation using depth sen-sors. In: Proc IEEE ICCV Workshops, pp. 1228–1234 (2011)

[Kinecttcp 2011] KinectTCP, Temple University (2011), `https://sites.google.com/a/temple.edu/kinecttcp/` (accessed September 2, 2013)

[Kraiss 2006] Kraiss, K.F.: Advanced man-machine interaction. Springer, Berlin (2006)

[Lang et al. 2012] Lang, S., Block, M., Rojas, R.: Sign language recognition using kinect. In: Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2012, Part I. LNCS, vol. 7267, pp. 394–402. Springer, Heidelberg (2012)

[Maulik and Bandyopadhyay 2002] Maulik, U., Bandyopadhyay, S.: Performance evaluation of some clustering algorithms and validity indices. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(12), 1650–1654 (2002)

[Marussy and Buza 2013] Marussy, K., Buza, K.: Hubness-based indicators for semi-supervised time-series clas-sification. In: Proc. 8th Japanese-Hungarian Symposium on Discrete Mathematics and Its Applications 2013, Veszprem, Hungary, pp. 97–108 (2013)

[Obdrzalek et al. 2013] Obdrzalek, S., Kurillo, G., Ofli, F., et al.: Accuracy and robustness of kinect pose estima-tion in the context of coaching of elderly population. In: Proc IEEE Annual International Conference on Engineering in Medicine and Biology Society, pp. 1188–1193 (2012)

[Ong et al. 2012] Ong, E.J., Cooper, H., Pugeault, N., et al.: Sign language recognition using sequential pattern trees. In: Proc. IEEE Computer Vision and Pattern Recognition, pp. 2200–2207 (2012)

[Ong and Ranganath 2005] Ong, S.C.W., Ranganath, S.: Automatic sign language analysis: a survey and the future beyond lexical meaning. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(6), 873–891 (2005)

[Oszust and Wysocki 2012] Oszust, M., Wysocki, M.: Modelling and recognition of signed expressions using subunits obtained by data–driven approach. In: Ramsay, A., Agre, G. (eds.) AIMSA 2012. LNCS (LNAI), vol. 7557, pp. 315–324. Springer, Heidelberg (2012)

[Radovanovic et al. 2010] Radovanovic, M., Nanopoulos, A., Ivanovic, M.: Hubs in space: popular nearest neighbors in high-dimensional data. Journal of Machine Learning Research 11, 2487–2531 (2010)

[Ren et al. 2011] Ren, Z., Meng, J., Yuan, J., et al.: Robust hand gesture recognition with kinect sensor. In: Proc. 19th International Conference on Multimedia, Scottsdale, AZ, USA (2011)

[Shotton et al. 2011] Shotton, J., Fitzgibbon, A., Cook, M., et al.: Real-time human pose recognition in parts from single depth images. In: Proc. IEEE Computer Vision and Pattern Recognition, pp. 1297–1304 (2011)

[Theodoridis and Koutroumbas 2008] Theodoridis, S., Koutroumbas, K.: Pattern Recognition, 4th edn. Academic Press (2008)

[Trojanowski and Wierzchon 2009] Trojanowski, K., Wierzchon, S.: Immune-based algorithms for dynamic optimization. Information Sciences 179, 1495–1515 (2009)

[Vezhnevets et al. 2003] Vezhnevets, V., Sazonov, V., Andreeva, A.: A survey on pixel-based skin color detection techniques. In: Proc. GraphiCon 2003, pp. 85–92 (2003)

[Vogler and Metaxas 1999] Vogler, C., Metaxas, D.: Toward scalability in ASL recognition: Breaking down signs into phonemes. In: Braffort, A., Gherbi, R., Gibet, S., Richardson, J., Teil, D. (eds.) GW 1999. LNCS (LNAI), vol. 1739, pp. 211–224. Springer, Heidelberg (2000)

[Xu and Wunsch 2009] Xu, R., Wunsch, D.: Clustering. Wiley-IEEE Press (2009)

[Zahedi and Manashty 2011] Zahedi, M., Manashty, A.R.: Robust sign language recognition system using ToF depth cameras. Computing Research Repository - CORR, abs/1105.0 (2011)

Multimodal Human-Computer Interfaces Based on Advanced Video and Audio Analysis

A. Czyżewski, P. Dalka, Ł. Kosikowski, B. Kunka, and P. Odya

Multimedia Systems Department, Gdansk University of Technology, Gdańsk, Poland
ac@pg.gda.pl, {dalken,kosiq,kuneck,
piotrod}@sound.eti.pg.gda.pl

Abstract. Multimodal interfaces development history is reviewed briefly in the introduction. Some applications of multimodal interfaces to education software for disabled people are presented. One of them, the LipMouse is a novel, vision-based human-computer interface that tracks user's lip movements and detect lips gestures. A new approach to diagnosing Parkinson's disease is also shown. The progression of the disease can be measured employing the UPDRS (Unified Parkinson Disease Rating Scale) scale which is used to evaluate motor and behavioral symptoms of the Parkinson's disease, based on the multimodal interface called Virtual-Touchpad (VTP) used for supporting medical diagnosis. The scent emitting multimodal computer interface provides an important supplement of the polysensoric stimulation process, playing an essential role in education and therapy of children with certain developmental disorders. The Smart Pen providing a tool for supporting therapy of developmental dyslexia is presented and results achieved with its application are discussed. The eye-gaze tracking system named Cyber Eye, developed at the Multimedia Systems Department employed to many kinds of experiments is presented including analysis of visual activity of patients remaining in vegetative state and their awareness evaluation. The paper is concluded with some general remarks concerning the role of multimodal computer interfaces applied to learning, therapy and everyday usage of computerized devices.

1 Introduction

Since the appearance of the first computers, people are constantly trying to find some better ways to communicate with the computers. To this day, using keyboard is the most popular text entry method while text is still read on a computer screen. But since the 70s' of the last century (i.e. from the moment the computer keyboard became popular) many other solutions have appeared on the market, only few with success. In the 90's, when computers enriched with a graphical user interface and screen resolutions began to rise, computer users had to learn how to use a computer mouse. Later, high expectations were set for speech communication with the computer. Unfortunately, it turned out that even today, comfortable bidirectional voice communication with machines remains the domain of science-fiction movies.

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_8,

Tablets showed that the mouse is no longer necessary to use a computer efficiently. Still, much remains to be done to help people to control a computer in a more natural manner, such as employing gestures, facial expressions or by gaze. It should be pointed out, that it is especially important for people with different form of disabilities. Nowadays, many researches are carried out to develop new forms of interaction with computers employing several modalities. Usually, these new forms are related nowadays with multimodal interfaces.

Multimodal interfaces are in the scope of research at Multimedia Systems Department from more than a decade. Gained experience in audio and video processing along with knowledge related to some medical applications allows for creating advanced solutions devoted especially to therapy and education applications. Some of these solutions are presented in the following sections of this paper. Special emphasis is put on interfaces created in the framework of the project entitled "Elaboration of a series of multimodal interfaces and their implementation to educational, medical, security and industrial applications".

2 LipMouse

LipMouse is the name of a novel, patent-pending, contactless, human-computer interface that allows a user to work on a computer using movements and gestures made with his/her mouth only [Dalka and Czyżewski 2010]. The target users for the tool are people who, for any reason, are not able or do not want to use traditional input devices. Therefore LipMouse is a solution enabling severely disabled and paralyzed people to use a computer and communicate in this way with the surrounding world. No user adaptation, such as placing marks on the face, is required in order to successfully work with LipMouse.

LipMouse is an application running on a standard PC. It requires only one additional hardware component: a display-mounted, standard web camera that captures images of the user face. The main task of LipMouse is to detect and to analyze images of user's mouth region in a video stream acquired from the web-camera. All movements of mouth (or head) are converted to movements of the screen cursor. LipMouse also detects three mouth gestures: opening the mouth, sticking out the tongue and forming puckered lips (as for kissing). Each gesture may be associated with an action, which in turn can be freely chosen by a user. Possible actions include clicking or double-clicking various mouse buttons, moving mouse wheel – both horizontally and vertically and others.

Fig. 1 presents the main window of the application. It allows a user to configure LipMouse according to his or her preferences. In the frames, the mouth region is marked with a rectangle, and the lip shape is marked with an ellipse.

Before a user starts working with LipMouse, a short calibration lasting about 30 seconds needs to be executed. During the calibration, the user is asked to perform some head movement and gestures according to the instructions seen on the screen. The purpose of the calibration is to tune LipMouse to detect gestures made by the actual user in the current lighting conditions.

Fig. 1 LipMouse application main window

The application uses advanced video processing algorithms. First, a user's face is detected in every image frame captured by a web camera. Further stages of the algorithm operation are restricted to the ROI containing the user's face. Then, the mouth region is localized using various color space transformations and its shift from the reference mouth position is calculated. This shift is directly used to drive a screen cursor. Simultaneously, a small region (blob) placed on user lips is found in mouth region. This blob is used as a starting condition for an iterative method for lip shape estimation with an ellipse. Lip shape and lip region image features are used by the intelligent decision system utilizing an artificial neural network to classify gestures made by a user.

A feed-forward ANN with one hidden layer is used to detect lip gestures. Each image frame is classified independently. The number of ANN inputs corresponds to the number of lip image features and is equal 168 or 171, depending on the chosen variant of lip region extracting [Dalka and Czyżewski 2009]. There are 4 outputs from ANN, related with types of gestures recognized by ANN. A type of the gesture is determined by the maximum value of the ANN outputs. However, in the HCI application it is crucial to minimize false-positive rate of detection of all three, intentional mouth gestures in order to prevent execution of actions not meant by a user. The false-negative rate is less important – if a gesture is not recognized in some frame, it will be most probably recognized in succeeding frames when a user moves his head a little or change their face expression.

In order to minimize number of false-positives, a post processing of ANN output vector o is performed in order to determine reliability of classification. The maximum value of ANN output o_{max} is converted according to the equation:

$$o'_{\max} = \frac{o_{\max}}{\sum_{i=0}^{3} o_i} \cdot o_{\max}, \quad o_i \in [0,1] \tag{1}$$

If $o'_{\max}$ is greater or equal to the threshold T, a gesture connected with the output o_{max} is returned as the recognized gesture; otherwise, the neutral gesture is returned which means that no real gesture is detected. This method assures that if the neural network output is not firm, no gesture is detected in order to minimize false-positives ratio. It can be noticed that T=0 turns off ANN output post-processing.

ANN was trained with a resilient backpropagation algorithm (RPROP) [Riedmiller and Braun (1993)]. Training data are acquired during calibration phase which was required at the beginning of every session with the application. The calibration consisted of 4 stages (equal to the number of gestures). During each stage, 60 frames containing gesture images were gathered (video rate is 15 fps). Feature vectors obtained from these frames formed training vectors (80% of all vectors) and validation vectors (every fifth vector). This means that total 192 feature vectors (48 for every gesture) were used for ANN training and 48 vectors were used to validate ANN after training (12 for every gesture). Five neural networks are trained based on the same data and the one with the smallest error rate of validation vector classification (with post-processing threshold T=0.5) was used for the lip gesture recognition [Dalka and Czyżewski 2010].

Experiments regarding gesture recognition accuracy have been performed using face recordings of 176 users. Detailed results of lip gesture classification are shown in Table 1. It is seen that increasing ANN post-processing threshold T improves effectiveness of neutral gesture recognition and worsens results of other three gestures classification. It is assumed that an optimal value of the T threshold is 0.5 which provides compromise between the effectiveness and the false-positive ratio of real gesture recognition.

Table 1 Results of lip gesture classification

Gesture	No. of image frames	Effectiveness of lip gesture classification for different ANN post-processing thresholds			
		T = 0	T = 0.25	T = 0.5	T = 0.75
Neutral (no gesture)	6120	92.9%	93.8%	94.9%	96.1%
Mouth opening	6120	95.4%	94.8%	92.4%	89.2%
Forming puckered lips	6120	92.5%	91.8%	88.2%	83.6%
Sticking out the tongue	6120	94.1%	93.2%	91.3%	85.6%
All gestures	24480	93.7%	93.4%	91.7%	88.6%

Achieved results of lip gesture classification are satisfactory. The total effectiveness of recognition over 90% means that on average three recognition errors appear every two seconds of algorithm working on consecutive video frames. Furthermore, due to ANN output post-processing extension, the majority of the errors emerge when the neutral gesture is recognized instead of other three gestures. These errors do not pose much inconvenience to a user and may be attenuated further by the means of simple time-averaging of lip gesture detection results.

3 Virtual Touchpad

Virtual Touchpad (VTP) is a multimodal interface enabling users to control a PC by hand gestures. The developed interface involves the software and the hardware part. Hand movements are captured by the webcam mounted on the specially designed stand located on the desk (camera's field of view contains a part of desk's plane). Hand gestures are recognized by the algorithm which is based on the SVM classifier. Six different hand gestures are detected i.e. closed/open fingers, pronation/supination hand and closed/open hand. More details related to the image analysis algorithms used by the VTP and some examples of its employment could be found in [Kupryjanow et al. 2010b].

Fig. 2 Results of the hand movements test

In this paper, an emphasis was put on the use of VTP in support of the diagnosis of Parkinson's disease (PD). The most visible symptoms of this disease are tremor, akinesia and gait disturbance. Quantitative evaluation of those symptoms is important for administering a dosage of drugs. Symptoms of the PD could be measured by the neurologist using some scales. One of the most popular is the Unified Parkinson Disease Rating Scale (UPDRS). It consists of the 42 various tests. Since the doctor assessment is highly biased by his/her experience, some objective measures have been proposed that could support this diagnosis. Three of UPDRS tests are associated with hand exercises, i.e. test No. 23 – Finger taps, No. 24 – Hand Movements and No. 25 – Rapid Alternating Movements of Hands. Basing on the VTP interface, relevant software was engineered allowing measuring the parameters of hand movements during those three kinds of activities [Kupryjanow et al. 2010a]. For all the exercises, the test statistics obtained for both hands are collected. Graphs presenting the detected gestures of left and right hand are generated. Additionally, number of detected gestures, mean gesture duration and total gesture duration are presented in two tables. Then, the specialist may analyze obtained results and compare them with data obtained from previously performed tests. In Fig. 2 sample results of the Hand Movements test are shown.

4 Auditory-Visual Attention Stimulator

The proposed method is a patent pending solution for lateralization training. The main idea of the Auditory-Visual Attention Stimulator is to perform a parallel stimulation of the sight and hearing senses using digital signal processing techniques. Modification of the visual and hearing stimuli is performed in order to relate perception by the appropriate hemisphere. In Fig. 3 the block diagram illustrating our approach is presented.

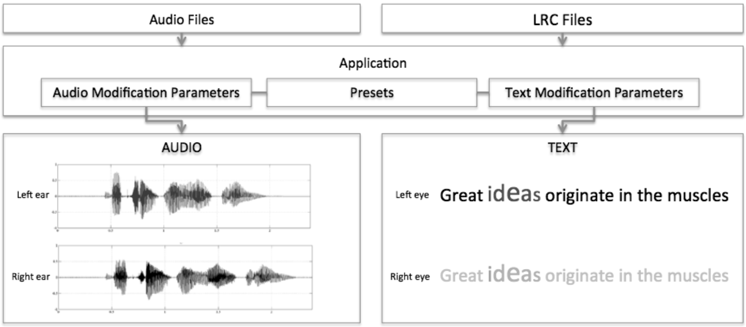


Fig. 3 Block diagram of the lateralization training system

The main assumption of the speech modification algorithm is to modify in real-time the duration of the different speech units using various time scaling factors. During the training the recorded speech and assigned text are used. The wave file is labeled manually in order to determine time stamps of the words. It is also possible to use smaller parts of speech, namely: syllables or even phonemes instead of words. The therapist can control the stretching factor, then after changing this parameter the DSP algorithm adaptation is required. This process is signalized to the user in an appropriate way.

To obtain a high quality audio signal, special speech stretching algorithm was developed. The proposed algorithm provides a combination of speech signal analysis algorithms, i.e. voice, vowels/consonants, stuttering detection and SOLA (Synchronous-Overlap-and-Add) based speech stretching algorithm. During fast speech articulation, vowels are stretched up with a greater stretching factor than consonants. During slow speech articulation only vowels are stretched while consonants remain unchanged. This method ensures constant low speed of the processed speech in both cases, which is necessary for children with CAPD (Central Auditory Processing Disorder). The stretching factor is varying during the processing and it depends on the speed of speech articulation and the speech content (vowel/consonant), according to equations:

$$\alpha_{cons}(t) = \frac{\alpha_a(t) ROS^{t}_{mean}}{(\alpha_{ratio}(t)+1)\Delta t_{vwoel} - ROS^{t}_{mean}} \tag{2}$$

$$\alpha_{ratio}(t) = \frac{ROS(t)}{ROS_d} \tag{3}$$

$$\alpha_{vowel}(t) = \eta \cdot \alpha_{cons}(t) \tag{4}$$

where $\alpha_{cons}(t)$ is the value of the scaling factor for the current frame (provided a consonant was detected), $\alpha_{vowel}(t)$ is the value of the scaling factor for the current frame (provided a vowel was detected , ROS_{mean}^{t} is the average value of speech rate in the interval used for the ROS estimation, $\alpha_{ratio}(t)$ is ratio between value of the ROS estimated for the current frame and the desired ROS_d, Δt_{vowel} is the duration of the vowel in the estimation interval, η is the ratio between the scaling factor used for the vowels and the scaling factor used for consonants [Kupryjanow and Czyżewski 2012].

The image modification is fully synchronized with speech signal. It could be achieved at the level of words or sentences. The developed system allows for a visual stimulation of the sight using an appropriate adjustment of the image parameters (Fig. 3): color; brightness; contrast; size of the letters.

In the enhanced version of the system, parameters associated with vision and hearing modification algorithms might be controlled by the eye-tracking system. The system, developed at Multimedia Systems Department (MSD) will be described in another section of the paper (Section 7). The eye-tracking system provides an opportunity to observe in the real-time where the patient focuses his/her sight. Sight reactions to stimuli are utilized in order to adjust the speech modification algorithms to patient's needs.

The proposed audio-visual stimulation method could also be used for an effective reading training. In such a scenario the same speech signal is reproduced to the both ears synchronously and the transcription of the speech is highlighted. Such a scenario was considered in the experiment. 37 children (24 girls and 13 boys) between 7 and 8 year participated in the experiment. Children's task was to find differences between the text (simple stories, adequate to the children's age) displayed on the screen and the speech heard in the headphones. When children noticed the error they had to press the space bar. During the test, four different scale factors were used: 1.0, 1.25, 1.5 and 2.0. It should be noticed that the first one corresponds to the unmodified speech rate. This value was used as a reference allowing for defining the overall reading comprehension skills of children. The whole test took about 5 minutes to complete.

The influence of the proposed audio-visual stimulation on the reading comprehension was analyzed while based on the number of undetected errors in the prepared testing sentences. In Fig. 4 the percentage of the errors in the speech transcription was presented using a box and whisker plot. The number of errors was calculated with the restriction that the child should find the error during the time when the wrongly transcribed word is heard in the headphones (this restriction is indicated as the 0 ms delay). The results show that the number of undetected errors decreased with the scale factor increase, i.e. when the speech rate is reduced.

In the further analysis, the number of NDE was analyzed for several values of the response time delay i.e. 500 ms, 1000 ms, 1500 ms and 2000 ms. The results show that when the response delay threshold rises, the number of errors decreases (when comparing the same values of scale factor) [Kupryjanow et al. 2013].

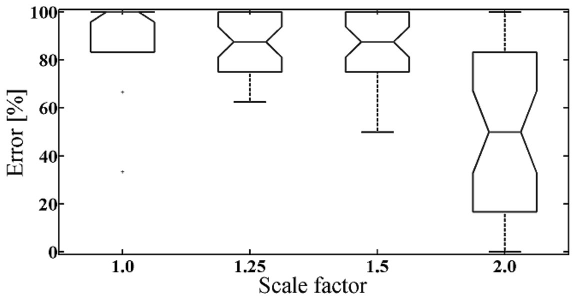


Fig. 4 Box and whisker plot presenting number of undetected errors for different scale factors

Results of the experiments indicate that the designed audio-visual stimulator could be used in order to improve the comprehension during the reading task. Independently from the assumed delay of the child response, the proposed method of stimulation had reduced the number of undetected errors, which correlates with the reading comprehension, since in order to detect the error children had to understand what they read.

This technology can be used also to enhance foreign language learning. Since the slowed-down speech is more understandable, the time required of effective learning becomes shorter. The preliminary experiments have been carried out in order to verify above thesis. The English language course has been prepared aimed at improving the following skills: comprehension, vocabulary, pronunciation, spelling, recall vocabulary and pronoun recall. 40 adult persons divided into four groups (10 persons in each group) took part in the experiment, 8 people have withdrawn from the course. All participants have declared to possess their English language skills at intermediate level. In case of three groups, different types of stimulation were used: audio and visual stimulation, only audio stimulation, and visual stimulation, only. One group was treated as the control group (without any stimulation). Results indicate that the auditory-visual stimulation is especially effective for improving recall vocabulary. The stimulation that affects only one sense (sight or hearing) is not sufficient to obtain better results in comparison to the results of the control group. Further research on this topic is planned.

5 Scent Emitting Multimodal Computer Interface

The main idea behind the developed interface was to enhance the polysensoric stimulation methods by adding a possibility of scent emission controlled by a PC computer. Various techniques of smell emission were overviewed. Finally, so-called Cold Air Diffusion was chosen. The technique consists in mechanically spreading tiny oil drops by using properly formed glass pipes and an air pump to produce underpressure over oil container. This solution can be easily controlled by a computer and makes it possible to spread the smells faster than those methods that use heating. Another advantage of this technique is very small dimensions of aroma molecules. Small molecules allow for easier filling with scent even big rooms. Moreover, small molecules easy combine with air molecules, thus in this way the scent remains in the room for a long time. In addition, the Cold Air Diffusion is protecting essential oils against overheating that could initiate chemical reaction within the oil and consequently, to reduce profitable aromatherapy effects.

Two versions of the scent emitting computer interface were developed in the scope of the project: the basic and the extended one. Both devices use USB or Bluetooth interfaces and they can work with common PC computers. The main difference between engineered versions is the number of available scent channels. The extended version (Fig. 5) is able to emit four scents independently and it is equipped with the following sensors: gas, temperature, humidity and atmospheric pressure. The basic version diffuses a single scent and it has the gas sensor built-in, only. However, it is possible to connect more than one device to the PC, simultaneously. As a result, the latter version of the interface gives therapists some more opportunities, e.g. they can built their own set of scents and locate the devices in various places of the room. More details related to the scent emission process and technical issues of the interface could be found in the previous authors' paper [Czyżewski et al. 2010].

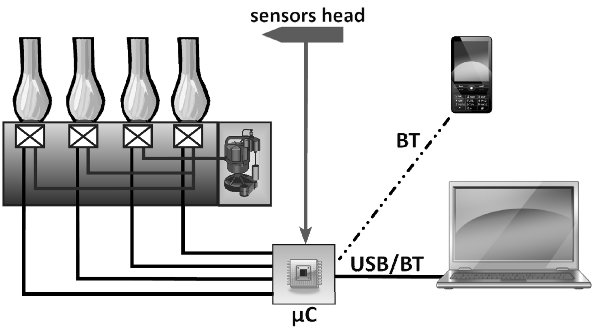


Fig. 5 Simplified block diagram of the extended version of the Scent Emitting Interface

It is assumed that the scent-emitting computer interface can be widely used in numerous health centers in much more repeatable and standardized ways than previously. Therapists will obtain a new tool that allows the emission of scents depending on particular needs. It also provides an opportunity to correlate the aromatherapy process (as a part of the polysensoric training) with images and sounds presented by the therapist.

Applications of the interface are not limited only to the polysensoric stimulation. It seems promising to use the interface as an important tool for diagnosing neurodegenerative diseases, e.g. Parkinson's disease. Olfactory disorders are found in about 90% of patients suffering from the Parkinson's disease, even in the very early period of the disease [Ponsen et al. 2004]. Currently used methods of the diagnosis, e.g. Smell Identification Test, do not allow for an accurate smell assessment. Additionally, patients often have problems with identifying some odors (e.g. skunk smell), as the tests were not standardized for European residents. The Scent Emitting Multimodal Computer Interface installed in a vented cabin can be used to carry out precise olfactory tests, even making it possible to find threshold values for some popular smells. Further research on this topic is planned.

6 Smart Pen

Numerous researches indicate that dyslexia and dysgraphia pose major problems in school learning, nowadays. Therapy for dyslexia is very important considering the fact that the number of people who suffer from dyslexia can be estimated even as 17%. This number varies according to studies conducted in different countries, e.g. it is close to 9-10% in Poland. Typical symptoms of the dyslexia are: improper holding a pencil/pen in the fingers, too much or too little pressure put on a surface (a sheet of paper), too much or too little pressure applied to a pencil/pen, crossing the lines or contours, changing direction in the drawings, and many others.

The Smart Pen was designed as a tool for supporting the therapy of developmental dyslexia, with a particular regard to dysgraphia. It comprises a LCD tablet and a specially designed pen equipped with pressure sensors. The LCD tablet gives an opportunity to measure the pressure put on its surface. It is necessary to obtain the correlation between a hand and an eye. For this reason, ordinary tablets could not be used: a person who uses such a tablet (e.g. drawing a picture) would have to see results of his/her work on screen placed far from the tablet.

The main advantage of the Smart Pen comes from equipping the tablet pen with three pressure sensors. These sensors are mounted on a special grip which helps to encourage the recommended tripod finger position when holding a pen. The grip is made with a plastic material that has properties comparable to hard rubber. The grip is similar to the ordinary pen (or pencil) grips used for dyslexic children who often reveal a poor pen grip.

The software allows for monitoring of some interface parameters that are important from the therapy point of view, such as pen grip or time taken to complete an activity being a part of an exercise. The interface designed allows for interesting (play and learn) activities to be performed with schoolchildren (i.e. learning of proper handling writing tools, basic writing etc.). Five different activity types are possible to choose: drawing, squiggles, coloring, mazes, joining the dots. All of these activities are similar to those ones suggested by dyslexia therapists [Frostig et al. 1964]. The prototype of the Smart Pen is presented in Fig. 6.

Fig. 6 Prototype of the Smart Pen (during coloring activity)

Tests that have been carried out with the participation of 22 pupils of the age range of 7-10 (primary school, classes I-III). Children with special needs (autism, ADHD, mental retardation) belonged also to that group. All pupils were asked to carry out 4 activities: coloring, mazes, joining the dots and various squiggles. Each session took about 20 minutes. Special attention was given to the proper pen handling during the all activities.

Among the whole children group, the reference group was chosen. Only children without any observable problems were selected. Additionally, the teachers did not observed any writing problems in this group.

To verify whether parameters measured by the Smart Pen can be used as indicators of graphomotorical problems, additional analyses were made. The correlation coefficient between information from teachers and therapist and results obtained by children was determined for this purpose.

In order to calculate the correlation coefficient, opinions of teachers and results were presented in the binary form – meaning that child met the criterion or not. The level of confidence was set to 5%, thus it turned out that the critical value of correlation factor was equal to 0.3598. The results are presented in Table 2.

The correlation coefficient between the pen grip below the threshold value and the children diagnose exceeded the critical value for all activities. The same situation is for the pressure put on the tablet. It can be stated that the system properly detects such cases.

In contrast, results in case of too strong pen grip and too strong pressure put on the table are not distinct. It may be explained by the fact that children could have been under stress. They tried to hold the pen in the most correct way possible. In result, they held the pen too strong. This observation was proven even by the children: after completion of activities they admitted that held the pen more than during normal writing.

Table 2 Correlation coefficients for different parameters

	Squiggles	Mazes	Joining the dots
Too weak pen grip	0.4109	0.5657	0.5181
Too weak pressure put on the tablet	0.6446	0.4384	0.4536
Too strong pen grip	0.0871	0.0987	0.1361
Too strong pressure put on the tablet	0.4324	0.3363	0.5804

Tests proved that using the Smart Pen it is possible to distinguish children without motoric disruptions and those who may be affected by dyslexia/dysgraphia. It can be done by analyzing different parameters, such as the pen grip and the pressure put on the tablet. Children with graphomotorical problems more frequently obtained worse parameters in activities that required precise pen movements. Furthermore, pupils taking part in the tests expressed their vivid interest in working with the system [Odya et al. 2010].

7 Cyber Eye

The Cyber Eye system is an eye-gaze tracking computer interface developed at Multimedia Systems Department. Applications of technology based on the control by gaze will become widespread in the near future. Analysis of the user's visual attention could be employed in many areas of everyday life. Hitherto, eye-gaze tracking systems have been employed in typical applications like usability of websites evaluation or communication with physically handicapped people, as well as in some more original ways – in stimulation of auditory-visual attention, including the support of dyslexia therapy, in the diagnosis and therapy of binocular depth perception [Kosikowski and Czyżewski 2009; Kosikowski and Czyżewski 2010], in studies of cross-modal phenomena related to visual-auditory perception [Kunka and Kostek 2010; Kunka and Kostek 2012] and in awareness evaluation of the post-comatose patients [Kunka et al. 2012]. A more detailed information on the Cyber Eye system can be found in some previous publications of the authors [Kunka and Kostek 2010; Kunka et al. 2012].

One of the most promising applications of the Cyber Eye is awareness evaluation of post-comatose patients. These patients who did not recover full consciousness are usually regarded as patients in the vegetative state (VS). It should be stressed that the ratio of misdiagnosis of VS patients touches 40% [Monti et al. 2010]. Thus, the Cyber Eye system makes a relevant diagnostic-therapeutic tool for physicians and therapists of those patients. A set of neuropsychological tasks was developed in a cooperation with the therapists of the “Light” Residential Medical Care Facility in Torun, Poland. The aim of the consciousness testing is to make a diagnosis followed by the stimulation of cognitive, linguistic and communication functions, memory, understanding, as well as logical and abstract thinking. For example, one of tasks consists in presenting three images during a specified time followed by covering them and asking the patient to indicate a specific one. The aim of this task is diagnosis and testing of the patient’s memory competence. Fig. 7 shows the post-comatose patient performing one of 11 neuropsychological tasks of the so-called “consciousness test”.

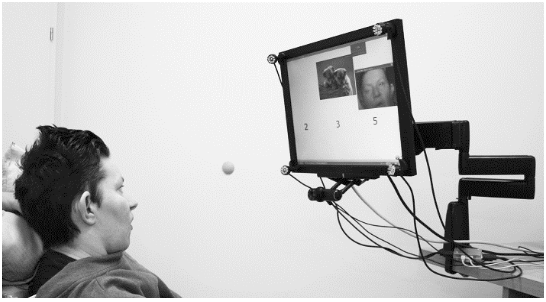


Fig. 7 A post-comatose patient during the “consciousness test” session employing the Cyber Eye

At the beginning, fifteen patients were recruited (most of them with traumatic brain injury): five females and ten males; mean age ± standard deviation (SD): 38.13 ± 11.99. Some patients breathe through a tracheostomy tube, and most of them had problems with excessive secretion impacting on his/her hyperactivity, which often required the discontinuation of the study. In addition, even though they are conscious, concentrating on one task is often problematic, and these indispositions influenced on size of final group of patients involved in the study (6 patients). The main criterion of the consciousness test was the correctness of test-related task performance. Therefore, the credibility of the data obtained from the Cyber-Eye system requires many repetitions of the test. The parameter ‘correctness’ was determined as a sum of correctly performed tasks of each examination session. Fig. 8 shows the box-whisker plot presenting the results of sessions based on the eye-gaze tracking during a 6-month observation period. It was observed that the subjects studied with the Cyber-Eye and diagnosed as VS patients demonstrated unquestionable and non-accidental signs of their consciousness.

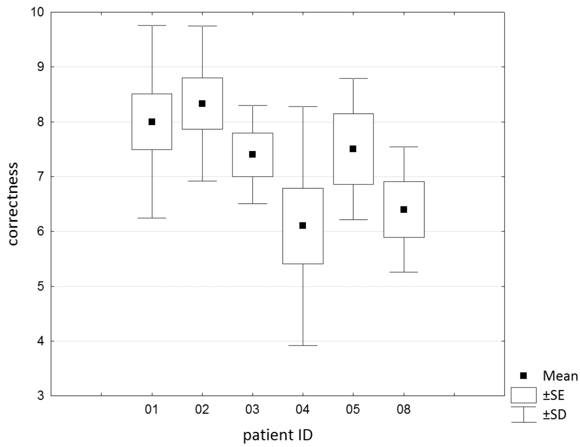


Fig. 8 Correctness of consciousness test performing within 6 months

Patient 01 achieved mean correctness value ± SD of 8 ± 1.68. This was a promising result, although patient 02 achieved a better outcome (8.33 ± 1.33). In other patients the results obtained were slightly lower: patient 03: 7.4 ± 0.8; patient 04: 6.1 ± 2.07; patient 05: 7.5 ± 1.12; and patient 08: 6.4 ± 1.02. Nevertheless, a global diagnosis of patients' condition requires the consideration of performing a subjective assessment by the therapist. In the case of patients: 02, 04 and 05, therapists evaluated their state based on gaze interaction and subjective observation. The subject 02 was diagnosed as an MCS (Minimally Conscious State) patient at the beginning of these experiments. A contact with this patient was subsequently enhanced, and he is currently regarded as an awakened patient. Moreover, he has been rehabilitated after the awakening, hence the verbal logic contact increased. Patients 04 and 05 are regarded as MCS patients.

8 Conclusions

Taking advantage of the recent progress in computing technology and of the increase of popularity of computer peripherals (e.g. USB cameras), it was possible to build new kind of computer interfaces. These interfaces adopt different types of modality and in this way they bring a possibility to use computers more efficiently. Especially, it proven to be evident in case of disabled people. The developed multimodal computer interfaces helped impaired people to communicate with other persons and decreased their social isolation, in this way. Multimodal interfaces can be also applied to therapy in case of pupils with different types of communication disorders, where traditional therapy methods proved to be rather not attractive for children, thus the progress achieved was not satisfactory in those cases.

The results of tests carried out show that users are very keen on employing the new tools to the education or therapy purposes. The feedback gathered so far shows that the technologies developed are received well by their designated users. Moreover, the tests carried out employing the users helped much to enhance the interfaces by adjusting them according to their particular needs.

Acknowledgment. Research funded within the project No. POIG.01.03.01-22-017/08, entitled "Elaboration of a series of multimodal interfaces and their implementation to educational, medical, security and industrial applications". The project is subsidized by the European regional development fund and by the Polish State budget.

References

[Czyżewski et al. 2010] Czyżewski, A., Odya, P., Smulko, J., Lentka, G., Kostek, B., Kotarski, M.: Scent emitting multimodal computer interface for learning enhancement. In: 4th International Workshop on Management and Interaction with Multimodal Information Content, Bilbao, Spain, pp. 142–146 (2010)

[Dalka and Czyżewski 2009] Dalka, P., Czyżewski, A.: Lip movement and gesture recognition for a multimodal human-computer interface. In: Proc. International Multiconference on Computer Science and Information Technology, 2nd International Symposium on Multimedia – Applications and Processing, Mrągowo, Poland, pp. 365–369 (2009)

[Dalka and Czyżewski 2010] Dalka, P., Czyżewski, A.: Human-computer interface based on visual lip movement and gesture recognition. Intern. J. Comp. Sci. and Appl. 3(7), 124–139 (2010)

[Frostig et al. 1964] Frostig, M., Miller, A.M., Horne, D.: The developmental program in visual perception: beginning pictures and patterns. Follett Publishing Company, Chicago (1964)

[Kosikowski and Czyżewski 2009] Kosikowski, Ł., Czyżewski, A.: Computer Based System for Strabismus and Amblyopia Therapy. In: Proc. International Multiconference on Computer Science and Information Technology, 2nd International Symposium on Multimedia – Applications and Processing, Mrągowo, Poland, pp. 493–496 (2009)

[Kosikowski and Czyżewski 2010] Kosikowski, Ł., Czyżewski, A.: Binocular vision impairments therapy supported by contactless eye-gaze tracking system. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) ICCHP 2010, Part II. LNCS, vol. 6180, pp. 373–376. Springer, Heidelberg (2010)

[Kunka and Kostek 2010] Kunka, B., Kostek, B.: Exploiting audio-visual correlation by means of gaze tracking. Int. J. Comp. Sci. and Appl. 7(3), 104–123 (2010)

[Kunka and Kostek 2012] Kunka, B., Kostek, B.: Objectivization of audio-visual correlation analysis. Arch. Acoust. 37(1), 63–72 (2012)

[Kunka et al. 2012] Kunka, B., Czyżewski, A., Kwiatkowska, A.: Awareness evaluation of patients in vegetative state employing eye-gaze tracking system. Int. J. Artif. Intell. Tools 21(2), 1–11 (2012)

[Kupryjanow and Czyżewski 2012] Kupryjanow, A., Czyżewski, A.: Methods of improving speech intelligibility for listeners with hearing resolution deficit. Diagn. Pathol. 7, 129 (2012)

[Kupryjanow et al. 2013] Kupryjanow, A., Kosikowski, Ł., Odya, P., Czyżewski, A.: Auditory-visual attention stimulator. In: Proc 134th AES Convention, Rome, Italy, Preprint no. 8810 (2013)

[Monti et al. 2010] Monti, M.M., Laureys, S., Owen, A.M.: The vegetative state. BMJ 341, c3765 (2010)

[Odya et al. 2010] Odya, P., Czyżewski, A., Grabkowska, A., Grabkowski, M.: Smart pen - new multimodal computer control tool for graphomotorical therapy. Intell. Decis. Technol. J. 4(3), 197−209 (2010)

[Ponsen et al. 2004] Ponsen, M.M., Stofers, D., Booij, J.: Idiopathic hyposomia as a preclinical sign of Parkinson's disease. Ann. Neurol. 56, 173–181 (2004)

[Riedmiller M and Braun 1993] Riedmiller, M., Braun, H.: A direct adaptive method for faster backpropagation learning: the RPROP algorithm. In: Proc. of IEEE International Conference on Neural Networks, San Francisco, USA, pp. 586–591 (1993)

Controlling and Coordinating Computers in a Room with In-Room Gestures

G. Tartari[1], D. Stødle[2], J.M. Bjørndalen[1], P.-H. Ha[1], and O.J. Anshus[1]

[1] Department of Computer Science, University of Tromsø, Norway
giacomo.tartari@uit.no
[2] Northern Research Institute, Tromsø, Norway
{jmb,phuong,otto}@cs.uit.no, daniel@norut.no

Abstract. To interact with a computer, a user can walk up to it and interact with it through its local interaction space defined by its input devices. With multiple computers in a room, the user can walk up to each computer and interact with it. However, this can be logistically impractical and forces the user to learn each computer's local interaction space. Interaction involving multiple computers also becomes hard or even impossible to do. We propose to have a global interaction space letting users, through in-room gestures, select and issue commands to one or multiple computers in the room. A global interaction space has functionality to sense and record the state of a room, including location of computers, users, and gestures, and uses this to issue commands to each computer. A prototype has been implemented in a room with multiple computers and a high-resolution, wall-sized display. The global interaction space is used to issue commands, moving display output from on-demand selected computers to the large display and back again. It is also used to select multiple computers and concurrently execute commands on them.

1 Introduction

The Global Interaction Space (GIS) system lets a user in a room of computers use one or several computers through in-room gestures without needing to walk up to a computer, or do remote logins to use it. It makes using multiple computers more efficient, including making it efficient to select any computer as the source for data to be visualized on another computer. The GIS system tracks users, analyzes their body movements to identify gestures like selecting a computer, and executes predefined scripts on a selected computer. A user can select multiple computers and with a gesture run a script on all of them. With a set of gestures it is possible to designate a computer to be the consumer and another computer to be the producer of data exchanged through a data network. This can be exploited to select multiple computers and with a simple gesture have them all send visualization output to a shared high-resolution display wall.

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,

DOI: 10.1007/978-3-319-08491-6_9,

Mobile and desktop computers have a set of technologies sensing actions done by a user. The technologies include keyboard, mouse, microphone, touch display, accelerometer, and a camera. The physical volume of space sensed is the local interaction space of the computer. In a room with several computers there are several local interaction spaces. A local interaction space has limitations. It is designed for a single computer, and is not aware of other computers and their respective interaction spaces. It is also typically range limited to covering the area near to a computer, and does not extend across a room. Consequently, a user must walk up to a computer to use it. Also, a user in a room cannot efficiently do interactions involving several of the computers in the room simultaneously. Local interaction spaces are also typically somewhat different from each other depending upon the operating system used. Operating systems and user interfaces implemented by Microsoft, Apple and Google, have distinct differences even if they are using many of the same basic ideas for interaction. Consequently, users have to learn how to use multiple local interaction spaces. During actual in-room use, this can be confusing and can lead to unintended user input.

To bridge the gap between computers and different interaction approaches, and to support interaction at a distance with computers in a room, the concept of a global interaction space is proposed. While users still can interact with individual computers through local interaction spaces, coordination between computers and actions meaningful for multiple computers in a room are done through a room's global interaction space. Tools like ssh and VNC/Remote Desktop also let a user interact with multiple computers. However, these tools merely extend the physical reach of the local interaction space of each computer. The user must still interact with each computer's local interaction space, without an efficient way of interactively controlling multiple computers in a room.

As an example of the use of a global interaction space, consider a room with multiple users and computers, and a shared display wall. The display wall is used to display output from one or several computers. The display wall has a local interaction space allowing users to stand in front of the display wall and interact using arm gestures. The display wall's local interaction space lets a user manipulate output rendered on the display wall, which is produced by one or more of the computers in the room, through a touch-like interface. However, to initialize output from a computer to the display wall a user may have to walk up to or remotely log into the computer and interact through its local interaction space to send output to the display wall. This must be repeated for each computer producing output to be viewed at the display wall. If the output from each computer is to be coordinated, the user must move over to each computer or do remote logins and set them up to do so. While doing remote logins may sound like a useful tool, the user must have a handheld mobile device or a laptop to do remote logins while standing in front of the display wall. This would impair the user's movements and his ability to use the display wall's touch-like interface. This is not efficient for a user standing at the display wall and is disruptive to the workflow of the user.

The global interaction space, on the other hand, will simply let a user standing in front of a display wall point to some computer to select it as the producer of output, and then point to a location on the large display to select where the output is to be displayed.

Finally, having a global interaction space in a room allows for having computers and displays in the room without their own local interaction spaces. The global interaction space can let users interact with such computers and displays. The Global Interaction Space was first presented in [Tartari et al. 2013].

2 Interaction Spaces

We define an interaction space as a volume within which user input can be detected. An interaction space exists orthogonally to the computers it acts on. Fig. 1 a) illustrates a room with a display wall and multiple interaction spaces. The tablet and laptop in the figure both have their own local interaction spaces. These interaction spaces only enable interaction with their corresponding computer. Next to the display wall canvas, a row of cameras creates another local interaction space [Stødle et al. 2009]. This interaction space acts on software running on the display wall, enabling multi-point, touch-free input by multiple users simultaneously. The display wall itself consists of a canvas, projectors and a display cluster. Since visualizations running on the display wall typically are parallel applications with some shared state, the touch-free display wall interaction space, while local to the display wall, acts on the state of all software running on all the computers in the display cluster and is in this sense a global interaction space for those computers. Finally, the global interaction space encompassing the room is illustrated.

An interaction space has several defining characteristics: (i) The approach taken to detect user input; (ii) number of simultaneous users; (iii) dimensionality of interaction; (iv) volume of space covered and (v) fixed or moving. The approach taken to detect user input includes camera-based detection of gestures, microphones listening for sounds, hardware buttons detecting clicks and optical sensors detecting mouse movement. Most existing input devices support only a single user at a time, however for an interaction space covering a large volume of space, support for multiple users is often necessary. The dimensionality of interaction determines how the user's input is perceived by the interaction space. A depth camera provides the basis for four dimensions of information (spatial 3D and time) about the user, a regular mouse provides 2D information while a volume knob provides just 1D information. The volume of space covered by different interaction spaces varies. A mouse attached to a computer by wire is rather fixed in place limited by the length of the cable and that a user must physically touch the mouse. However, if the mouse were wireless the resulting interaction space, while still small, would be moving rather than fixed at one single location. A camera or a microphone can effectively cover larger volumes. The touch-free interaction space for the display wall's multi-point input system is fixed in space, but covers a large area. A simple mobile interaction space is akin to a wireless mouse or the touch sensor on smartphones. A more elaborate movable interaction spaces can be constructed using steerable cameras [Stødle et al. 2007].

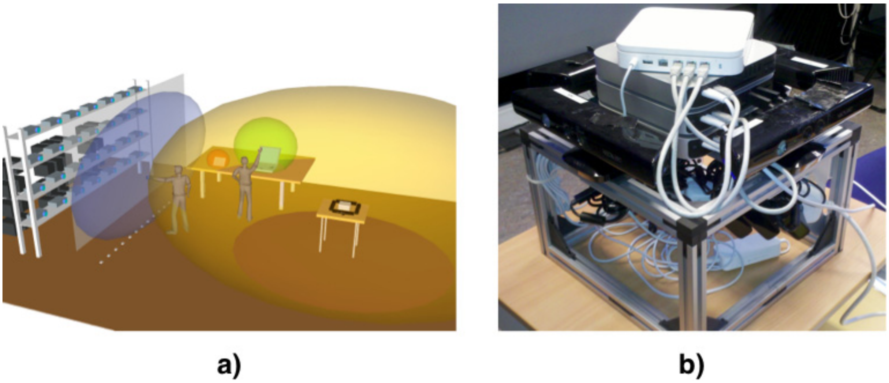


Fig. 1 a) Three Local and one Global Interaction Spaces. b) Hardware prototype of sensor suite with processing for the Global Interaction Space.

3 Global Interaction Space Interface

We define gestures as movements of limbs and body that express some predefined meaning. To detect gestures we need to sense and track movements of a user, and then analyze these to discover if there are any gestures in them.

If we can translate some of a user's movements into character sequences, match these sequences with predefined regular expressions within an acceptable delay; we can build a system applicable for an in-room global interaction space with a limited vocabulary and relaxed interactive demands.

Assuming a technology to obtain a 3D scan of the user, such as a 3D camera, and assuming that the user is facing the camera, we can detect and track six well defined points of the user: the closest point to the camera, the farthest point from the camera, the top-, bottom-, left- and rightmost points of the user. Given that a plane is defined by a point and a normal we can combine these six points with the reference axes of the camera coordinates to obtain an axis aligned bounding box enveloping the user, see Fig. 2 a). Assuming also that we are tracking the points used to generate the bounding box, we are in effect surrounding the user with six virtual touch screens always in contact with the user, Fig. 2 b). The movements of these points are then used to produce character sequences and matched with regular expressions to detect gestures. No single body part is tracked, such as hands or head; the user is free to use any part of his body to perform the gestures, Fig. 2 c).

In our prototype the points on the planes behind and below the feet of the user are harder to track, but the four remaining planes have proven to be sufficient for a minimal gesture set.

To illustrate the generation of the string characters consider Fig. 2. First of all we define which user movements we want to be translated into characters. We call this set a dictionary. Fig. 2 d) illustrates the motion dictionary. The arrows represent the movements the system will look for, and the character each movement will produce. For simplicity and efficiency we considered only motions along the major axes and the diagonal between them. For example, an up-arrow implies an arm or body part moving straight up and being interpreted as a character, in this case n.

Fig. 2 e) illustrates a circular motion and the characters produced by doing it. A full circular movement of an arm/hand produces the characters “nuersdwln”. The system takes into account character repetition and speed of the user movement so that characters are produced only in a predefined speed range.

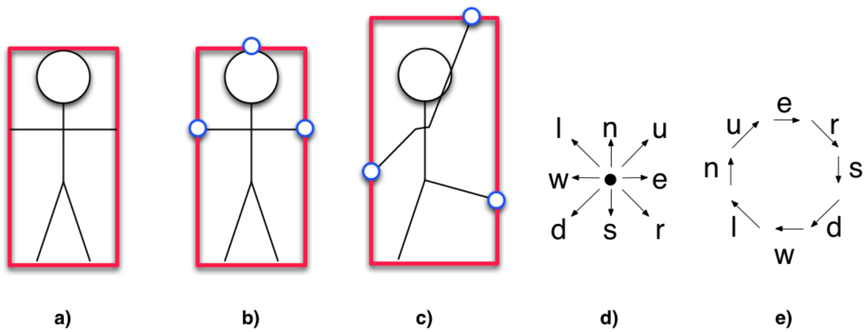


Fig. 2 Bounding box and regular expression examples: a) Bounding box surrounding the user. b) Bounding box with points used to generate it in evidence. c) Possible use of a bounding box. No single body part is tracked just the point generating the bounding box. d) Motion to character dictionary e) Example of circular gesture character production.

To actually track a users body and define a bounding box, depth images from a 3D camera, like the Kinect, is suitable. A 3D camera makes it simple distinguish a user in front of the camera from the background. A bounding box is calculated with a single scan of a depth image.

Through practical use of the GIS system, we have identified two types of GIS gestures of special importance: context selection, and execution. The context selection type comprises a gesture for selecting a computer, and a gesture for selecting an action (a script) on the computer. The execution gesture starts execution of the selected script on the selected computer.

Coordination between computers can be done using the two basic types: to direct a data stream from one computer to another the user selects both computers and the relevant script on each, and then performs an execute gesture.

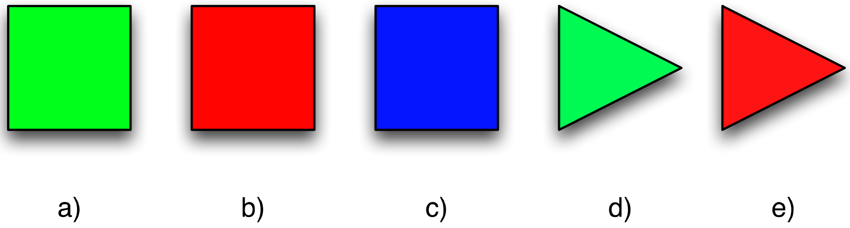


Fig. 3 Visual Feedback a) no selection, b) single selection, c) multiple selection, d) script running, e) error while running a script

Giving the user clear visual feedback on which gestures have been detected by the system, and which computers are selected and what actions will be done, is important to aid the user and achieve the intended goals. The Global Interaction Space uses simple colored geometrical shapes to let a user across the room see the effects of gestures. These shapes are displayed on a display associated with each computer in the room Fig. 3 illustrates the idea of using visual feedback to the user doing gestures. A green square is displayed when a computer is ready for selection, but has not been selected. A red square is displayed when a computer is the only computer selected. If multiple computers were selected when a single selection gesture is done, the other computers are deselected. A blue square is displayed on each computer selected by a gesture allowing multiple selections.

When a computer has been selected by a selection gesture, it displays in addition to the appropriate square, a set of different colored circles indicating the scripts that can be selected (not shown in Fig. 3). In practice, the number of scripts needed for an in-room global interaction space is low, allowing for just a few actions, and using colors is enough to distinguish between just a few scripts.

A computer running a script displays a green triangle during the execution of the script, and a red triangle in case some error occurred during the execution. If no errors occurred the computer returns to its previous state of selection when the execution terminates.

4 Related Literature

There is a large body of research literature on interactivity and interaction; here we focus on techniques and tools to build interaction spaces. In [Stødle et al. 2009] a distributed optical sensor system implements a device free interaction space. The system uses an array of commodity cameras as 3D multipoint input to track hands and other objects. The tracked coordinates are used to provide 3D input to applications running on wall-sized displays. We extend this system to suit a multi display environment, we also use commodity 3D cameras to provide room wide depth information and gesture detection.

LightSpace [Wilson and Benko 2010], uses multiple depth cameras and projectors. The projectors and cameras are calibrated to real world coordinates so any surface visible by both camera and projectors can be used as a touch screen. Adding the 3D world coordinates of the detected users to this multi-display installation allows for different multi-touch body gestures, such as picking up an object from a display and putting it on another. The body of one or more users can be used to transfer objects to different displays by touching the object on the first display and then touching the other display. As in LightSpace we use the free space in a room to interact, but we focus on providing a system where users can interact with multiple stand-alone computers in the room as well as new computers entering the room, and fulfill actions involving several computers.

Another relevant work is [Van den Bergh and Van Gool 2011], where a novel algorithm is presented that detects hand gestures using both RGB and depth information. The prototype is used to evaluate the goodness of hand gesture detection techniques using a combination of RGB and depth images. To evaluate the algorithms, a device free interaction system is developed and tested. In the same context [Kim et al. 2012] presents Digits, a personal, mobile interaction space provided by a wrist worn sensor. The sensor uses off-the-shelf hardware components. Digit can detect the pose of the hand without instrumenting it. Both [Van den Bergh and Van Gool 2011] and [Kim et al. 2012] provide interactive input by gestures, but none of them, to our knowledge, provide interaction with many computers, which is the focus of our work.

In [Bragdon et al. 2011] a system designed to support meeting of co-located software developers explores the space of touch and air gesture in a multi-display multi-device environment. The system is composed of a 42" touch-display, two Kinects, a smartphone and a tablet. Mid air gestures, like pointing to an object on the bigger display, are supported through the Kinect. In combination with a touch-enabled device, the hand gesture can be augmented to address some of the problems of gesture detection, such as accidental activation or lack of tactile response. Our prototype is not tailored to a specific task and promotes interaction between many computers not necessarily with a shared display and many handheld devices. Our system also focuses on executing actions on the involved computers as a form of interaction, letting the users interact and coordinate with different computers in the room.

[Ebert et al. 2012] describes a touch free interface to a medical image viewer used in surgery rooms. The system uses a depth camera and a voice recognition system as a substitute for keyboard and mouse input in an environment where touching those devices can compromise the operation. As in our system, there is an interaction space enabling the use of a computer with hand gestures, but we differ in the number of computers and in the more general solution not targeted to any specific software or scenario.

5 Architecture

A room has multiple computers with displays. Some computers are primarily fixed in place, while others are mobile and frequently change location in the room as users move about. Users move to and from any computer in the room and interact with applications running on one or several of the computers through each computer's local interaction space. Users can interact with multiple computers through the room's global interaction space.

A global interaction space comprises a set of functionalities, see Fig. 4, divided into a global side and a local side.

The global side comprises: Room Global State Monitoring (RGSM): In-room users are sensed through a sensor suite covering all or parts of the room. The location of the computers in the room can also be determined by the sensors (however, presently we use a static map telling the system the in-room coordinates of the stationary computers in the room).

Room Global State Analysis (RGSA): The sensor data is analyzed on the fly to detect the state of the room, and in particular detecting users and gestures.

Gesture to Action Translator (GAT): Data representing users and gestures are then translated into actions according to the mapping given by the pre-defined Gesture to Action Dictionary (GAD).

The local side comprises: Action Executor (AE). It executes actions on a computer in a room. Each action is defined by the system, Actions Definitions by Room (ADR), and given to each computer when it enters the room.

Visual Feedback (VF): The users can see the status of the local side and in particular the status of the actions execution on the monitor connected to the computer, if there is one. The Visual Feedback (VF) renders the visual representation of the local side status.

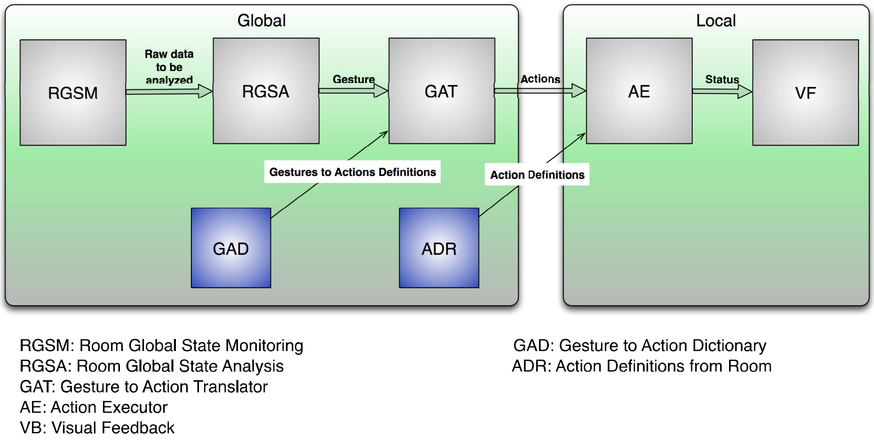


Fig. 4 Architecture

6 Design

The design of the GIS prototype is combination of the functionalities identified by the architecture into components, and is organized as a pipeline (Fig. 5). Each stage of the pipeline does some processing on the data to refine it gradually, from raw sensor data, detecting gestures, turning gestures into actions, and then executing the actions on the relevant computers in the room. A stage does a blocking receive to wait until data from the previous stage arrives, does its processing, and finally does a non-blocking send of the refined data to the next stage in the pipeline.

The design distinguishes between functionality that should always be executed on the same computer, and functionality that can be executed on any computer.

The definition of actions, ADR, is served by a server to the computers of the room. The system then only has to send action identifiers to each computer, and the computers will use the action definitions to actually do the actions.

The system is split in two main sides: the global side representing the room, and the local side representing a computer. Two main components, Sensor and Room comprise the global side. There is only one Room component, while there can be one or several Sensor components. This design decision reflects the fact that multiple sensors can be part of the room and they can be distributed around the room to better cover its entirety. For this reason the system must be aware of the number of Sensor components. In addition the Sensor components handle gesture detection, which is a potentially CPU intensive task. In this way the system can distribute the computational load to the different computers driving the sensors.

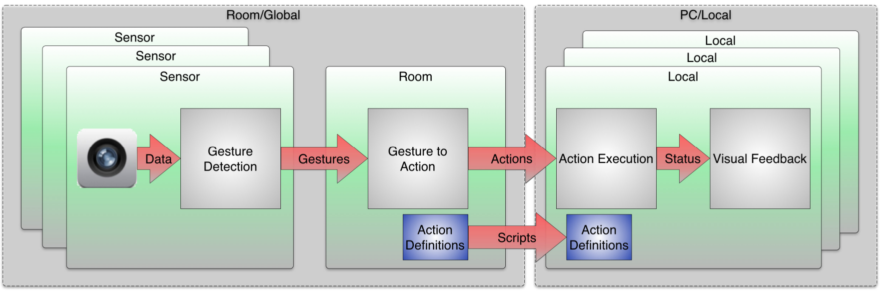


Fig. 5 Design (description given in the text)

There is only one Room component in the system. The motivation behind this choice is to keep the global state of the room in one place. This simplifies the management of both Sensors and Local components. The detected data about body movements flows from the sensors to the Room component that routes commands to the correct computers.

Having the global status of the system available in one point simplifies the logic to compute the above-mentioned steps. Another advantage of a single Room component is that the action definitions repository can be in one place that all the other components of the system already know.

The local side is represented in Fig. 5 as multiple instances of the Local component running on each computer that is part of the room. The Local component receives the action definitions from the Room component and stores them for later execution. It also listens for actions sent from the Room component and, upon receiving one, executes the appropriate action.

7 Implementation

The prototype system is implemented as a pipeline (Fig. 6), primarily using the Go programming language. Go's CSP like [Hoare 1978] features are well suited for the implementation of a pipeline allowing isolation of the stages of the computation and communication through channels. Fig. 1 b) shows a hardware prototype of the sensor suite.

The ADR server is implemented as an http server, and the ADR client is an http client. The action definitions are scripts written in a scripting language assumed to be widely available; the implementation uses Python.

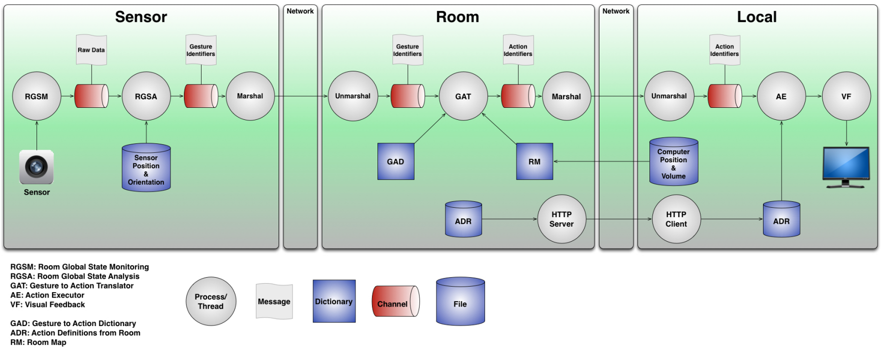


Fig. 6 Implementation (description given in the text)

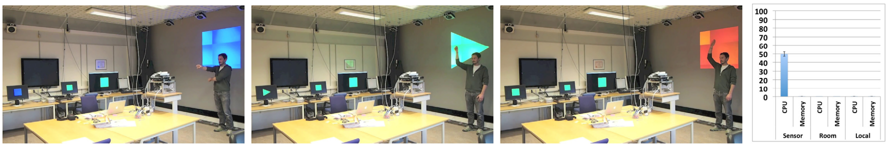
The current implementation assumes both sensors and computers in the room to be stationary, which simplifies handling the geometry of the room and its initialization. At startup, Sensor and Local components read a configuration file containing position, orientation and bounding box of Sensor and Local components. The Sensor component uses the position and orientation of the sensor to transform the gesture coordinates from sensor coordinates to room coordinates. This relieves the Room component of such calculations and allows adding any number of sensors.

The Local component uses the bounding box to communicate its position and volume of interaction to the Room component. The Room component uses this information to send actions to the Local component in which a gesture is detected or to which a gesture is directed. The bounding box is stored on the Room component into the Room Manager dictionary along with the TCP socket to the Local component it came from. In this way a Local component is bound to a volume of the room.

A prototype of the Sensor component is shown in Fig. 1 b), the prototype is built using four MS Kinect devices as sensors and two Mac Minis, each controlling two of the four Kinect. The sensors are oriented in opposing directions to cover as much space as possible in the room. The other components, Room and Local, are also running on a Mac Mini each. All the Mac Minis used have an Intel i7 CPU at 2.7 GHz and 8GB RAM interconnected to the same Gigabit Ethernet switch. To detect users gestures we have not used any of the proprietary drivers for the MS Kinects, such as MS Kinect SDK or OpenNI, but we choose to use the open source drivers provided by the OpenKinect community. The motivation for this is efficiency, portability, and access to source code. The open source drivers, however, do not provide any skeletal recognition as the proprietary ones, but only make available a depth image. A depth image is an image in which each pixel corresponds to the distance form the camera to the detected object.

8 Experiments

A set of performance experiments has been conducted on the prototype system (Fig. 7). For the experiments, we configured the system with (i) two computers, each with two cameras, and each running the RGSM and RGSA; (ii) one computer running the GAT; and (iii) multiple in-room computers, each running the AE and the VF. All computers were Mac Minis at 2.7GHz and 8GB RAM interconnected to the same Gigabit Ethernet switch.



a) b) c) d)

Fig. 7 Experiment setup and performance measurements

We used the Python psutil module, to measure each computer's CPU utilization, amount of physical memory in use, and network traffic. We also measured the delay from when a gesture was initiated by a user to the global interaction space, and until the corresponding action was started at the in-room computer. We measured this by capturing a video at high frame rate of the user and of a display behind the user. We counted the frames from when the user starts moving his right arm and until the display shows that the corresponding action of drawing a triangle on the display happens.

The results show that the CPU utilization for the computers running the RGSM and RGSA is close to 50%, see Fig. 7 d). This is not an unexpected result given that the computers process the output of two cameras at 30 frames per second. The result of the processing is then translated to strings of characters representing the user's movements. The strings are then matched against the regular expressions for gesture recognition. The computer running the GAT has a very low CPU utilization of 2%.

The network traffic is insignificant in all cases. The frequency of sending gesture and action identifier data is less than one per second because the user cannot do more gestures/sec and recognize the visual feedback. The system also transmits less than 100 Bytes per gesture or action. This is because we primarily transmit identifiers for gestures and actions between the computers. This is possible because we customize the in-room computers with all action scripts when they enter the room instead of sending the scripts to them every time they are needed. If we assume we have 64 scripts to download, and that each script is 16KB, we need to download only 1MB to a computer as a one-time download. Assuming we have many computers needing downloads simultaneously, even a Wi-Fi network will easily have enough bandwidth to do this sufficiently fast for users not to be noticeably delayed. In practice we expect the number of scripts to be much less than 64, and their size on average to be less than 16KB.

The latency from a gesture is completed and recorded by the RGSM until the corresponding action is issued by the AE is 824ms. This is interactively fast, and users should not notice delays in practice.

We conclude that the prototype is not CPU, memory, or network limited, and interactively fast. This is because we intentionally do very simple sensing and gesture detection. The system has most of its CPU and memory resources available for more advanced functionalities and for other workloads. With more advanced sensing and gesture detection we expect a higher CPU and memory load.

9 Conclusion

We work daily in a room with both a large, high-resolution display wall and many computers. We have extensive experience with what we need to improve in the interaction with the resources of the room to achieve an efficient workflow. We frequently need to be able to rapidly select computers for coordinated commands including place shifting their display output onto the display wall. Traditionally, we do this by manually doing remote login or walking up to computers and issuing commands to them. While this works, it involves multiple steps for the user, and it is disruptive to the workflow.

The global interaction space in the laboratory allows us to do this both simpler for the user, and with fewer disruptions to the workflow. It is also quite satisfying to effortlessly steer multiple computers. We have not implemented complicated gestures, and gesture detection is easily done interactively fast.

We have found that very simple gestures, like raising an arm above your head, can be applied in multiple settings. Simple gestures are surprisingly useful while being cheap and fast to detect.

We expect that complex gestures and the corresponding actions can significantly increase the delay. The gestures suitable for a room are not as interactively demanding as gestures done on, say, a tablet. A few hundred milliseconds is a long time for an interactive interface, but may be more than fast enough when selecting computers in a room and issuing commands to them.

It is possible to give the user the possibility of creating gestures by giving the system user defined regular expressions, as seen in [Kin et al. 2012].

The local side is customized by the global side by being given a set of actions that the global side is willing to offer to the users. We have found the principle of customization to be a simple way of making the local side do exactly what the global side has defined. To customize a computer entering the room, a one-time overhead is taken when downloading action scripts to the computer. This reduces the traffic between the global and local side when low latencies matter the most, which is during actual use of the global interaction space.

The Global Interaction Space system makes it more efficient for a user to control and apply multiple computers in a room, at the same time, and from across the room. The user of the system is through in-room gestures able to select one or multiple computers, select action scripts on each computer, and start execution of the scripts. The gesture detection uses regular expressions encoded from the user movements inside the room.

Acknowledgment. This work was funded in part by the Norwegian Research Council, projects 187828, 159936/V30, 155550/420, and Tromsø Research Foundation (Tromsø Forskningsstiftelse). We would like to thank Ken Arne Jensen for helping us build the sensor suite and make it look like something done by the creatures in Alien. Thanks also to Joseph Hurley for proofreading.

References

[Bragdon et al. 2011] Bragdon, A., DeLine, R., Hinckley, K., Ringel, M.: Code Space: Touch + Air Gesture Hybrid Interactions for Supporting Developer Meetings. In: Proc. ACM International Conference on Interactive Tabletops and Surfaces, pp. 212–221 (2011)

[Ebert et al. 2012] Ebert, L.C., Hatch, G., Ampanozi, G., Thali, M.J., Ross, S.: You Can't Touch This Touch-free Navigation Through Radiological Images. Surgical Innovation 19(3), 301–307 (2012)

[Hoare 1978] Hoare, C.A.R.: Communicating Sequential Processes. Commun. ACM 21(8), 666–677 (1978)

[Kim et al. 2012] Kim, D., Hilliges, O., Izadi, S., Butler, A.D., Chen, J., Oikonomidis, I., Olivier, P.: Digits: Freehand 3D Interactions Anywhere Using a Wrist-worn Gloveless Sensor. In: Proc. of the 25th Annual ACM Symposium on User Interface Software and Technology, pp. 167–176 (2012)

[Kin et al. 2012] Kin, K., Björn, H., DeRose, T., Agrawala, M.: Proton: Multitouch Gestures as Regular Expressions. In: Proc. SIGCHI Conference on Human Factors in Computing Systems, pp. 2885–2894 (2012)

[Stødle et al. 2007] Stødle, D., Bjørndalen, J.M., Anshus, O.J.: A System for Hybrid Vision- and Sound-Based Interaction with Distal and Proximal Targets on Wall-Sized, High-Resolution Tiled Displays. In: Lew, M., Sebe, N., Huang, T.S., Bakker, E.M. (eds.) HCI 2007. LNCS, vol. 4796, pp. 59–68. Springer, Heidelberg (2007)

[Stodle et al. 2009] Stødle, D., Troyanskaya, O., Li, K., Anshus, O.J.: Tech-note: Device-free Interaction Spaces. In: IEEE Symposium on 3D User Interfaces (2009)

[Tartari et al. 2013] Tartari, G., Stødle, D., Bjørndalen, J.M.: Phuong Hoai Ha, and Anshus OJ Global Interaction Space for User Interaction with a Room of Computers. In: Proc.: 2013 The 6th International Conference on Human System Interaction (HSI), pp. 84–89 (2013)

[Van den Bergh and Van Gool 2011] Van den Bergh, M., Van Gool, L.: Combining RGB and ToF Cameras for Real-time 3D Hand Gesture Interaction. In: 2011 IEEE Workshop on Applications of Computer Vision (WACV), pp. 66–72 (2011)

[Wilson et al. 2010] Wilson, A.D., Benko, H.: Combining Multiple Depth Cameras and Projectors for Interactions on, Above and Between Surfaces. In: Proc 23nd Annual ACM Symposium on User Interface Software and Technology, pp. 273–282 (2010)

A BPM-Based Approach of Human Task Support within Automated Workflows

S. Holzmüller-Laue[1], B. Göde[2], and K. Thurow[1]

[1] celisca- Center of Life Science Automation, University of Rostock, Rostock, Germany
{silke.holzmueller-laue,kerstin.thurow}@celisca.de
[2] Institute of Automation, University of Rostock, Rostock, Germany
bernd.goede@uni-rostock.de

Abstract. The contribution addresses the potential of model-driven workflow automation that allows new kinds of human system interaction. A control of the end-to-end process includes the seamless control of human tasks in conjunction with distributed automation solutions in addition to the simplified application-technical system integration. The human task support for applications in the life science automation increases significant the efficiency in the execution of experiments, serves to reduce stress for involved employees, and provides important contributions to the improvement of the quality management in the research. Mobile devices assist interesting opportunities for notifications and provision of information. The presented approach takes up the standard BPMN 2.0 published in 2011 as the graphical modeling and automation language for any business processes.

1 Introduction

The generalized life science automation (LSA) includes all application areas of the laboratory automation, e.g. healthcare, pharmaceutical developments, chemistry and biology. The current workflow automation in the R&D laboratories is limited by the structure of the subprocesses. On so-called automation islands special applications are performed automatically controlled by appropriate process control systems (PCS). The automated workflow is defined in a manufacturer-specific method on such automation island. Not all tasks can be performed by machines. Manual activities (e.g., preparatory lab activities, documentations, experiment evaluation with result computation, interpretation, knowledge extraction and reporting) are supported by various information systems, but without embedding of these „Human Tasks“ in a sufficiently continuous workflow control function. There are regulations, e.g., so-called Standard Operation Procedures (SOP), which are basis of the daily experimental work. But nobody in research labs checks and documents the compliance with these regulations. To a large degree, the documenting information systems are isolated from the solution of the process automation such as the control functions in Process Control Systems (PCS) or the method management of lab device systems.

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_10,

The flexibility in the research currently limits the automation level particularly with regard to the process documentation and the data handling. A large, reliable documentation level is an important tool for quality assurance not only in the regulated area but also in the research, where the extensive process documentation ensures the availability of the process know-how. Due to the growing amount of data resulting from the process automation (e.g. by complex lab robots for high sample throughput) the automation of data handling is of enormous importance. It also provides important contributions to improve efficiency and quality assurance.

The challenges and goals of the work presented here address the integration of all involved subprocesses - whether automated or manually performed, independently in which organizational unit – in a centrally controlled workflow, which is called "end-to-end-process". One of the key tasks is the flexible system integration of all used IT systems and automation islands of the high hybrid and heterogeneous laboratory environment. This paper focusses especially the subtheme of the integration of manual processes, which is not considered in common automation concepts. The human task integration includes the consideration in an automated time- and state-dependent control of the workflow, notifications of pending tasks, and the selective allocation of relevant information and functions. This can speed up the workflows and relieve the laboratory staff, who is often involved in several simultaneous processes.

The solution proposed in this paper is based on process automation with methods of the advanced business process management (BPM). BPM provides now an interesting approach to the fusion of process analysis including numerous documentation options with the generation of executable process flow models and in the end automation solutions and execution systems with process engines. The connection of various automation islands with different manufacturer-specific process control systems into one process description requires a uniform notation. The proposed solution uses the Business Process Model and Notation (BPMN) that is established to model and perform business processes. A model-based automation concept provides the flexibility of research processes with short lifecycles. The opportunities of BPM-assisted workflow control benefit the efficiency and the quality of end-to-end-processes in R&D in addition to time savings and stress reduction for involved employees.

The presented paper is organized as follows: section 2 explains the background regarding to BPM and the process notation BPMN followed by remarks to typical processes of the application domain LSA. In section 3 explanations of appropriate automation strategies for selected classes of subprocesses follow. Finally, an application example in which the focus is on manual activities shows results of a performed proof of concept and highlights the potential of BPM for the human task interaction.

2 Background

2.1 BPM and BPMN

The scientific and business discipline Business Process Management provides industry-independent concepts, methods and technologies to support the design, simulation, execution, monitoring and the analysis of business processes. A business process is an arbitrary sequence of activities that go from an initial state to one or more final state(s). The process logic combines the activities with events and gateways to route processes. The activity can be executed automated, semi-automated or manual. The human task support occurs via automated messages, signaling to members of a role, user forms or a just-in-time support of required integrated information systems for a suitable working environment.

For the definition of such processes the graphical notation standard BPMN has been established in the recent years, which pursues the goal to overcome the understanding gap between IT-specialists, automation engineers and business users [OMG 2011; Allweyer 2011; Silver 2011]. With the BPMN 2.0 published in 2011 a direct executable notation is available (ISO/IEC standard 19510:2013). Until now transformations from one notation into another (e.g. BPEL) are necessary to execute the process. BPMN 2.0 displaces this problematic and time-consuming task. A graphical notation of the process flow leads to a simple changeability. This is an important requirement in the flexible daily work in the life sciences. An executable BPM process model consists of the BPMN diagram as the graphical description and all necessary information to perform the model such as process variables, roles, actors, decisions, event definitions, user interfaces for tasks, input and output data for tasks including their mapping to process variables, messages and interfaces to external system, or the business logic for tasks. The process engine is capable of executing this model directly. Thereby users will be notified of pending tasks, user interfaces are provided and external systems are accessed via defined interfaces. Business Process Management Systems (BPMS) connect process modeler and process engine in one platform enhanced by functions for process monitoring, process administration and for the human task management. Such systems usually provide extensive opportunities for the integration of third-party systems (databases, information systems etc.). The understandable notation, smarter tools and process libraries with preconfigured process models allow non-technical users to discuss, create and update powerful workflow projects.

2.2 Characteristic of Processes in the Life Science Automation

Typical, recurring subprocesses of experimental workflows in the life sciences are on the one hand manual activities such as the preparation and the follow-up of experiments: the design and planning of experiments; the supply of chemicals, tools, and labware; the storage and removal of substances including the monitoring of storage conditions; the sample preparation; the evaluation and interpretation of experiments as well as the reporting. And on the other hand there are the typically automated tasks and subprocesses such as reaction processes (synthesis of substances, biological screening or elementary steps like heating, cooling, shaking, stirring) or the characterization of samples in the chemical analytics. Depending on the used materials and existing devices, process steps like sample preparation, sampling or dosage are performed by both: manually by laboratory staff or automatically by appropriate laboratory equipment. In each sub process different software systems are used for process control, data capture, data management, and data processing. These subprocesses are connected to structured process chains partly multiple. This can result in recursive processes, e.g. sampling and chemical analysis of samples during the running reaction with feedback effect to the reaction parameter.

Service tasks or subprocesses (here as term for the process structuring, not BPMN elements) that are modeled machine-readable and performed by an automation system (assisted by PCS or instrument control system (ICS)) represent the typical island automation of a laboratory application. Table 1 shows a summary of exemplary benefits of the process-oriented approach for manual as well as IT- and automation-assisted activities related to LSA with varying complexity.

Table 1 Workflow Components and their Automation Effect

Components	Examples	Automation Effect
Elementary manual	Supply of labware; Prepare, observe, abort process on automation island; Documentation; Transport; …	Multiple information; Control of timing; Control/ signaling of status; Selection and supply of resources; Work lists; Documentation during processing
Elementary IT-assisted	Raw data transfer Raw data processing; Process documentation; Data analysis: …	Automated raw data management; Automated evaluation of measurements / extraction of results / statistical experiment planning; Automated data capture and tamper-proof storage
Complex IT-/automation assisted	Syntheses robots; Liquid handling systems; Analytical measurement system; Cell culturing	Multiple information and control of timing; Configure/ start of subsystem; Automated control of status; Exception handling within workflow automation

3 Methods

The target workflow automation for LSA does not address a fully automation by machines primarily, but a continuous IT-support of process control along comprehensive process chains with full involvement of human tasks. The last one is supply by a widespread provision of information and a just-in-time-IT-support. Improving the interoperability accelerates the handling of processes, helps to avoid errors and improves the traceability of processes, results and findings. The data flow, which is necessary for the workflow automation, can be done automatically or with the addition of human interaction. Thereby existing established information systems act as a communication buffer or central information storage for experimental data, context data, documents, and results. Such information systems are to be integrated into the overall workflow control. The benefits of the proposed workflow automation result from the integration of all process steps into a homogeneous concept for overall process control, whose definition is accessible for all end users in form of a process model. BPMN is a suitable instrument to do this. The executable process model merges all service tasks, callable subprocesses and manual activities to an end-to-end-workflow control.

3.1 Integration of Human Task-Control and Monitoring

Even in a laboratory equipped with modern automation technology, there are many manual works. Until now application rules like SOPs (standard operation procedures) define what, when, until when and with what a specific preparative or experimental work is performed. But, who does control and document the execution? The integration of human activities in the control flow provides exactly this control and the synchronization with automated tasks. For the lab personnel that results in a guided workflow. The process control sends notifications about the next tasks and provides all execution-relevant information at the right time - without paper, without mistakes. The implemented workflow automation replaces the previous time management using kitchen timer.

BPMN distinguishes two types of human tasks: the “manual task”, which provides no system interfaces to the performer, and the “user task”. For executable BPMN models “user tasks” are preferred because they are performed via computer interface which allows to capture results, status or comments and to publicize process-relevant information (parameter, method, device …) or documents (SOP, sample lists ...). External interactive IT systems or their components are called.

BPMS usually offer a Human Workflow Management with user management and roles assignment or allow the integration of existing central Identity Management Solutions (e.g., LDAP server). Based on this, work lists are created for each user to assign the pending tasks priority-driven. The performer assignment is made according to the role, the mapping of a specific user or by a parameter-based query evaluated at runtime. A good example is the supply of compounds for a screening test by the lab staff (marked in Fig. 1 with 2). Initially, the notification with the information about the compounds to be supply is sent to all employees with the role „lab staff“. A laboratory technician takes the order and executes it. This technician is the owner of this task. In the further workflow only this user gets a request to confirm the execution and to comment the results. For all other members of the role „lab staff“ this request is not relevant. With the knowledge about the task owner the confirmation request can assign to the correct technician and the execution can be documented automatically with a time stamp and the name of the person immediately after the confirmation, e.g., who does supply the compounds at which time.

Another application example for this conditional assignment of users is the pre- and post-preparation of the automation island. For reliable results it is useful the same technician prepares and cleans up the automation system (Fig. 1 – 3/4). Another variant of the task assignment is to iterate the performer over a list of users. This strategy is applicable for instance in an interoperator study within a method validation (see the example).

In addition to the task lists and corresponding dialog components, traditional communication channels such as e-mail or SMS are available to provide required information. This is especially relevant for transmission of larger amounts of information, which will be processed by the users. Thus an e-mail to start the step „barcode labeling“ includes the needed codes in the attachment or as a copyable section in the e-mail text area (Fig. 1 - 1).

The use of mobile devices is very interesting, just for the integration of manual activities into the central process control. It helps to further increase efficiency but also to relieve the lab staff. Work lists and e-mail accounts can be accessed anytime and anywhere, execution confirmations can be sent via appropriate web clients or apps. Field data can be captured on the spot (in sterile rooms or sampling points without desktop PC too). Furthermore the camera function of mobile devices can be used for visual documentation.

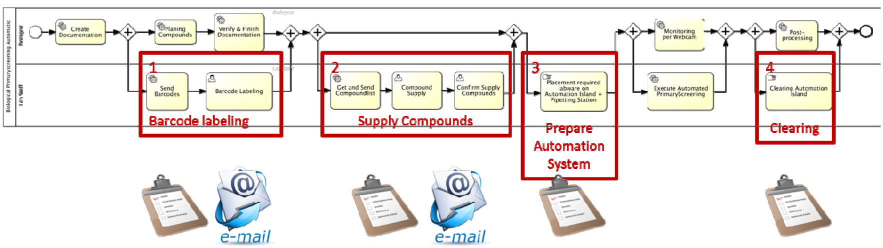


Fig. 1 Workflow with human tasks for pre- and post-preparation of experiments

3.2 Integration of Interactive Computer Systems

Besides of manual experimental work there are different human tasks using diverse IT applications. Typical examples are the experiment planning, the experiment documentation, the data visualization and analysis or the data archiving. Goals for the workflow automation are the provision of selective functions and components of the IT systems and data in the respective form and place.

Especially in the context of a detailed process documentation within an information system an IT-assisted documentation opens a considerable potential for process optimization and quality assurance. Laboratory Information Management Systems (LIMS) are established information systems for result management and documentation of experiments in the laboratories. A LIMS developed by the authors is used at celisca [Thurow 2004]. The effort-reduced, semi or fully automated completion of the process documentation with detailed information about the process flow results in fewer gaps within the process documentation and increases the traceability of processes and results. Measurements and results are integrated in the context of their generation. Functionalities of the used LIMS are integrated via web service interface to consider the authentication and the verification of access rights. In less complex information systems a direct writing access to the database of the information system with appropriate SQL-statements is possible.

The navigation effort within the information system can be significantly reduced if the current process documentation is linked directly from the process instance. Suitable dialog components of the information system can be used with appropriate context information. In Fig. 1 the task “Planning Compounds” uses the LIMS-internal user interface to assign compounds from the library to the experiment documentation. The above mentioned LIMS has a library of process documentation templates, which may be basis of new process documentations. The functionality for template-based creating can be used via web service. Current parameters are considered. In addition to the elimination of this often tedious manual work this brings the particular advantage that the documentation style of all process instances is the same, and it is guaranteed that the equal process steps and parameters have the same name. Therefore process variables and results are researchable and can evaluate together automatically.

Beneficial for transparency and reproducibility is also a reference from the LIMS documentation to the corresponding BPMN model. Thus shows exactly how the data is created. This can be done in different ways depending on the specific possibilities of the used BPMS. The possibilities range from the inserted image of the model, using a special class of information "BPMN model" (BPMN stored in the database in conjunction with BPMN editor to display) to a link to the process model in the model repository with versioning of the BPMS.

In consideration of the fact that the main objective of all laboratory activities is the generation of data, information and knowledge, the automation of the data flow is of particular importance. The aim is to make the process data, generated anywhere in any subsystem, available to all other participating systems without any delay. This requires not only data transfer mechanisms but also data transformations are necessary.

Current semi-automated analytical application systems (e.g., ICS, chromatography data systems (CDS)) of laboratory automation often offer only a data interface. In part, input parameters such as sample lists etc. can be imported and the results are available in report files in a predefined directory. The sample list is a result of a previous experiment and stored anywhere in the file system. Already simple command-line directory replication commands such as the Robocopy ("Robust File Copy" is a standard feature of current Windows, Microsoft) allow with several options for flexible copying of directories, including sub-directories on the network an automated file transfer. If the CDS report files of individual experiments are requested in its original format for data analysis, they are automatically copied in this manner to a central location too. An enterprise service bus (ESB) can automate data extractions and transformations. An ESB offers numerous ways to read data from different sources, for data selection and data extraction or format transformation, and supports the data-driven decision making related to the further process routing. The integration and automation of these rather trivial processes save time considerably and mainly avoid error in practice.

3.3 Integration of Automation Systems

To integrate the various applications of the heterogeneous lab system environment and BPMS, several interfaces have to develop individually. Between the functional integration (e.g. remote procedure call, web service) and the data integration (such as shared files or database, message- oriented middleware, enterprise service bus) can be distinguished. The application domain focuses here functionalities to start a device-specific method, which is defined in a PCS/ ICS, and the transfer of data to parameterize this method as well as the collection of results (measurements, status or location of result files). Primarily these functions and the integration strategies, which are described in the following, relate to the control of PCS/ICS- assisted automation islands. But they are applicable for automated transport steps or for predefined data processing as well.

If a laboratory system provides no possibility for external access to functions or methods, and to transfer data, then a manual initiated start of the method have to be used. In this case an authorized person receives all the necessary process-instance-specific information with the request to execution of a run on an automation system. The process engine notifies the user via the work list and/ or other communication ways. The confirmation of the execution with the resulting status allows continuing the process depending on the status (e.g. error handling by relevant experts) as well as an IT-assisted documentation of this process step. Further simplifications of the integration conditions in the structured laboratory automation by SOA interfaces are expected in the next future.

For elementary, automated functional coupling with short execution times can be accessed synchronously by a web service call or a remote procedure call on the system to be integrated. Applications of life sciences often process an experiment for a long time. For example, an automated screening can run several hours or days. Using message queuing allows an asynchronous communication and thus a loose coupling of the involved systems. Ordering and processing reaction are separated in time. The main advantage of the message queuing is the guaranteed delivery regardless of the current availability of the receiver. A message will be stored in a buffer until the receiver picks up the message. The sender can continue the work after sending a message regardless of how long the receiver needs to process the message. The systems linked by messaging have to know nothing of each other apart from syntax and semantics of the messages themselves. On this way for example a status query can be realized, which reduces the communication load on the part of the PLS considerably.

The linking of diverse automation islands to the central-controlled overall workflow increases the automation degree. This fact results in changing working environment for the laboratory personnel with a growing part of monitoring tasks. The workflow control supports this with monitoring activities using webcams or signalizations about errors or critical situations during the automated sub process.

4 Results

Typical processes of life sciences were selected for a proof of concept. These examples represent the various activities of table 2 and are used for a representative cross-evaluation of advanced BPM methods in LSA. The following example focusses the new interaction concepts for setup, configuration and execution of lab workflows provided by the BPMS-based workflow automation.

The selected process describes validation experiments in the context of the development of methods for the analysis of chiral compounds. Steps in sample preparation, dosage and evaluation of the experiment build the frame around the device-level automation of the measurement processes using a CDS (MassHunter, Agilent Technologies, Böblingen). The process described here results from the development of a high throughput chemical analysis for the determination of chiral compounds based on mass spectrometry [Fleischer 2011]. The solution can be analyzed as part of the validation methods used to evaluate the precision of interday (repeated measurements on different days), intraday (repeated measurements within a day) as well as for inter- operator experiments (repetitions by different operators). These experiments are repeated whenever the suitability of the method has to be checked for another class of materials. Intraday and interday experiments differ only in the waiting time between two experiments. In the inter-operator variant the actor assignment iterates over a user list. For these reasons the proposed model-driven workflow automation is extremely useful for this application.

The investigator is already supported in the experiment planning. The integrated LIMS provides a feature for automated template-based setup of test documentations via web service. This subprocess template will be parameterized by a number of repetitions of experiments. The supply of samples is automatically documented in the LIMS with timestamps and the name of the operator, immediately after the execution confirmation by the operator. The orders for each step will be published in the work lists of users in the human workflow management of the BPMS client. The client is also available on mobile devices. As a result, employees are immediately notified of pending tasks. Together with these task

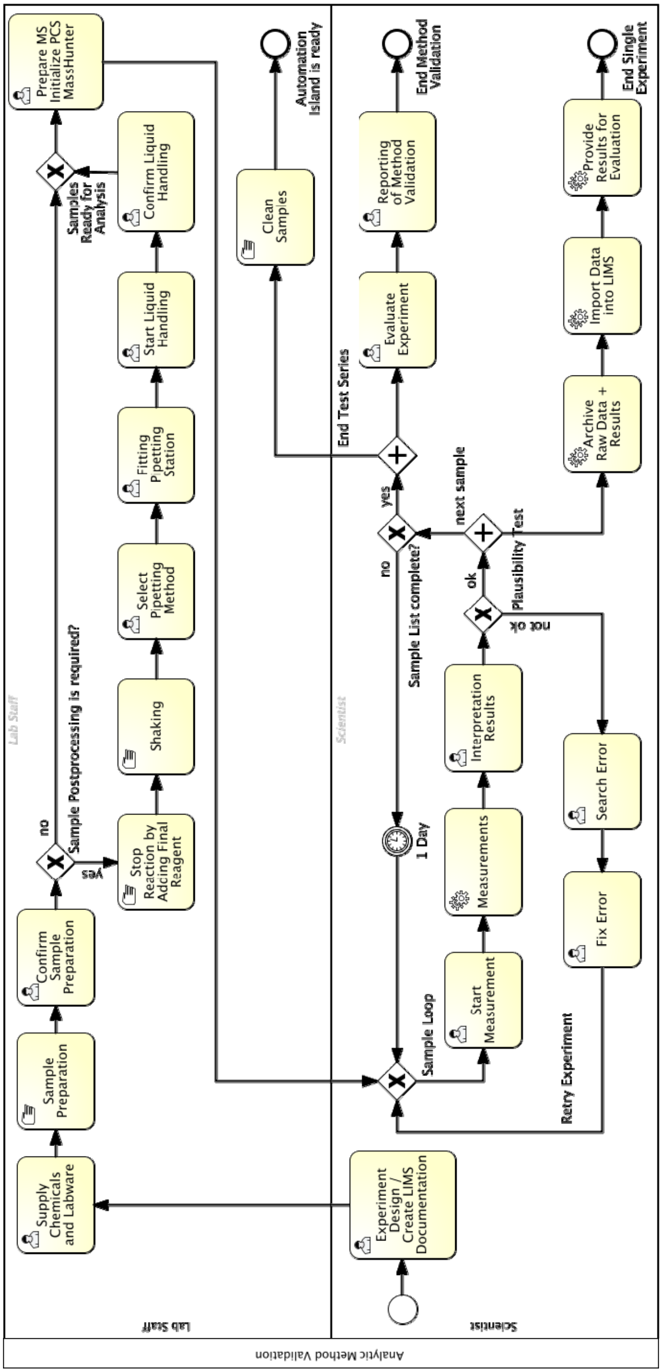


Fig. 2 BPMN process diagram of the example process

requests functions of involved information systems required in the respective process step are provided via URL calls (selection of test series in LIMS for information about experiment setup, help desk for master data information, remote desktop of ICS, knowledge base, etc.). The just-in-time method of correct pre-selection of functions and information result in noticeable reliability and time savings. This process-related human task interaction leads to the expected significant relief of laboratory staff who are involved in multiple, simultaneously running processes.

A direct online communication with the CDS is not possible due to a lack of appropriate interfaces of the CDS. The CDS-method has to be parameterized and initiated manually. After the measurement the raw data of the daily measurement are imported automatically by an in-house developed, Excel-based evaluation tool, in order to determine the mean values of the replicates, standard deviations and coefficients of variation of the enantiomeric excess. After a positive result of the plausibility test by the senior chemist, these files with derived data are copied automatically to a central backup server (file moving by Robocopy). The next step of the workflow automation controlled by BPMS integrates an ESB-job which monitors the report directory and processes incoming files. The ESB-subprocess selects the interesting data from the result file of the daily experiment and distributes them to various target locations: firstly in the database of LIMS and secondly additionally in another Excel template for the complete experiment evaluation. Until now, this task has been performed by manual copying. Now the time-consuming, error-prone manual data capturing and copying is eliminated by BPMS control. The use of the described process in all causes of method validation results in always the identical process flow, which increases the plausibility and reproducibility, especially in the method development and validation.

5 Conclusions

The tools and methods of the BPMN 2.0-based BPM open new automation targets at the workflow level in life science laboratories. BPMSs as the central component offer a comparatively open approach to adaptive workflow automation. In hybrid and heterogeneous environments of the established structured laboratory automation BPMSs provide an economic integration platform for both automated and manual subprocesses. BPMN 2.0 is confirmed as a proper process notation.

The article focuses the aspect of human system interaction of modern model-driven workflow automation from concept to effects in examples. The end user benefits from the presented model-driven approach by a much better IT support using the just-in-time method for all required IT components. The integrated documentation methodology relieves the users of an annoying part of their workload. The achieved increased degree of documentation with lower expenses for the documentation has enormous importance for life science laboratories to fulfill quality standards. The example application points out that the integration of the manual tasks in the time control by a process engine. Beside the accurate documentation of the process in the LIMS it allows a safe time-critical process handling, reducing the execution times, and finally the traceability to detect errors. In the field of complex life science applications, the combination of graphic, arbitrary detailed directly executable process models as well as a link between the process documentation and the process model allows an until now not achieved quality of know-how processing, storing and sharing.

Acknowledgment. The German Research Association (DFG) is gratefully acknowledged for funding this project. We also wish to thank the German Federal Ministry of Education and Research (BMBF) and the Ministry of Education, Science and Culture of the State Mecklenburg-Vorpommern for the financial support of celisca.

References

[Allweyer 2011] Allweyer, T.: BPMN 2.0. Introduction to the standard for business process modeling. Books on Demand, Norderstedt (2011)

[Fleischer 2011] Fleischer, H.: Rapid enantiomeric excess determination of D- and L-proline using electrospray ionization-mass spectrometry. American Laboratory 43(9), 32–36 (2011)

[OMG 2011] OMG, Business Process Model and Notation (BPMN) Version 2.0 (formal/2011-01-03) (2011), `http://www.omg.org/spec/BPMN/2.0` (accessed October 11, 2013)

[Silver 2011] Silver, B.: BPMN method and style. With BPMN implementer's guide, 2nd edn. Cody-Cassidy Press, Aptos (2011)

[Thurow et al. 2004] Thurow, K., Göde, B., Dingerdissen, U., Stoll, N.: Laboratory Information Management Systems for Life Science Applications. Organic Process Research & Development 8(6), 970–982 (2004)

A Case Study on Implementation of an Audience Response System in the Fundamentals of Computer Science Course

L. Jackowska-Strumiłło[1], P. Strumiłło[2], J. Nowakowski[1], and P. Tomczak[1]

[1] Institute of Applied Computer Science, Lodz University of Technology, Łódź, Poland
{lidia_js,jacnow}@kis.p.lodz.pl, p.tomczak85@gmail.com
[2] Institute of Electronics, Lodz University of Technology, Łódź, Poland
pawel.strumillo@p.lodz.pl

Abstract. The article presents a concept of using an audience response system (ARS) or "clickers", which can run on a mobile, tablet, PDA or PC for enhancing interaction between a lecturer and students in large classrooms. The developed interactive audio response software is based on Joomla! content management system (CMS) and its JVoteSystem component. The article compares results of students' learning outcomes for the students taught with and without question based learning methodology and also with and without using clickers software.

1 Introduction

Audience response systems (ARS) are becoming more and more popular tools for managing audience response or voting feedback. Currently, they are more commonly used in education, mostly to engage students during lectures in large classrooms and give the lecturer prompt anonymous response from the audience. ARS or "clickers" offer a specific concept of enhancing interaction between a lecturer and students. They make the lectures more interesting and interactive for the listeners and offer a rapid feedback to the lecturer about how the students comprehend new lectured material [Halloran 1995; Knight and Wood 2005; Draper et al. 2002]. Clickers are the handheld small transmitters commonly used in an audience response systems [Simpson and Oliver 2006]. A large number of clickers technical solutions, both hardware and software, are known [Carnaghan et al. 2011] and one of their limitations is the cost of the lecture room equipment for the ARS. Nowadays in the era of Internet and modern mobile technologies this inconvenience can be easily avoided.

The paper presents an Internet-based ARS application for teaching support, which can run on cell phone, tablet, PDA, laptop, PC etc. The system was developed in the Institute of Applied Computer Science at Electrical, Electronic, Computer and Control Faculty at the Lodz University of Technology (TUL) and introduced for teaching of the subject ***Fundamentals of Computer Science*** for the students of ***Computer Science*** course of study.

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_11,

The aim of the study described in the paper was also to compare the results of students' learning outcomes for the students taught with and without question based learning methodology and also a comparison of the results obtained for questions based learning implemented with and without clickers.

2 Audience Response System

Interactive audience response systems are extremely useful for activating students in large classrooms [Caldwell 2007; Morling et al. 2008]. The use ARS, however, requires preparing new types of lecturing materials, e.g. the Power Point presentations including interactive elements, like questions and quizzes. Questions can have a single correct answer or be designed without any "right" answer in order to encourage debate and discussion. Clickers can play either a secondary or central role during a lecture. They can be incorporated into a standard lecture to increase interaction between students and a lecturer or used as a part of a significant change in teaching style that promotes active learning in a class [Beekes 2006; Duncan 2005].

During traditional lectures, some students do not dare to answer questions or participate in discussions. Clickers' answers are anonymous, so the students are encouraged to give the answer even if they are not confident of their choice. For the lecturers, on the other hand, immediate response about understanding difficult problems may be helpful to pay special attention to certain problems identified by the students [Patterson et al. 2010; Greer and Heaney 2004]. Also, a style of teaching changes. Talk, dialogue and dispute during lecture, often with enthusiasm and emotion, is important for motivation, processing of facts and reasons, and for understanding and memorizing material from the class [Yourstone et al. 2008].

Clickers are ideally suited to bring about more student involvement through peer instruction or peer discussion. In this approach, teachers use clickers to survey student answers to a thought-provoking conceptual question [Duncan 2005]. If the classroom response system indicates a diversity of opinion, teachers give students several minutes to discuss the question with their neighbours in the lecture hall. It has long been known that teaching someone else helps to understand an idea. Duncan [2005] reports that this relatively easy-to-implement technique can significantly increase student learning outcomes.

Influence of clickers methodology for students' learning outcomes was tested by many researchers and teachers for different groups of students (of different size and age), for different education levels and disciplines. They reported a positive [Yourstone et al. 2008], a small positive [Morling et al. 2008] or neutral [Patterson et al. 2010] impact on learning outcomes on test scores. Some authors [Yourstone et al. 2008] cautioned that the clickers may not have been the reason for improvement in test scores and further research is necessary.

Also, the methods of testing were different, i.e. by comparing the results for two [Caldwell 2007] or more [Morling et al. 2008] independent groups of students taught the same subject parallel in different classes or by comparing the entrance examination scores with a final examination [Yourstone et al. 2008].

2.1 Typical Clickers Questions

Typically, ARS questions are introduced as a part of preparing lecture notes. Lecturers can also add questions "on-the-fly" during the class, but typically two to five questions per standard 45 minute lecture seem to be optimal [Caldwell 2007]. There are many types of questions. They can be subdivided with respect to the following teaching objectives:

- To increase or manage interaction, through questions that:
    - start or focus discussions [Halloran 1995],
    - require interaction with peers [Knight and Wood 2005],
    - collect votes after a debate [Draper et al. 2002],
- To find out more about students, by [Caldwell 2007]:
    - polling student opinions or attitudes,
    - probing students' pre-existing level of understanding,
    - asking how students feel about clickers and/or active learning,
- For diagnostic assessment, through questions that:
    - assess students' understanding of material in the lecture [Halloran 1995],
    - determine future direction of the lecture, including the level of detail needed,
    - test students' understanding of the previous lecture notes,
- For preparing for class or labs,
- To make lecture fun.

Keeping in mind the three principal educational principles, i.e.: question based learning, active testing and feedback, one may come to conclusion, that use of clickers during classes can help in achieving these principles.

2.2 ARS Software and Hardware at TUL

Internet-based ARS application for learning support (for cell phone, tablet, PDA, laptop, PC, etc.), developed at TUL, can be used in lecture rooms with Wi-Fi access and students can use their laptops or mobiles. The disadvantage of the presented Internet-based solution is, that not all students have laptops or other mobile devices. Nevertheless, they can participate in traditional question based learning without clickers or work together with their colleagues equipped with mobile devices.

The developed interactive audio response system is based on Joomla! content management system (CMS) and its JVoteSystem component. The software is installed in the Institute of Applied Computer Science at TUL on HP ProLiant DL180 server with Intel Quad-Core XEON E5620 Processor, 12GB RAM and 2TB HD. The server works under FreeBSD/amd64 operating system. It is connected into main network of TUL with fiber-optic cable of 1 Gb/s. The server supports the LAN with Fast Ethernet standard, equipped with CISCO components and connecting more than 220 workstations. It serves staff accounts, email and www. Wi-Fi access is available in the TUL campus precinct via eduroam (education roaming) access service [Jackowska-Strumiłło et al. 2013].

Joomla! is an award-winning open source CMS, which enables to build Web sites and powerful online applications in a short time. It was designed by the use of object oriented paradigms. Thanks to its module-based architecture Joomla! is easy in use and highly extensible programming tool. Joomla! uses a server design pattern, in which most of the data is processed on a Web server, and a formatted page is passed to the browser. This software architecture is particularly suitable for mobile applications.

The developed software uses Joomla! 2.5.6 with PHP 5.3, MySQL 5 data base and Apache 2 Web Server. Joomla! extension JVoteSystem component was customized for use during lectures for interactive answering to questions, which were introduced to encourage students to more active participation in the lectures.

JVoteSystem is a component for Joomla! designed to create polls. It allows Web site users to vote, answer questions, add comments and create own question answers. Main Panel of jVoteSystem is shown in Fig. 1. Overview tab is the starting page of the jVoteSystem component. Is contains the main component menu which is intuitive and user friendly. Poll tab provides an overview of all available surveys and the ability to create new ones. To add a new profile, a new poll must be created and then configured. In the first step the information in the forms in the General tab must be fulfilled, such as the title for the poll, alias or alternate short name used to identify survey questions, the categories to which the poll is to be assigned and the possible answers (their number can be freely defined). In the next step, the poll settings are configured.

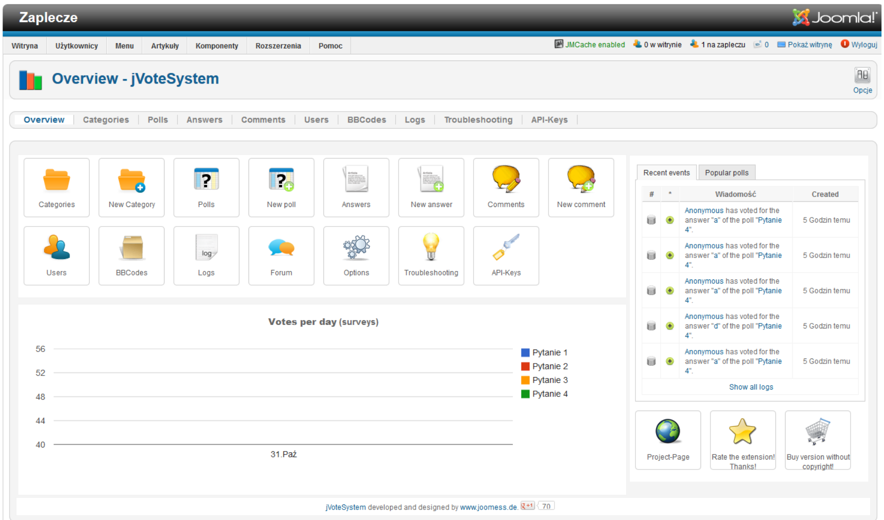


Fig. 1 Main Panel of the jVoteSystem

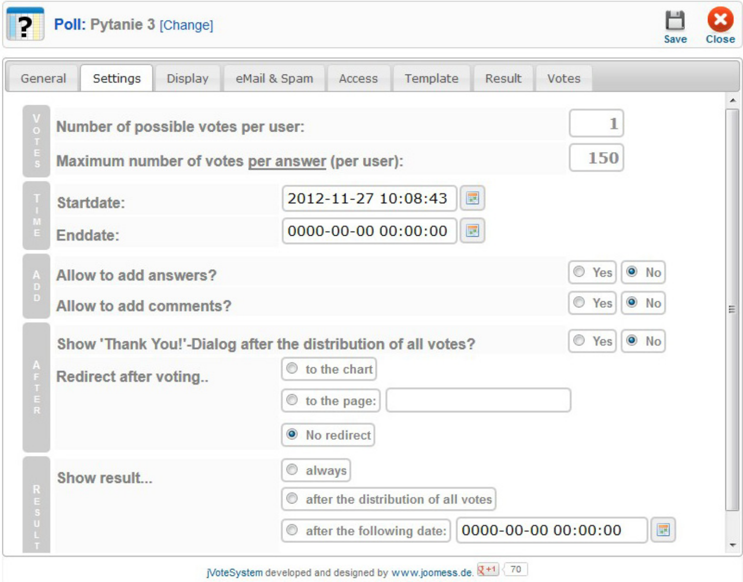


Fig. 2 Poll Settings tab - Configuring the poll

Poll Settings tab with the configuration of the poll in clickers application is shown in Fig. 2. The number of possible votes per user is set to 1, and the number of maximum possible votes per answer is set to 150. Adding own user answers and comments is not allowed. Results of the survey are shown by the teacher after the voting is finished.

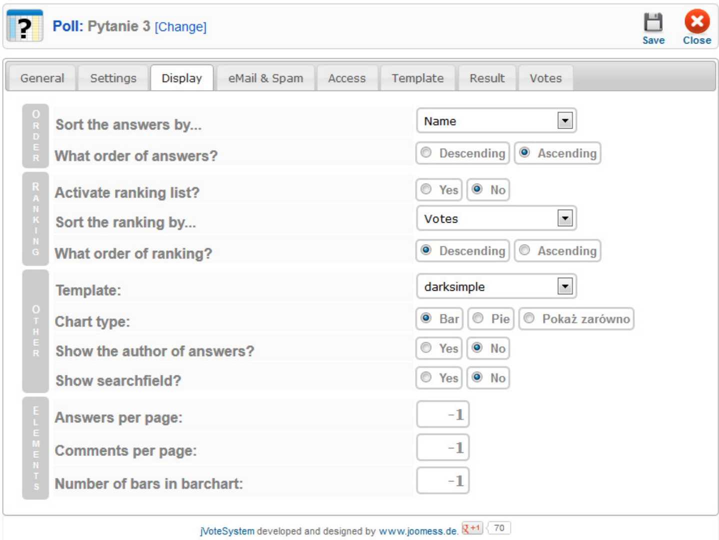


Fig. 3 Poll Display tab - Configuring the appearance of the poll

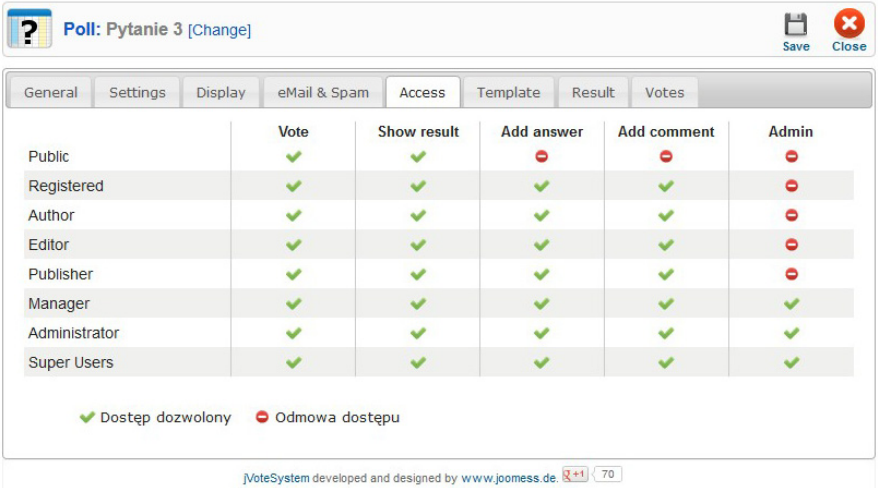


Fig. 4 Poll Access tab - Configuring the access rights

Poll Display tab (shown in Fig. 3) was used to configure the graphical user interface of the poll. A darksimple template was chosen. The number of elements, such as responses and comments, which are available on the website is not limited - the default value is “-1”, i.e. without restrictions.

Poll Access tab was used to set user groups access rights to vote, displaying the results of the survey, adding answers and comments. These settings are shown in Fig. 4. General layout of the question panel for a student is shown in Fig. 5. The questions were presented in a form allowing selection of one from four possible answers. An example summary of the audience response for the lecturer is presented in Fig. 6.

Fig. 5 Student question panel

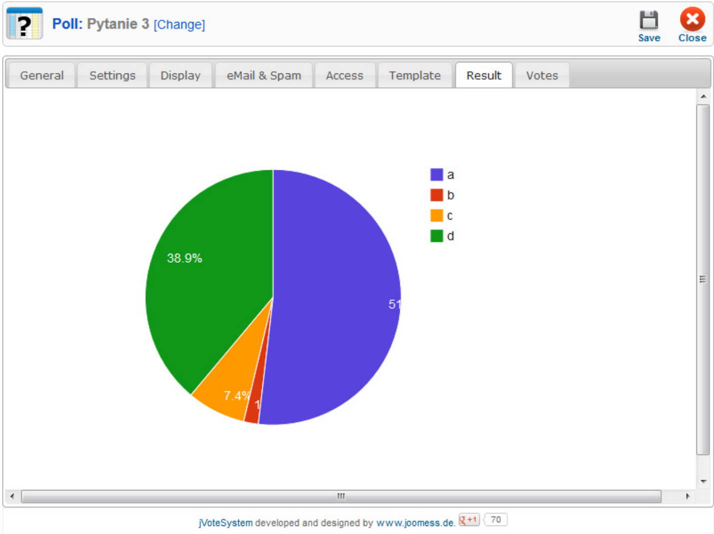


Fig. 6 Summary of the audience response for the lecturer

In the developed clickers application four questions are displayed on a single webpage. Implementation of this functionality was possible by using jVoteSystem in the form of a plug-in. JVoteSystem component is available in the form of the module and plug-in. However, if it is used as a module, a single survey can be displayed on a single webpage.

The Joomla! based ARS was used during lectures on “Fundamentals of Computer Science” given for the first course students.

3 Question Based Learning Implementation

The idea of question based learning behind clickers is not new. Teachers have used interactive, instructive questioning to teach students since at least the time of Socrates. This style of interaction, however, becomes very difficult as class size increases.

This idea was introduced firstly without ARS in the lectures for Computer Science students in 2010. Fig. 7 & 8 show examples of lecture slides with questions.

The answers are displayed on the right hand side of the slide after the discussion with the audience is completed and the answers are given. It appeared, that such a simple form of activity of students during lecture strongly increased their scores (see Table 2). Detailed discussion of assessment results is presented in the following section.

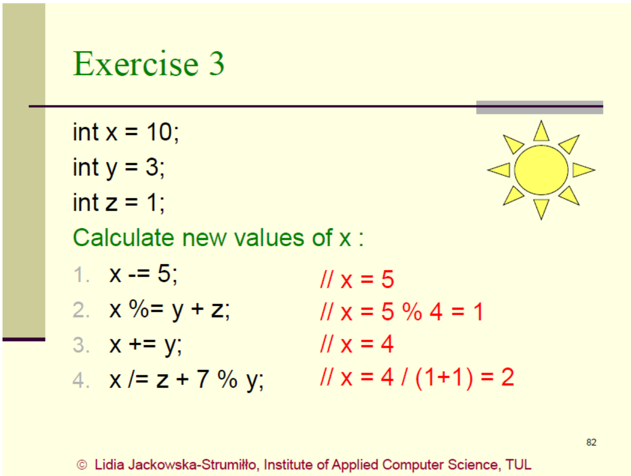


Fig. 7 An example of interactive lecture slide with questions for Test C1

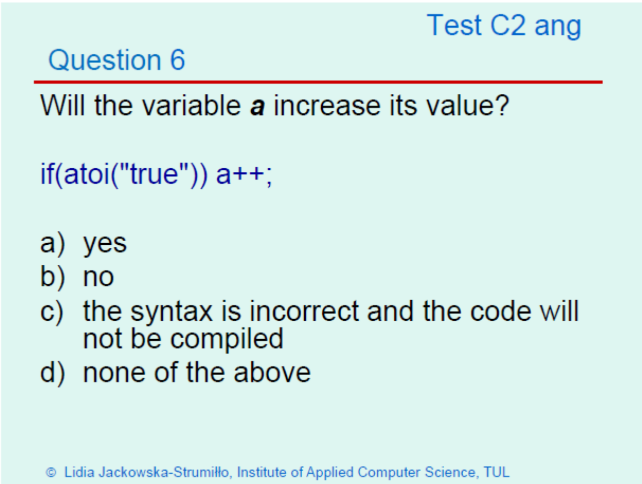


Fig. 8 An example of the interactive lecture slide with a question for Test C2

3.1 Effects of Question Based Learning

Table 1 lists mean grades obtained by the students at “Fundamentals of Computer Science” course from the “Test C1" during last six years. Question based learning methodology described in the previous section (Fig. 7) was used in the first part of a semester. The Test C1 from C programming language was performed in the middle of the semester by the use of electronic colloquium system in the Intranet of the Institute of Applied Computer Science [Jackowska-Strumiłło et al. 2002].

Fig. 9 shows distributions of grades of the students' groups in years 2010-2012. It can easily be seen, that mean grade has strongly decreased in years 2010-2012, also the number of the best grades (5) at that time was reduced. During the same period, the number of unsatisfactory grades (2) was almost doubled in years 2010-2011. One can ask an open question: is the general diminishing of grades a common tendency in higher technical education because students rely on commonly available lecturing materials and do not participate in lectures and therefore obtain lower grades? The answer is not clear, but perhaps clickers can avert this tendency.

Table 1 Students' mean grades from "test C1" in years 2007-2012

Year	Mean Grade	Std. Dev	No. of students
2007	3,47	0,74	157
2008	3,45	0,98	130
2009	3,48	0,92	155
2010	3,10	0,90	160
2011	3,10	0,89	168
2012	3,32	0,86	171

Question based learning methodology was not used in the second part of the semester. It was introduced firstly without ARS in 2010 (Fig. 8) and with ARS system in 2012. Results of the C2 tests show significant influence of question based learning methodology introduced in 2010. As it can be seen from Table 2 mean grade raised from 2,55 to 3,81 – the best result in last 6 years. While, Fig. 10 shows, that use of clickers technology during lectures did not increase significantly students' grades, but obtained notes were distributed more uniformly and there was more good and very good marks.

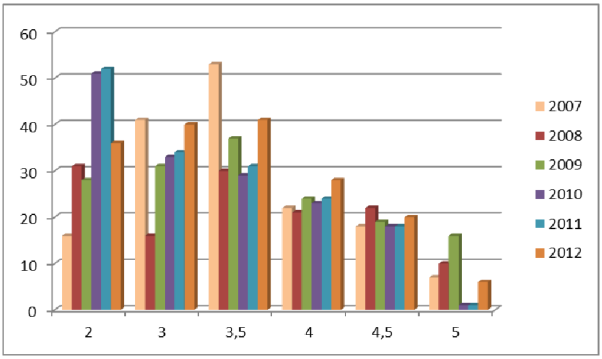


Fig. 9 The histogram of student grades of Test C1

Table 2 Students' mean grades from "test C2" in years 2007-2012

Year	Mean Grade	Std. Dev	No. of students
2007	2,92	0,83	140
2008	2,61	0,86	117
2009	2,55	0,77	149
2010	3,81	1,00	151
2011	3,53	1,01	163
2012	**3,60**	**0,98**	**136**
2012 – lectures with clickers			

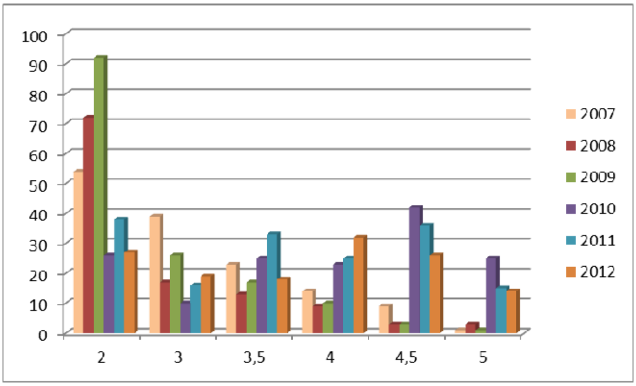


Fig. 10 The histogram of student grades of Test C2

We conclude that, despite a little influence on students' grades, the Clickers methodology is a great help for a lecturer. Anonymous feedback from the lecture audience directs a lecturer to pay special attention to problems, which can pose difficulties for students.

4 Conclusions

The presented data about final grades obtained by students can lead to the following conclusions and thoughts.

Comparing the results of Test C1 and Test C2 for the same groups of students it is well seen, that even for the students groups in years 2010-2012, for which the mean grade on Test C1 has strongly decreased, the mean grade on Test C2 raised significantly, i.e. from 2,55 to 3,81. It shows that question based learning methodology has strongly improved students learning outcomes.

Clickers technology for sure encourages students to active participation in the lectures. When it was introduced to the lecture, it was perceived by the students as a new and interesting solution. Nevertheless, some of the students complained about the need to attend the lectures with their notebooks. Consequently, at some classes only half of the students actively used the clickers system.

The authors of the article will develop and enhance the Clickers system and introduce it to other lecturers. We hope, that this interactive teaching methodology will help to improve the quality of the teaching process. We hope also, that rapid development of technology will make the smartphones widely available among students.

Acknowledgment. The authors wish to thank Professor Reidar Kvadsheim from the University of Applied Sciences in Oslo and Akershus for drawing attention to Clickers as the teaching methodology.

References

[Beekes 2006] Beekes, W.: The "Millionaire" method for encouraging participation. Active Learn. Higher Educ. 7(1), 25–36 (2006)

[Caldwell 2007] Caldwell, J.E.: Clickers in the large classroom: Current research and best-practice tips. CBE-Life Sciences Education 6(1), 9–20 (2007)

[Carnaghan et al. 2011] Carnaghan, C., Edmonds, T.P., Lechner, T.A., Olds, P.R.: Using student response systems in the accounting classroom: Strengths, strategies and limitations. J of Acc. Ed. 29(4), 265–283 (2011)

[Draper et al. 2002] Draper, S., Cargill, J., Cutts, Q.: Electronically enhanced classroom interaction. Aust. J. Educ. Technol. 18(1), 13–23 (2002)

[Duncan 2005] Duncan, D.: Clickers in the Classroom: How to Enhance Science Teaching Using Classroom Response Systems. Addison Wesley and Benjamin Cummings, New York (2005)

[Greer and Heaney 2004] Greer, L., Heaney, P.J.: Real-time analysis of student comprehension: An assessment of electronic student response technology in an introductory earth science course. Journal of Geoscience Education 52(4), 345–351 (2004)

[Halloran 1995] Halloran, L.: A comparison of two methods of teaching: computer managed instruction and keypad questions versus traditional classroom lecture. Comput Nursing 13(6), 285–288 (1995)

[Jackowska-Strumiłło et al. 2002] Jackowska-Strumiłło, L., Bieniecki, W., Brożyński, T.: Interactive computer system for checking students' knowledge in the Intranet. In: 10th International Conference on Information Technology Systems SIS 2002, Lodz, Poland, pp. 84–89 (2002) (in Polish)

[Jackowska-Strumiłło et al. 2013] Jackowska-Strumiłło, L., Nowakowski, J., Strumiłło, P., Tomczak, P.: Interactive question based learning methodology and clickers: Fundamentals of Computer Science course case study. In: 6th International Conference on Human System Interaction, Sopot, June 6-8, pp. 978–971 (2013) IEEE Catalog Number CFPI32ID-CDR, ISBN 978-1-4673-5636-7

[Knight and Wood 2005] Knight, J.K., Wood, W.B.: Teaching more by lecturing less. Cell Biol. Educ. 4, 298–310 (2005)

[Morling et al. 2008] Morling, B., McAuliffe, M., Cohen, L., DiLorenzo, T.: Efficacy of personal response systems ("clickers") in large, introductory psychology classes. Teaching of Psychology 35(1), 45–50 (2008)

[Patterson et al. 2010] Patterson, B., Kilpatrick, J., Woebkenberg, E.: Evidence for teaching practice: The impact of clickers in a large classroom environment. Nurse Education Today 30(7), 603–607 (2010)

[Simpson and Oliver 2006] Simpson, V., Oliver, M.: Electronic voting systems for lectures then and now: A comparison of research and practice. Australasian Journal of Educational Technology 23(2), 187–208 (2006)

[Yourstone et al. 2008] Yourstone, S.A., Kraye, H.S., Albaum, G.: Classroom questioning with immediate electronic response: do clickers improve learning? Decision Sciences Journal of Innovative Education 6(1), 75–88 (2008)

Multi-disciplinary, Global Student Collaboration

A.E. Milewski[1], K. Swigger[2], and F.C. Serce[3]

[1] Department of Computer Science and Software Engineering,
Monmouth University, West Long Branch, NJ, USA
a milewsk@monmouth.edu
[2] Computer Science and Engineering Department,
University of North Texas, Denton, TX, USA
kathy@cs.unt.edu
[3] Department of Information Systems Engineering,
Atılım University, Ankara, Turkey
fcserce@atilim.edu.tr

Abstract. The goal of this study is to understand the dynamics of collaboration within globally-distributed teams working in a realistic Human-System Interaction (further called HSI) environment and Software Engineering context. Quantitative data on communications were collected by capturing virtually all of the communications between the team members. Qualitative data were collected through the interviews conducted by the involved instructors. The results reveal some of the challenges associated with working in interdisciplinary and global settings and suggest areas of caution for such HSI educational experiences in the future.

1 Introduction

Teaching students the skills required for successful Human-System Interaction (further called HSI) design and evaluation is an interdisciplinary task, intensely collaborative and requiring broad experience with respect to its relevant application contexts. We describe a classroom project that sought to provide undergraduate students of software engineering with experiences to exercise skills in these key areas and at the same time to gather data to understand the mechanisms of HSI process and teamwork.

First, the interdisciplinary nature of the field is a common topic in many HSI courses; the field has inputs and interactions with a wide diversity of professions [Rogers et al. 2001]. However, it is often difficult to arrange such a class so that students experience this diversity. One way to provide for a more complete understanding of interdisciplinarity is to compel students to experience the entire lifecycle of a product, from design through development and testing, and therefore interact with others working in different roles. While many current HSI courses include a class project involving the design of a product user interface, our course also included software development and testing of the design in order to show

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_12,

students how the different parties – including designers, developers, testers and users – must interact to produce a good product. Together, students from the two courses planned, designed and built a complete software product and jointly followed a common lifecycle process.

Second, accomplishing HSI properly requires close collaboration with other project members and in recent years, that collaboration has increasingly meant collaboration with globally distributed team members. It is now common for HSI designers to collaborate – as a critical skill – with project team members in other countries [Chavan and Kaushik 2004] and to overcome some of the well known difficulties associated with globally distributed work [Herbsleb et al. 2001]. Our project provided global experience by forming teams of students enrolled in a HSI course at Monmouth University (MU) in New Jersey (US) with students taking a software development course at Atılım University in Ankara (Turkey).

Finally, it is crucial to experientially demonstrate to students the breadth of content involved in the HSI field, to show it includes much more than just, for example, web design. Our students had the opportunity to follow recent trends by designing the user interface and implementing the software development for an Android App. From the standpoint of HSI training, while the general principles of design remain the same as with any other product, many of the critical issues, the design guidelines and prototyping tools were specific to mobile applications [Kjeldskov and Paay 2012].

In addition, the goal of this study was to go further and understand in detail how the students reacted to these experiences–especially those related to distributed collaboration. Therefore, we collected measurements of all communications between team members with the goal of better understanding how much students in such a context communicate, what types of communication they engage in, and when during the lifecycle they communicate. Similar analysis have been reported elsewhere [Serce et al. 2009; Swigger et al. 2010; Swigger et al. 2012] but we have augmented these quantitative measures with surveys and qualitative interviews to explore students' cultural backgrounds and impressions of the experiences, their expectations and also how their perceptions of collaboration is compared with the actual measures.

2 Project Description

MU students in the HSI class worked on teams with the Atılım students enrolled in a software development class devoted to mobile application development. It was expected that all members of each team would collaborate on many aspects of the project, for example, to agree on features, user types, and use cases. Students were given several project assignments and were told to share them within the entire team and to discuss them to whatever extent was felt necessary. Students were encouraged to have at least one real-time, interactive chat session, but schedule mismatches between countries precluded that in all but one team. All other interaction was at the team members' discretion. The interactions were kept confidential and the students were debriefed at the end of the semester about the findings.

The MU HSI designers followed a variant of the Usability Engineering Lifecycle Model [Mayhew 2007], including creating (i) User Profile/Persona (ii) Critical Tasks (iii) Usability/User Experience Goals (iv) User Conceptual Models (v) Low Fidelity Prototypes, (vi) Quick and Dirty Evaluation, (vii) High Fidelity Prototypes and (viii) Scenario Testing. Atılım students were responsible for developing the android application, preparing a report including sample screens from the application and uploading the resulting android executable to a shared course management system called Redmine. Five projects were selected for this study. Students were provided a high–level description of their assigned projects, which ranged from a medication reminder application to a personal course scheduler application – all for android devices. Students enrolled in the course received about ten percent credit as part of their overall course grade for completing the project.

Teams were composed of one MU student and between one and three students from Atılım. The language for communication within the project teams was English. For each project assignment, student teams were asked to communicate with members through Redmine. This system provides for realtime/interactive chat and multiparty conferences, as well as asynchronous means of communications such as discussion forums, emails, file sharing and wiki's. The system also records all teams' interactions as well as time-stamps when the interaction occurs, and all interactions were recorded for further analysis [Swigger et al. 2010].

3 Data Collection and Analysis

This was an exploratory study of many aspects of student teamwork, including communication amount, type and temporal dynamics as well as student perceptions and insights. Therefore, it was appropriate to approach the data collection task with multiple measures and to triangulate between the quantitative and qualitative measures to draw conclusions [Jick 1979].

3.1 Communication Types

Quantitative data on communications were collected by capturing virtually all of the textual communications entered by team members. A single communication activity was defined as a single asynchronous post to a forum, a message sent, a wiki posted, or a file uploaded. These communication data were analyzed with text analysis procedures described in [Swigger et al. 2010]. The procedures consisted of two researchers/instructors manually tagging/categorizing each separate communication as: Contributing, Seeking–Input, Reflecting, Monitoring, Social Interaction and Planning. The main analysis was based on the tagging of the one experienced researcher. The other researcher's tagging was used to assess reliability. The initial inter–rater agreement based on tagging done by the researcher after two-pass training was approximately 94%. Analyses of these data were based on visual inspection and numeric description only.

3.2 Communication Pace

We were also interested in determining how the various pacing within a team varied over the length of a project. A team's pace is defined as the length of time between different communication activities within a group. After the projects were completed, the communication activities (chat, wiki, etc.) were extracted from the independent database. A single communication activity was defined as a single asynchronous post to a forum, a message sent, a wiki posted, or a file uploaded, etc. Synchronous communications were also captured and labeled as chats or chat room activities. All of these communication transmissions were selected and organized by team and then listed in chronological order.

An elapsed time, or time interval between communications, was then computed for each transmission based upon the start time for a particular team. For this study, the time interval between communications was determined by computing the total time that it took for one team member to respond to another team member's posting, chat, wiki post, etc. Team members responding to their own posts were not included in these counts. The rationale for this decision was that a self-response was not really indicative of group interactions and teaming so it should not be included in our statistics. Thus, all self-responses were eliminated from the time-interval temporal data.

We calculated the communication activities that occurred during each day, and then categorized the communication according to whether the response took place within less than 24 hours or more than 24 hours. This normalization procedure allowed us to compare work patterns between the projects.

3.3 Time Perspective

Previous work has demonstrated that people often show that they have temporal patterns in the workplace, and that these temporal patterns are often linked to an individual's culture or country [McInerney 2004]. As a result, efforts to define these time patterns and their link to various cultural norms have grown considerably in the past several years. In an effort to broaden the scope of temporal research, Zimbardo created a Time Perspective Inventory [Zimbardo and Boyd 1999] that attempts to create individual temporal profiles that may affect how individuals relate to different people, work or situations.

We used a short version of Zimbardo's inventory [D'Alessio et al. 2003; McInerney 2004]. In this survey, there are three factors. First, the Future factor reflects a person's general orientation to future events. An example of an item in this category is "I am able to resist temptations when I know that there is work to be done." In contrast to being future-oriented, there are two ways to be present-oriented: Hedonistically and Fatalistically. People oriented toward hedonistic tendencies tend to be self-indulgent, shirking most work-oriented activities. At the same time, hedonistic people are often involved in activities that require high energy. Finally, people oriented toward fatalism feel that their lives are dominated by outside forces. An example of a question in this category is "Much of my life is determined by fate."

3.4 Qualitative Data Analysis

Following the project, twelve students were interviewed to gather qualitative impressions regarding their impressions about global teams, communication difficulty, workload distribution, expectations about who would do what in the team and team cohesion. Interviewees were asked to sketch, on a project timeline, their perception of expected and actual levels of the measure in question as probes to guide further in–depth questions. Analyses of these data are based on qualitative "theory–building procedures".

4 Results

4.1 Communication Amount

The quantitative data on communication levels for members of all eleven teams across the project weeks is shown in Fig. 1. The level of communication is somewhat lower than that reported for previous similar projects [Swigger et al. 2012]. The leftmost colored section represents a 5–day Holiday at Atılım (11/4–11/11), and the right colored section shows MU final exams and the semester's end (12/12). As can be seen, communication levels started low, briefly rose and then decreased gradually until early November when they peaked near the start of the Atılım Holiday. They then dropped to near zero and gradually increased until the middle of December when MU students had exams. There was large variability in the number of communications between team members. Nearly 40% of all communications were provided by two team members.

The qualitative data on communication level concur: 10 of 12 team members noted that communication level was low or could or should have been higher. For all five of the interviewed teams, a majority of their members drew an "Expected Communication" figure that was consistently higher than their "Perceived Communication" figure. Interview data suggested that the reason for low communication levels was that students didn't feel there was much to communicate about. The requirements and tasks for the project were well explained because of their different roles, student felt separated and independent from their teammates. Four interviewees blamed their remote teammates' unresponsiveness for low communication although none of these reported attempting to increase interactions.

For most of the interviewees, the early–November peak and subsequent dip in communications is reflected in the qualitative "Perceived Communications" data, but for some it was not. The expected and perceived communication levels for a single, sample team were often not synchronized because interviewees clearly forgot the exact dates of some events and there was a great deal of variation in the specific pattern of "Perceived Communications. Still, five of the 12 interviewees show a peak in "Perceived Communications" within 2 weeks in either direction of Nov 7- the measured peak in communications. Nine interviewees drew a curve for

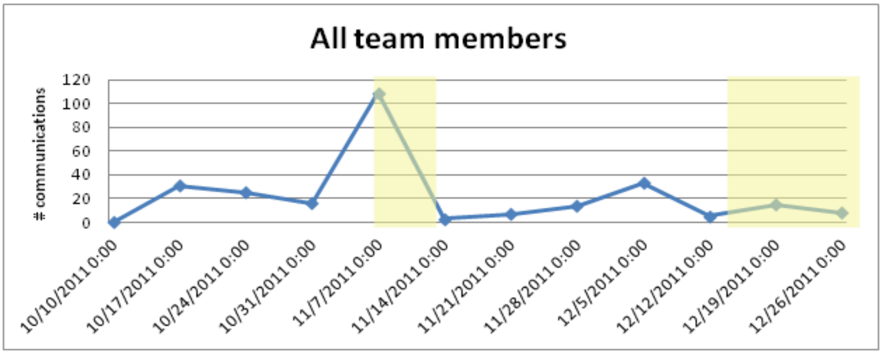


Fig. 1 Communication Levels

"Perceived Communications" that had lower communication levels in the middle of the project than near the beginning. Five of the interviewees offered that communication "fell apart" during the Atılım Holiday and "never recovered".

Interestingly, for all five interviewed teams, the HSI Designer from MU showed a "Perceived Communication" estimate that was lower than that for the Atılım developer(s). Since role and culture are confounded in this study, it could be that the HSI members perceived less communication because they felt they needed more communication due to the nature of the HSI design task. Alternatively, it may be that there are cultural differences in the tendency to say positive/optimistic phrases, so that during the interview, the Turkish members may have been hesitant to veridically report low communication levels to the researcher. Indeed, it may be that the Turkish members actually felt that there was more successful communication because they felt more in control of the project than did the U.S. Students (see the section: Time Perspective and Culture).

The quantitative data shows that MU students' average number of communications per member was higher than Atılım students' (MU=6.9 vs Atılım =3.3). This difference largely comes from the two top communicators, both from the same team and one from each location.

4.2 *Communication Type*

Fig. 2 shows the percentage of communications accounted for by each type for each week. While the general levels of communication were lower, there are many similarities between these results and those of [Swigger et al. 2012] with respect to communication types although this comparison is based on inspection rather than statistical analysis. For example, there seems to be very little activity among the teams at the beginning of the software development process (i.e., 10–30% stage), but most is socializing and planning. As the project progresses, these diminish, but contributing and information seeking behaviors increase. Planning behaviors are at their peak just before the project's midpoint as social interactions have dropped away.

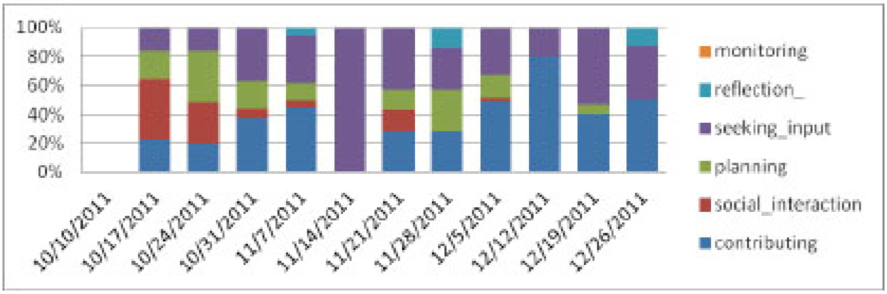


Fig. 2 Percentage Accounted for by each Type of Communication

While there are similarities with previous projects, there are differences and they appear to be the result of the specific context and schedule of this project. Specifically, the gap in communications beginning with the Atılım Holiday obviates the rise in contributing and seeking behaviors observed previously. Similarly, the rise in communication near the end of the project could not occur easily because, in the present project, the Monmouth students were preparing for and taking final exams while the Atılım students were completing the Droid apps.

Also, the current project appeared to show fewer planning behaviors than previous projects. The interviews suggested that team members viewed the project as well–organized and not difficult so that extensive planning was not required. But again, the separation of roles was commonly cited as a reason for a low level of collaborative activity while planning the project.

4.3 Communication Pace

After extracting all the transmissions from all the groups, we looked at how quickly team members responded throughout the project. Fig. 3 shows the response times that were less than 24 hours versus those that were more than 24 hours. Students' temporal response patterns seem to vary considerably over the course of the project. The response rates are much faster at the beginning of the project. These rapid response rates can be found again after about one-third of the project has been completed. While students are responding rapidly to some posts, they are also waiting to respond to others. These slower response rates peek about half-way through the project. There are some short bursts of both rapid and slower response rates towards the end of the project. These patterns are similar to what was found in the previous discussion of the communication categories.

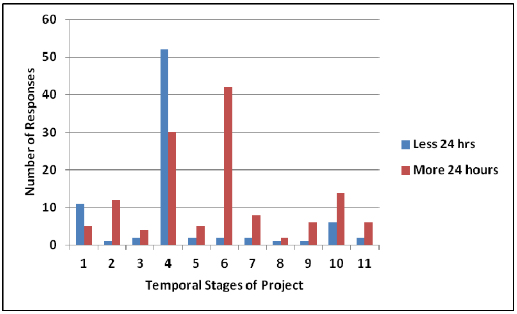


Fig. 3 Faster and Slower Response Times Across the Project

The tendency to respond more quickly to other team member's communication appears to covary with the type of communications engaged in. Specifically, planning behaviors are associated with faster responding while contributing is associated with slower responding.

4.4 Time Perspective and Culture

All students who participated in the project completed the short version of the Time Perspective Inventory. Table 1 summarizes the means and standard deviations for each country on each of the scales and also reports the t-test results comparing nationalities for all three scales in one table. In general, mean scores for all students were higher on the Future scale than on either of the present-oriented scales, consistent with the common expectation that technology students are achievers, efficient, concerned about the consequences of their action and responsible. There were no statistically significant differences between Turkish and US students in either the Future or Hedonism scales. However, there was a difference between countries in the Fatalism mean scale scores (F = 2.226, p = .118). US students had significantly higher mean scores than Turkish students (23.5 for US students versus 19.142 for the Turkish students), which indicates that US participants responded higher on the Fatalistic scale than Turkish participants.

Table 1 STPI Results across Team member Nationality

Scale	Country	N	Mean	sd	T	F
Future	Turkey	21	28.95	9.080	-1.11	.1370
	USA	11	32.36	6.423		
Hedonism	Turkey	21	27	6.148	-.65	.2603
	US	11	28.64	7.941		
Fatalism	Turkey	21	19.142	5.525	-1.52	0.0422*
	US	11	23.5	6.451		

Contrary to expectations, US students tend to believe that they have less control over their lives than Turkish students and that their lives are more dominated by external forces.

4.5 Perception of Workload

The qualitative estimations of workload (i.e. number of hours worked) for each team member were assessed during the interviews. As might be expected, "Perceived Workload" estimates generally depended on the team member's role: HIS designers perceived themselves to be working harder around the time that the persona and UE design artifacts were due while the developers perceived themselves as working harder around the time that the Droid app was due.

For seven interviewees, "Expected Workload" and "Perceived Workload" was about the same. For five of the 12 interviewees, their sketched estimation of "Perceived Workload" was generally lower than their "Expected Workload". There were two common explanations from those that rated "Perceived Workload" lower than expected. First, several developers said that when they began the project, they didn't know the tools and were concerned that learning them would take large effort. After they learned the tools, they judged the project to be easier than expected. The HSI members, who often rated expected communications to be higher than they were, on the other hand, proposed that failure to communicate a lot made the project easier, so they had extra time to accomplish other tasks.

4.6 Perception of Group Cohesion

Interviewees' perception of group cohesion was also assessed. Cohesion was defined during the interviews as the team members sense of "closeness" with other team members. MU students could only judge cohesion with remote members. Surprisingly, cohesion with local members was not consistently judged higher than with remote members. In fact, in four of five teams, the Atılım developers judged cohesion with remote and local members to be nearly identical. The one exception was a developer whose local partner dropped the class. This interviewee judged cohesion with the remote HSI designer to be higher than the absent local developer. This finding seems contrary to earlier findings suggesting lower cohesion and trust for remote team members [Al-Ani et al. 2011]. One potential explanation might be based on the assumption that cohesion was generally quite low owing to the low level of communications and the dearth of social interaction. Indeed, several interviewees admitted that social interaction and a group–based attitude was very far from reality. Given that cohesion might not have been in the interviewees' mindset, they might have drawn something to please the interviewer. This also might have been influenced by cultural differences.

The latter explanation based on cultural differences is consistent with another interesting finding: that the MU HSI designers, in all cases, judged group cohesion with Atılım to be lower than did the Atılım developers with MU. Again, because of confounded factors and cultural differences in willingness to report negative sentiments, it is difficult to explain this with certainty. It may be due to the MU HSI designers being geographically alone, while the Turkish developers often had a collocated team member, since isolated workers often feel less cohesion with their group than do those working with both local and remote team members [Milewski & Mullinix, 2005]. On the other hand, the higher degree of fatalism of MU students, as measured by the Zimbardo survey, may have led them to be more negative and less likely to rate their team's performance highly.

5 Summary and Conclusions

This study set out to provide students in an undergraduate-level HSI course with experiences in collaborating both across discipline/role boundaries and across the geographical, temporal and cultural distances associated with global work. Students were enthusiastic about the mobile applications that were designed and developed as a result of this project, and they appreciated the opportunity to work in a interdisciplinary and global setting that simulated the realities of industry.

Although the sample of students was small and there were many uncontrolled variables, the study also suggested some more general insights that can be applied to future HSI education. First, there were differences between US/MU/HSI members and Turkey/Atılım/Development members in measures of communication, perceived communication, perceived group cohesion and fatalistic attitude. Some of these differences might be construed as cultural differences that researchers must be careful about it doing future studies. This study has difficulty separating cultural differences from those associated with geographic distance, but, some cultures do either tend to be more positive or are more hesitant to report negative sentiments than are others. As HSI training becomes more global, instructors and their assessment plans need to take culture into account. Second, it may be that students need training specifically in interdisciplinary and multi–role teams. Several of our results could be explained by positing that student team members charged with different responsibilities may tend to collaborate less closely than more homogeneous teams found in the typical class project. Students could benefit from explicit training on methods of interacting with members of different skills.

Third, although cultural differences may affect students' perceived cohesion (or their willingness to report on this), there appeared to be relatively little cultural differences between Turkish and American students attitudes about time. Both the US and Turkish teams had similar views about the importance of getting things accomplished in a timely manner. Surprisingly, fatalism mean scale differences were significantly higher for the US students as opposed to the Turkish students. This was a surprising finding, since the fatalism factor is often associated with developing countries with large religious populations. Perhaps the after affects of the 2008-2012 Great Recession has caused students in the US to be much more fatalistic about their future, as opposed to students in Turkey, a country that actually experienced some prosperity during the economic downturn.

Finally, this project demonstrates how unsynchronized schedules associated with time zone and holiday differences in global student projects can result in gaps that can cause lasting difficulties by decreasing communication and cohesion. Instructors need to take special precautions to "jump-start" the project after the gap or prepare students to continue the project in some specific way.

This study has implications outside the classroom as well. First, the study demonstrates that textual analysis of student team communications can provide an interesting window into the amount and type of collaboration. Furthermore, it confirms earlier work showing a progression, through a project lifecycle, of different types of communications [Swigger et al. 2012] and extends those findings to show implications for the speed with which students respond to one another. Finally, the study is a good example of how the combination of careful quantitative measurement and qualitative data from interviews can provide a richer understanding of collaborative work than either alone.

References

[Al-Ani et al. 2011] Al-Ani, B., Wilensky, H., Redmiles, D., Simmons, E.: An understanding of the role of trust in knowledge seeking and acceptance practices in distributed development teams. In: The 6th Int. Conf. on Global Softw. Eng., Helsinki, Finland, pp. 25–34 (2011)

[Chavan and Kaushik 2004] Kaushik, G., Chavan, A.L.: Collaborating on ethnography & design research: center for ethnography & contextual innovation at HFI. In: CHI 2004 Extended Abstracts on Human Factors in Computing Systems (CHI EA 2004), Vienna, Austria, pp. 1069–1170 (2004)

[D'Alessio et al. 2003] D'Alessio, M., Guarino, A., Pascalis, V., Zimbardo, P.G.: Testing zimbardo's stanford time perspective inventory (stpi)|short form. Time and Society 12(2/3), 333–347 (2003)

[Herbsleb et al. 2001] Herbsleb, J.D., Mockus, A., Finholt, T.A., Grinter, R.E.: An empirical study of global software development: distance and speed. In: Int. Conf. on Softw. Eng., pp. 81–90 (2001)

[Jick 1979] Jick, T.D.: Mixing qualitative and quantitative methods: triangulation in action. Administrative Science Q 24(4), 602–611 (1979)

[Kjeldskov and Paay 2012] Kjeldskov, J., Paay, J.: A longitudinal review of mobile HCI research methods. In: Proc 14th Int. Conf. Human-Computer Interaction, pp. 69–78 (2012)

[Mayhew 1999] Mayhew, D.: The usability engineering lifecycle: a practitioner's handbook for user interface design. Kaufmann, New York (1999)

[McInerney 2004] McInerney, D.M.: A discussion of future time perspective. Educational Psychology Review 16(2), 141–151 (2004)

[Milewski and Mullinix 2005] Milewski, A.E., Mullinix, B.: Solitary collaboration across cultures: a qualitative examination. In: Aykin, N., Peerce, J. (eds.) Internationalization, Online Communities and Social Computing: Design and Evaluation. Lawrence Erlbaum, New York (2005)

[Serce 2009] Serce, F.C., Alpaslan, F.N., Swigger, K., Brazile, R., Dafoulas, G., Lopez, V., Schumacker, R.: Exploring collaboration patterns among global software development. In: Proc 4th IEEE Int. Conf. on Global Softw. Eng., Limerick, Ireland, pp. 61–70 (2009)

[Swigger et al. 2010] Swigger, K., Brazile, R., Dafoulas, G., Serce, F.C., Alpaslan, F.N., Lopez, V.: Using content and text classification methods to characterize team performance. In: Int. Conf. on Global Softw. Eng., Princeton, pp. 192–200 (2010)

[Swigger et al. 2012] Swigger, K., Hoyt, M., Serce, C., Alpaslan, F.N., Lopez, V.: The temporal communication patterns of global software development student teams. Computers in Human Behav. 28(1), 384–392 (2012)

[Zimbardo and Boyd 1999] Zimbardo, P., Boyd, J.N.: Putting time in perspective: A valid, reliable individual-differences metric. J. Personality and Social Psychology 77(6), 1271–1288 (1999)

Part II

Health Monitoring and Disabled People Helping Systems

Intelligent Sensing and Learning for Assisted Living Applications

P. Augustyniak

AGH University of Science and Technology, Krakow, Poland
august@agh.edu.pl

Abstract. Achievements of today telemedicine and surveillance techniques offer exciting possibilities for developing intelligent assisted living systems for elderly or disabled. This chapter presents an universal sketch of multimodal health monitoring systems complying with regard to a paradigm of ubiquitous and personalized medicine. The proposed design combines advantages of context-reprogrammable sensors, flexibility of reconfigurable networks built on the surface of human body or embedded in premise infrastructures and automatic, subject-sensitive decision making based on presumptions and experience represented in artificial intelligence. Considering these key features leads to a concept, technical design and a prototype of a system suitable for majority of human surveillance purposes including home care, hospices, rehabilitation and sport training. The prototype was tested in several experimental setups intended to simulate volunteers' homes for seamless monitoring with no limit of indoor and outdoor mobility and for recognition of normal, suspected and dangerous events. The results confirm the expected adaptability of the system to human-dependent and environment-specific relations.

1 Introduction

Telemedical applications were proposed recently in vast numbers as a cost-effective alternative for treatment of diabetic, asthmatic or cardiac patients in home care conditions [Sun et al. 2009]. Current technological and methodological efforts in the latter domain aim at extraction of behavioral information focusing beyond medical purposes. The representation of motion, feelings or sleep in cardiovascular signals, makes the cardiac monitoring sensors useful also for behavior tracking [Augustyniak 2007]. This opportunity was already recognized and applied for fitness and sport training purposes. Some applications recently introduced to the market are useful for everyday monitoring of humans and yield pertinent data to estimate the subject's behavior.

In majority, commercially available e-health services make use of commonly available data sensors, telecommunication infrastructure, archiving and automated interpretation [Hristova et al. 2008]. Due to a closed-architecture-based design they feature very limited range of adaptation at sensor, infrastructure or decision-making levels. Such systems are tailored to an average patient, provide a patient-independent

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_13, © Springer International Publishing Switzerland 2014

diagnostic information and consequently are far from complying to the paradigm of personalized medicine.

Surveillance systems, in a common meaning of this term, are based on recognition and tracking of objects. In majority they are based on still or motion picture analysis (including code bars), detection of presence (e.g. radio-frequency identification tags) or interactions with static elements of the environment [Rougier et al. 2007; Liao and Yang 2008]. Since image-based systems are not performing well for detecting a human in unstable lighting conditions, they are often complemented with radio wave or ultrasound-based tactless detectors and with sensorized elements of the subject's environment.

Although a traditional meaning of surveillance is focused on objects, in our research we understand this term as a method for monitoring the human safety, with regard to internal (health-based) and external (environmental) risk factors. For this purpose, methods and experiences from medical diagnostics, particularly telemedicine, and from security-oriented technologies, such as object recognition and motion prediction were combined in one system. Consequently, our universal design dedicated to a wide target in ageing societies, consists of a network of facultative intelligent components and may be easily adapted by the software to particular needs.

The novelty of presented approach consists in a complementary application of wearable and infrastructure-embedded sensors providing information about selected diagnostic parameters, about subject's gesture and motion and about his or her interaction with the elements of surrounding infrastructure. The system uses different sensing modes and makes benefit of information redundancy supporting a seamless monitoring in case when selected data are not available (e.g. the subject is outdoor). The coinciding information is also used to identify the subject, allowing the system to support multiple subjects in a common area.

Since the proposed design is expected to combine features of telehealthcare and traditional object surveillance, it makes a simultaneous use of wearable and embedded sensors. Wearable sensors accompany each particular subject in his or her activity, thus each action he or she performs in a variable environment is represented in collected data set. Embedded sensors, being part of premise infrastructures capture parameters of given environment and the data represent its change caused by the presence of subjects. In the proposed concept, the wearable sensor network constitutes a subject-related independent system with own data interpretation, management and transmission gateways. If the subject is outdoor, the long range digital communication (e.g. GPRS) is used favoring wide availability of the carrier over the transmission bandwidth. If the subject is detected in the range of a wideband local area network (e.g. WiFi) of his or her home or office, the sub-network of wearable sensors is automatically appended to its infrastructure [Augustyniak 2011]. To maintain the integrity of wearable sensor network and correct target identification in case of multiple subject tracking, the communication of all wearable sensor-originated data passes through the wearable server.

Although the prototype specification was presented, implemented and tested for supervision of one subject in one premise, it may easily be extended to supervision of multiple subjects in a common area or even in multiple environments such as homes, offices, vehicles or supermarkets.

The set of wearable sensors is supervised by a wearable server through an ultralow range energy-efficient reconfigurable wireless network (Bluetooth class III). The network consists at least of two lead electrocardiogram front-end, an accelerometer-based motion sensor, a GPS receiver and a patient button. The wearable server is based on a selected smartphone featuring a built-in accelerometer and a GPS receiver and providing with two communication gateways (long distance: GPRS, HSDPA and short distance: WiFi). Depending on the patient needs, the wearable sub-network may be extended to up to eight nodes for support of other specific sensors (breath, SpO2, arterial pulse and blood pressure, single limb acceleration, glucose level etc.) [Otto et al. 2006].

The complementary set of embedded sensors is also organized in a local area network and supervised by a dedicated PC-based server providing also a short-range wireless interface (WiFi) to the subject-related wearable networks. The surveillance is based on video sequences providing information about the subject's presence, short distance mobility and safety. This data are complemented by detection of specific sounds (snoring, snorting, footsteps and falling [Mikrut and Smoleń 2011; Czopek 2011; Czopek 2012]). For these purposes the basic configuration of the embedded network include intelligent cameras and microphones. Additional information on how the subject is interacting with the environment may be acquired from smart sensors embedded in house equipment: taps, cooker, microwave oven, electric kettle, iron, chairs, toilet seat, bed etc.). These sensors belong to the embedded infrastructure of the building, however, for the purpose of commodity some of them use wireless interfaces.

2 General Assumptions about System Design

Dynamic configuration of the network, multimodal data acquisition and simultaneous support of different subject-specific detection rules require high flexibility of every system components. This can be achieved with programmable sensors and communication rules, adaptive detection of subject's action and intelligent classification of events as control- or health status-related. In order to make the adaptation independent from one aspect to another, services expected from the surveillance system were classified into three main categories: measurement, reconfiguration and decision making. Each category of services within the system constitutes a functional layer allowing for separate optimization of sensors, infrastructure and artificial intelligence [Augustyniak 2012]:

- The input layer consists of intelligent sensors acquiring the signals, performing the interpretation and providing a context-based semantic description of the subject.

- The infrastructure layer is based on a dynamic network of sensors using software-selectable architecture and data carriers; it optimizes the measurements and validates redundant data.
- The decision layer applies premise-dedicated rules and subject-derived behavioral habits to identify potentially dangerous events and initiate an adequate action.

Following the layered design, a single system accepts wide range of sensors, various indoor-outdoor setups and easily learns the subjects' habits. Premise rules or subject's habits may be easily translated to appropriate system architectures and/or to usage of different sensors for data acquisition. Another benefit from the layered design is easier management of data flow within the system (fig. 1) independently on the sensors, subjects or environments under supervision.

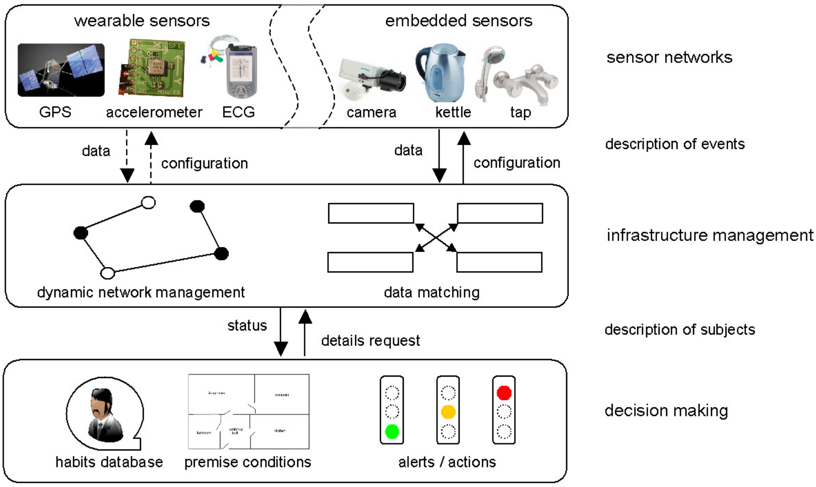


Fig. 1 Block diagram of the layered design and data flow in the assisted living system

2.1 Configuration and Networking of Sensors

The input to the sensor networks are lists of parameters expected for description of each subject. The output of the sensor networks is the information about events that was captured by the network. Each event contains its quantitative description as well as attributes of time and position. Sensors are expected to provide only the data requested by the infrastructure manager, accordingly to the surveillance goals and available communication resources. This approach assumes that in a regular reporting mode, basic processing of collected data is performed by the sensor which yields a semantic, low data volume description of the subject.

Since wearable networks are designed to collect data from a specific subject, the particular configuration depends on the features of each individual. For example, in some patient-dependent circumstances the ECG sensor can be expected to provide the heart rate only, while in some others a full disclosure ECG is necessary. The wearable sensors accompany the subject in every activity, so the data issued are automatically labeled with the appropriate identifier. As battery-operated devices, wearable sensors subject to a power economy-oriented design, aiming at compromise between light weight and reasonable autonomy time.

Sensors embedded in the premises' infrastructures constitute a complementary group and support simultaneous tracking of multiple subjects. Although any item of the house equipment may be interacted by different individuals, we assumed that each sensorized appliance is operated by a single person at a time, so related sensor data can be temporarily attributed to a particular subject. Some extent of mobility of embedded sensors results from their mode of use (e.g. electric kettle or iron). Consequently, despite of not following the subject, they require the use of wireless data interface. The most sophisticated type of embedded sensor are intelligent cameras. Consisting of a lens, image sensor and dedicated processor, these devices provide a semantic description of visible objects, including their unique ID, shape variations and position.

Wearable and embedded sensors play complementary roles in the surveillance system. Generally, the wearable sensors capture the active (voluntary or not) subject's behavior, while the embedded sensors record the environment with the subject as a passive participant. Reason-effect relations may be often revealed from cooperation of two networks:

- An action initiated by the subject (e.g. EMG-measured muscle contraction acquired by wearable sensor network) and its effect in the environment (e.g. displacement of the subject's body acquired by embedded sensor network),
- A change made in the environment (e.g. start of dog barking detected by embedded sensor network) and its influence to the subject's parameters (e.g. heart rate rise measured by wearable sensor network)

The content of sensors' report is individually programmed by the configuration service, and depends on the sensor type:

- Interaction sensors (tap, kettle, chair, etc.) return a binary value or single measurement value averaged in the report interval.
- Vibration sensors (accelerometers on the body or bed) return an averaged or peak amplitude or raw values from sequence of measurements made with programmed sampling frequency (e.g. 100Hz).
- Signal sensors (ECG, EMG-based motion, sounds) return processed results (respectively RR intervals, muscle activation level, brain activation features or detected acoustic effects) accordingly to the programmed processing level or raw sequences of measurement made with programmed sampling frequency (e.g. 250Hz for ECG).

- Visual sensors (cameras) return processed information about the objects, their features (size, shape variations, motion direction and speed) or raw sequence of color images complying with PAL video standard. The visual sensors are also recording the motion picture sequence in a memory loop buffer and provide raw visual information accompanying a particular event.

2.2 Infrastructure Management Services

Three services commonly described as infrastructure management are responsible for supervising of subjects, information flow and redundant data interpretation. The input of the infrastructure management is a list of subjects and their individual measurement requirements. The output is the information about the subjects, their positions and probability-ordered list of status. Accordingly to this task, the subjects are identified in the range of operation of embedded sensors (i.e. indoor), the sensors are activated with respect to actual information requirement and the network topology is set respectively.

The local area network of a premise supports multiple permanent nodes resposable for communication with embedded sensors and temporary nodes being gateways to personal servers of subjects' wearable networks. The detection of an indoor subject results in appending the wearable server to the infrastructure of embedded sensor network as a single node of given measurement features (fig. 2).

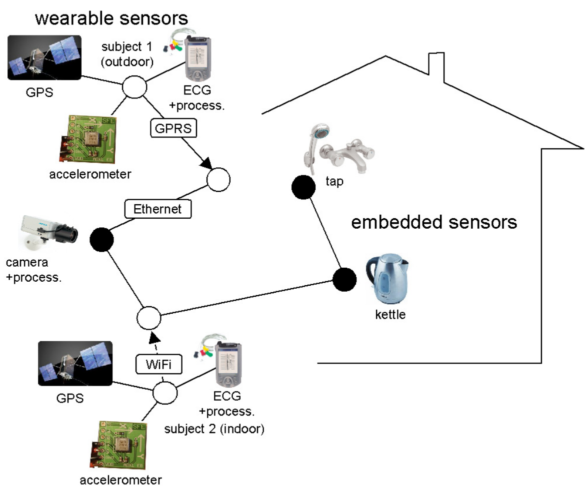


Fig. 2 Sensor types and network infrastructure

Regardless the position of the subject is indoor or outdoor, the data from wearable sensors are interpreted by the wearable server. This facilitates the identification of sensors in multiple subjects, unifies the communication rules of the wearable network and maintains data integrity when the subject moves outdoor [Smoleń et al. 2011]. The use of intelligent sensors and on-site data processing, significantly reduces the required network throughput, however occasional access to the raw data (e.g. ECG or video stream) may significantly raise the data flow.

Depending on the particular setup, data from different sensors can be correlated or redundant. Cross-matching of these data makes an opportunity for substitution of sensors, automatic data validation or common labeling of subjects. Since wearable sensor data are rather specific for the subject, while embedded sensors data are in general specific for the premise, the rules of interpreting the data representing a common event were defined as follows:

- The synchronicity of subjects' motion or sounds means the identity of subjects.
- Simultaneity of a signal from operating a sensorized appliance and corresponding visual information reveals the identity of acting subject.
- The correlation of motion data from wearable sensors and visual sensors helps to label the visible object by the subject's identifier and helps to predict the reliability of motion estimate when the subject is outdoor.

At the end of processing, the data from each subject are supplemented by information on subject's localization and interactions. Time series of such information structures for a particular subject are known as behavioral record.

2.3 Classification of Human Actions

The behavioral record provided by the surveillance infrastructure in result of consecutive measurements of selected vital parameters and interactions is a multimodal timeline log of subject's status. Classification of human actions as regular or possibly dangerous events requires referencing of current measurement to two databases:

- Premise-related database describing the purpose and topological connections between pieces in subject's premise [Ślusarczyk and Augustyniak 2010].
- Subject-related database describing usual behavioral patterns and their variants for each part of the day [Augustyniak et al. 2010].

The behavioral pattern is a statistically processed section of behavioral record including: probability-ordered list of current status, average and standard deviation values of expected duration, probability-ordered list of subsequent status and optional pointers to detailed subject-related sensor data.

For preserving the specificity of subjects’ actions, respective patterns are entered to the system database as examples recorded by the same system that is used for surveillance. Since recording of dangerous or critical patterns from the subject is not feasible, we record patterns of safe behavior available in almost all records of everyday life. The behavior classification module runs either in setup, learning and discrimination modes (fig. 3):

- In the setup mode, the human supervisor defines the premise-specific permission list, redundant sensors and sensors overlapping areas.
- In the learning mode the system records the subject's behavior and calculates statistics of behavioral patterns. In case the permission list or subject-related limits are exceeded, or in case the subject presses the button, detailed data from selected sensors are recorded for future review.
- In the discrimination mode the system records the subject's behavior, calculates statistics of behavioral patterns and performs programmed actions. For all events classified as ‘suspicious’ the detailed data from selected sensors are recorded.

The recorded detailed data are screened during the individual interview regularly appointed by the supervisor with the subject. The interview is an opportunity to manually adjust the system response to the individual needs.

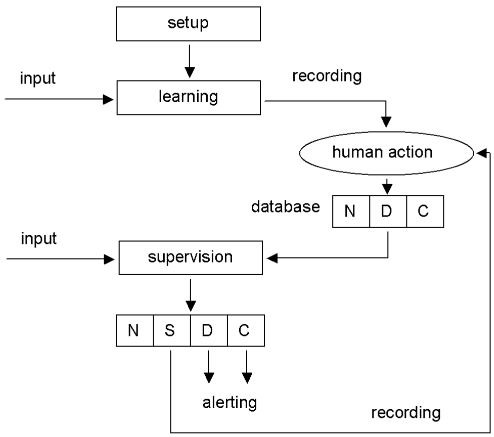


Fig. 3 Block diagram of the life cycle of the assisted living system

3 Implementation and Test Conditions

Basic tests of a prototype implementation of the assisted living system were focused on revealing possible problems with data transmission and synchronization, with management of dynamic infrastructure of the network and with the AI-based learning of subjects' behavior.

3.1 Implementation Details

The prototype wearable set of sensors consisted of two channels ECG, accelerometer, GPS and smartphone-based wearable server. The wearable network was supervised by a wearable server (LG-760 with Android OS) over a Bluetooth-based (IEEE 802.15.1 ver. 2.0, 2.1 Mbps, class II, 2.5 mW) piconet using a RFcomm-based protocol. The two channels ECG front-end (8 kbps) and three axes accelerometers (4.8 kbps) were nodes of the wearable network, while a GPS receiver and patient's button functionality were provided by the smartphone. The wearable server was connected as a client to the main network with the use of two alternative gateways: WiFi-based broadband (for indoor subjects) and GSM/GPRS-based (900MHz, device class 10 with up to four downlink slots and up to two uplink slots). The wearable server is running a circular data buffer compensating for up to 25 min of wireless data carier absence, dedicated procedures for limited-range processing of the ECG, for determination of the heart rate and quantitative motion parameters.

We also simulated the human living environment supervised by a set of embedded sensors consisting of: two cameras, a microphone matrix and two potentiometer-based sensors (water tap and electric kettle). Each intelligent camera was simulated by a regular camera connected to a dedicated PC for the reason of faster adjustment of image data processing routines. Each device runs a custom-programmed real time application based on the optical flow method and calculates centers, edges, size and shape variations of objects. For the reason of huge information volume in the test mode, the prototype intelligent cameras were connected to the AP with Ethernet cables. The microphone matrix was mounted in the bedroom and also connected to a dedicated PC for detection of snoring and fall sounds.

The embedded sensors were connected to a Ethernet-based network and WiFi (IEEE 802.11b, 2.4 GHz, 11 Mbps). The network was supervised by a SUSE Linux-based PC with a static IP, running also the infrastructure and the decision making software. The management and the AI parts were custom-programmed in C++, while the database and network parts were based on PHP and MySQL technologies.

3.2 Testing Conditions

A two piece apartment was arranged in the laboratory consisting of:

- A 'bedroom' with a sofa bed, a camera (with infrared lighting) and microphones (this setup was also used for sleep studies [Smoleń et al. 2010]).
- A 'kitchen' with two sensorized appliances (tap and electric kettle), and a camera.

This arrangement was attributed with specific set of rules: for example no lying was allowed out of the bed, the kettle could be removed from the stand for no longer than 1 minute, the tap could remain opened only when the subject is close to it etc.

Four young volunteers tested the prototype system by performing actions as: lying, getting up, sitting, walking and working, organized in sessions accordingly to 17 schedules differing by state order and duration. Each session lasted for ca. 20 minutes and aimed to simulate an excerpt of everyday life. Some schedules contained events violating the permission list of the apartment, a subset of schedules included also outdoor activities as walk, short run and rest. Behavioral records were collected continuously, the system was expected to recognize the subject's status and - with respect to its temporal changes - to classify each behavioral pattern into one of four possible categories: {normal, suspicious, dangerous, critical}.

4 Results of Tests

The subject's status detection was performed with the accuracy ranging from 89.1 % (for outdoor working) to 96.8 % (for outdoor walking) - in both cases detection was based on accelerometers.

The WiFi-based data transfer was made with no measurable delay, whereas the GPRS-based connection delay ranged from 3.1 to 17.1 s (worst case) [Augustyniak 2010]. The success ratio for the gateway switching was 93 % for switching from GPRS to WiFi, and only 73 % in the opposite direction. In conditions of our experiment the subject kept losing the long-range carrier at the entrance to the building, and needed 55-75 s until he or she reached a WiFi-enabled premise.

The identification of body position based solely on visual methods using Manhattan metric ranged from 9.6% (for bending pose) to 97.6% [Mikrut et al. 2012]. Identification of subjects based on a correlation of data from different recording modes was found reliable: the average value of motion signal correlation from the same subject was 91 % and only 68 % from two different subjects. In all cases the highest correlation was achieved between the data from the same subject.

5 Conclusions

This paper proposes general rules of design of open-architecture systems for health- and behavior-based surveillance of human. The system was prototyped and tested in one-to-one network cooperation model, but after slight modification may also be useful for surveillance of several people in various environments.

The presented system may easily be adapted by the software to meet particular needs of a supervised individual, nevertheless it addresses a wide range of population. The scale effect raises opportunities for commercialization and due to the use of state-of-the-art technology, the presented architecture for assisted living is a cost-effective, evidence-based and reliable alternative for direct care provided everyday by nurses or volunteers in hospitals, hospices and nursing homes.

Presented generalized design rules for assisted living systems take in consideration multiple personalized wearable and residential parts, conditionally cooperating in the surveillance of subjects. The system supports the transfer of wide range of human description data in variable connectivity conditions. As verified by the design of prototype system, the design rules are complete and consistent.

In the near future the prototype system will be tested as an alerting tool for assistants of welfare aid providing care to the ill or elderly. This stage of territorially limited pilot implementation is planned to highlight any social and organizational problems with the monitoring and to demonstrate its potential for improving the quality of life of an ageing society.

Acknowledgment. The author expresses his gratitude to A. Broniec, K. Czopek, M. Smoleń, E. Kańtoch, Z. Mikrut and P. Pleciak for their cooperation and pertinent remarks on each stage of the presented work. This Scientific work is supported by the AGH University of Science and Technology in year 2013 as a research project No. 11.11.120.612.

References

[Augustyniak 2007] Augustyniak, P.: Validation of automatic ECG processing management in adaptive distributed surveillance system. In: Kurzyński, M., et al. (eds.) Computer Recognition Systems 2, pp. 574–580. Springer, Heidelberg (2007)

[Augustyniak 2010] Augustyniak, P.: Personalized building-embedded and wearable monitoring infrastructures with mediated data transfer. In: Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 2121–2127 (2010)

[Augustyniak et al. 2010] Augustyniak, P., Smoleń, M., Broniec, A., Chodak, J.: Data integration in multimodal home care surveillance and communication system. In: Piętka, E., Kawa, J. (eds.) Information Technologies in Biomedicine. AISC, vol. 69, pp. 391–402. Springer, Heidelberg (2010)

[Augustyniak 2011] Augustyniak, P.: Compound Personal and residential infrastructure for ubiquitous health supervision. In: Hippe, Z.S., Kulikowski, J.L., Mroczek, T. (eds.) Human – Computer Systems Interaction. AISC, vol. 98, pp. 523–538. Springer, Heidelberg (2012)

[Augustyniak 2012] Augustyniak, P.: Layered design of an assisted living system for disabled. In: Piętka, E., Kawa, J. (eds.) ITIB 2012. LNCS, vol. 7339, pp. 498–509. Springer, Heidelberg (2012)

[Czopek 2011] Czopek, K.: Evaluation of breathing dynamics using the correlation of acoustic and ECG signals. Computing in Cardiology 38, 33–36 (2011)

[Czopek 2012] Czopek, K.: Cardiac activity based on acoustic signal properties. Acta Physica Polonica A, Acoustic and Biomedical Engineering 121(1-A), 42–45 (2012)

[Hristova et al. 2008] Hristova, A., Bernardos, A.M., Casar, J.R.: Context-aware services for ambient assisted living: A case-study. In: First International Symposium on Applied Sciences on Biomedical and Communication Technologies, ISABEL 2008, pp. 1–5 (2008)

[Otto et al. 2006] Otto, C., et al.: System architecture of a wireless body area sensor network for ubiquitous health monitoring. Journal of Mobile Multimedia 1(4), 307–326 (2006)

[Liao and Yang 2008] Liao, W.H., Yang, C.M.: Video-based activity and movement pattern analysis in overnight sleep studies. In: Proc. 21st ICPR, Pattern Recognition (2008)

[Mikrut and Smoleń 2011] Mikrut, Z., Smoleń, M.: A neural network approach to recognition of the selected human motion patterns. Automatyka 15(3), 535–543 (2011)

[Mikrut et al. 2012] Mikrut, Z., Pleciak, P., Smoleń, M.: Combining pattern matching and optical flow methods in home care vision system. In: Piętka, E., Kawa, J. (eds.) ITIB 2012. LNCS, vol. 7339, pp. 537–548. Springer, Heidelberg (2012)

[Rougier et al. 2007] Rougier, C., Meunier, J., St-Arnaud, A., Rousseau, J.: Fall detection from human shape and motion history using video surveillance. In: Proc. 21st International Conference on Advanced Information Networking and Applications Workshops, pp. 875–880 (2007)

[Smoleń et al. 2010] Smoleń, M., Czopek, K., Augustyniak, P.: Sleep evaluation device for home-care. In: Piętka, E., Kawa, J. (eds.) Information Technologies in Biomedicine. AISC, vol. 69, pp. 367–378. Springer, Heidelberg (2010)

[Smoleń et al. 2011] Smoleń, M., Kańtoch, E., Augustyniak, P., Kowalski, P.: Wearable patient home monitoring based on ECG and ACC sensors. In: Jobbágy, Á. (ed.) 5th European IFMBE Conference. IFMBE Proceedings, vol. 37, pp. 941–944. Springer, Heidelberg (2011)

[Sun et al. 2009] Sun, H., De Florio, V., Gui, N., Blondia, C.: Promises and challenges of ambient assisted living systems. In: Proc. 6th Inter. Conf. on Information Technology: New Generations, pp. 1201–1207 (2009)

[Ślusarczyk and Augustyniak 2010] Ślusarczyk, G., Augustyniak, P.: A graph representation of subject's time-state space. In: Piętka, E., Kawa, J. (eds.) Information Technologies in Biomedicine. AISC, vol. 69, pp. 379–390. Springer, Heidelberg (2010)

Designing and Evaluating Online Telehealth Systems for Seniors

J.S. Dhillon[1], B.C. Wünsche[2], and C. Lutteroth[2]

[1] Department of Graphics and Multimedia, Universiti Tenaga Nasional,
43000 Kajang, Selangor, Malaysia
jaspaljeet@uniten.edu.my
[2] Department of Computer Science, University of Auckland,
Private Bag 92019, Auckland, New Zealand
{burkhard,lutteroth}@cs.auckland.ac.nz

Abstract. The increasing cost of senior healthcare represents a serious challenge to most developed countries. Telehealth has been widely promoted as a technology to make healthcare more effective and affordable, however, current telehealth systems suffer from vendor lock-in and high cost, and are designed for managing chronic diseases rather than preventing them. We propose a novel framework for telehealth systems aimed at overcoming these shortcomings through principles such as extensibility, accessibility, social support and motivational feedback. The framework was employed to develop a patient-centric telehealth system, *Healthcare4Life*, which was evaluated in a user study to draw conclusions about the feasibility and acceptance of such a system by seniors. The results indicate that seniors are capable and motivated to use a web-based telehealth system if it provides suitable health applications. Furthermore, the results suggest that such a system has the potential to affect seniors' attitude towards their own health positively.

1 Introduction

In many developed countries healthcare systems are struggling with an increasing number of seniors, more chronic diseases affecting them, a declining potential support ratio, mounting senior healthcare costs, and a shortage of health professionals. The three main emerging public health issues are: providing access to affordable healthcare, solving chronic health problems and preventing diseases [Edelman 2009]. One promising approach to address these issues is to empower health consumers to better manage and monitor their health [Hibbard and Cunningham 2008]. If patients play an active role in preventing diseases instead of being passive recipients of treatments after getting sick, they are more able to take control of their health.

Home-based healthcare applications such as telehealth can enable users to track their health status and to actively participate in treatment regimens and preventive strategies. Telehealth can be defined as [HRSA 2013]: "*the use of electronic information and telecommunications technologies to support long-distance clinical*

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_14, © Springer International Publishing Switzerland 2014

healthcare, patient and professional health-related education, public health, and health administration." From the definition it is clear that telehealth is about receiving care from a distance. Research suggests that seniors generally prefer to age independently in their own homes rather than moving to an institutional care setting [Botsis et al. 2008]. Telehealth meets this preference and could therefore be a viable approach for helping seniors to take charge of their health. However, in order to be successful, telehealth systems need to be widely available, affordable and extendable. Most importantly, their design should reflect the needs of the "patients", rather than those of clinicians.

Internet-based solutions are capable of reaching a wider section of the senior population at a low cost, with promising interventions that could affect desired lifestyle changes. Seniors are often said to experience anxiety and usability issues when dealing with technology, but the current trend is changing this. The number of seniors going online is increasing steadily. Recent Internet demographics suggest that more than half (53%) of U.S. seniors are online today [Zickuhr and Madden 2012]. The continued growth in the number of seniors going online indicates an increased utilisation of web-based resources, thereby providing opportunities to address the shortcomings of current telehealth systems by making them more accessible, add social aspects and stimulate integration of third-party applications. Therefore, we choose an approach that leverages the Web as a medium to deliver healthcare services to seniors.

The principle aim of the study presented here is to develop and test new concepts to address the shortcomings of current telehealth systems and Web 2.0 health applications. The intention is to contribute to a novel design for web-based telehealth systems that is closely aligned with consumers' preferences, i.e. telehealth from the perspective of the senior consumer. We propose a framework for an accessible patient-centric telehealth system with social aspects, develop a prototype implementing key concepts, and explore these concepts using the prototype. A patient-centric approach was employed in developing the prototype by working closely with seniors from the outset. Several user studies were conducted with representatives of this target user population.

First, we describe the shortcomings of current health informatics applications, building an understanding of the extent they fulfill the healthcare needs of users. Based on this we propose a novel framework for web-based, patient-centric, affordable and extendable telehealth systems and describe the key design decisions that were applied to create the Healthcare4Life (HC4L) system prototype. Our implementation uses OpenSocial and Drupal to provide social networking functionalities and a Facebook-like plug-in architecture for third-party content [Dhillon et al. 2012]. A six-week user study indicates that seniors welcome the opportunity to become more proactive in managing their health, and that a system such as HC4L can empower users by giving them more control over managing their health.

Section 2 describes requirements and motivates our work by summarising the shortfalls of current consumer health informatics applications. Section 3 presents the proposed framework for the design of patient-centric telehealth systems.

Section 4 describes the evaluation of the HC4L prototype, and Section 5 concludes the paper with a summary of our contributions and future work.

2 Establishing Requirements

This section first provides an overview of current consumer health informatics applications to motivate the development of a novel web-based telehealth system. Then, we describe a small-scale interview study conducted to identify the user interface and functional requirements of a web-based telehealth system.

2.1 Consumer Health Informatics Applications

In recent years an increasing number of healthcare applications has been developed, including consumer health informatics applications designed to interact directly with consumers, with or without the presence of healthcare professionals. In our previous work [Singh et al. 2010], we evaluated common consumer health informatics applications (i.e. telehealth systems, health record management systems, Web 2.0 health applications, serious games and exertainment applications) from the patient's perspective and discussed their strengths and weaknesses for healthcare. Our analysis identifies a number of shortcomings constraining widespread use and health outcomes, as summarised below.

Commercial telehealth applications and most health record management systems are centred on clinical users, health service providers and vendors' interest of generating a continuous revenue stream. Generally such applications perform well in collecting, analysing and monitoring health data, but there is little support for patients to change their lifestyle positively. Most of these applications are expensive and suffer from vendor lock-in, making it difficult to add new content. They usually do not fit into the regular activities of a user, making them disruptive. Most telehealth applications are designed for monitoring and treating patients with chronic diseases, but do little to prevent them.

Health information websites offer an impressive range of information, but it can be difficult to assess its reliability, meaning and implications. Web-based discussion and support groups can provide a more personal experience and add a social factor that can help patients with coping and commencing positive lifestyle changes. Web 2.0 health applications and services (also known as Health 2.0) are rapidly gaining attention from patients and professionals as they extend traditional healthcare delivery models, facilitate patient self-care and provide social support. However, most of these applications are fairly costly, limited in their functionality, and target younger health consumers [Dhillon et al. 2011a].

Serious games and exertainment applications are arguably the most patient-centric consumer health informatics applications. However, evaluations of their effectiveness report mixed results, especially for long-term use. One of the main problems is their limited content, which means applications can become repetitive and boring, so that lifestyle changes are only temporary [Owens et al. 2011]. In most cases content is controlled by a single vendor and must be purchased. Furthermore, monitoring data is not shared between different games, so that a continuous recording of health parameters and activities is not possible.

In order to promote a more widespread use of health informatics applications and achieve better health outcomes, applications need to offer a wide range of content and functionalities, there needs to be a quality control of content, and different applications must be integrated to deliver a more complete picture of health over time. Furthermore, application should be low-cost (preferably free), easy-to-use, concentrate on the prevention of diseases, and provide motivation and support. The latter suggests the incorporation of social networking features. Such features can also help to reduce loneliness, which has been shown to result in serious health problems [Miller 2011].

2.2 Interview Study

In order to identify requirements for a more accessible patient-centric telehealth system, we performed an interview study with eight potential senior users [Dhillon et al. 2011b]. The study was designed as a qualitative inquiry focusing on seniors' perceptions and preferences. First, a semi-structured interview was conducted. Then, a paper-based prototype of HC4L and screenshots of different existing health informatics applications were presented for evaluation.

The results suggest that the Internet is a suitable platform for the delivery of telehealth applications. In the interviews, the participants suggested several applications such as exercises for different health problems, diet control, and simple network games. They generally favoured user interfaces with a simple layout such as a single horizontal menu at the top, making it easy to identify and choose key functionalities. Buttons with suitable icons or text were preferred over hyperlinks. The results indicate that social support through Facebook-like features can be useful in a community of like-minded users. Social interactions may help users to reduce loneliness, motivate each other, and share experiences.

3 Conceptual Framework

In this section, we describe a novel conceptual framework and key design principles to aid development of patient-centric telehealth systems. The framework presents the big picture (i.e. the overall vision including possible future work) for overcoming the challenges of existing consumer health informatics applications in empowering health consumers to take charge of their healthcare.

3.1 Framework

Based on the above requirements we developed a framework to aid development of patient-centric telehealth systems, illustrated in Fig. 1, which is accessible, extendable by third parties, contains social aspects, encourages cognitive engagement, and aims to put the user in control [Singh et al. 2010]. The framework has an open Facebook-like architecture enabling third-party developers to contribute new content and functionalities. Examples are applications for monitoring, health information, mental fitness and education. Physical fitness and rehabilitation can be supported by adding applications using consumer-level sensing devices [Dhillon et al. 2011a]. Accessibility is facilitated by making the system web-based, so that it can be accessed on desktop computers, tablets and mobile phones. In order to assess the quality of content we propose to employ a ranking system displaying user satisfaction and popularity of each service. The ranking system should contain separate scores from patients and registered clinical/academic users.

In order to share data between different applications, unifying data elements must be introduced that relate different types of data. This can be achieved by using a triplestore database: data entities are composed of subject-predicate-object triples, where the predicate represents the unifying element. For example, different exercise games might use different measures to record users' physical activities, which could be unified using a "calories burned" or "perceived-level-of exertion" scale [Borg 1982] similar to those used in gym equipment. The unified data can then be used by monitoring applications to enable users to design activity plans and track progress.

While commercial telehealth systems put an emphasis on clinical networks, we utilise social networks to help users get in touch with their family, make new friends, and discuss medical complaints with peers and support groups. The aim is to improve emotional health, which is essential for the overall wellbeing. Social networks can also help with motivation, e.g., through family support or competing/exercising together via a video link or in a virtual environment.

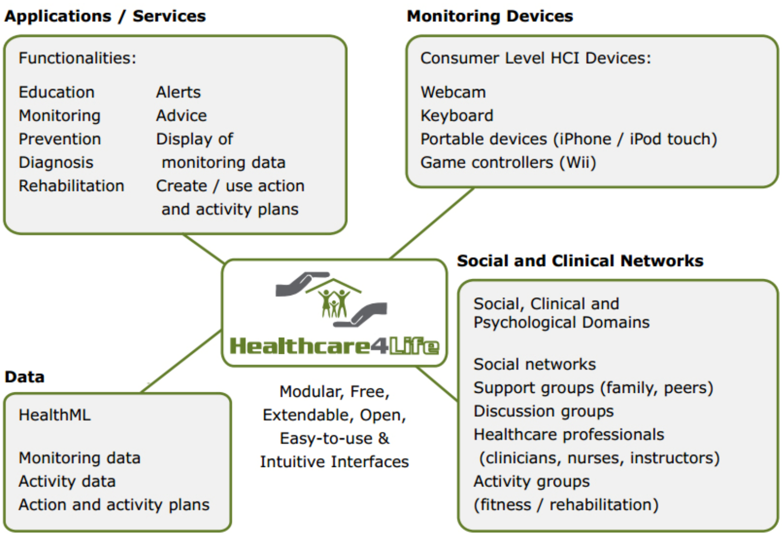


Fig. 1 Conceptual framework for HC4L [Singh et al. 2010]

3.2 Key Design Principles

The above analysis of consumer health informatics applications and the interview study resulted in the following key design principles (ordered by importance):

- **Open and extensible:** The system should offer a wide variety of health-related applications that cater for the various needs and preferences of patients. If the available content is too limited, it will become repetitive and boring for users and result only in short-term lifestyle changes. The system should be "open", so that third parties can contribute and adopters can avoid vendor lock-in.
- **Accessible and affordable:** The system must be accessible with a common web browser on different devices, at low cost or free of charge.
- **Social and emotional support:** The system should foster a caring community, enabling users to provide and obtain social support to and from other users. Social features should be incorporated to reduce loneliness, e.g., help users to become acquainted with other users, to perform health-related activities together (e.g., playing a memory game), and to motivate each other.
- **Feedback and motivation:** The system should provide users with feedback on their health progress and motivate them to become more proactive, e.g., make positive lifestyle changes. Visual feedback via easy-to-understand graphs and charts is particularly important for users with limited health literacy.
- **Privacy control:** User privacy must be maintained and users must have control over their personal data. Sensitive data such as those relating to symptoms, diagnoses and treatments should not be disclosed without explicit consent.
- **Personalised user interface:** Users should be able to customise the system with applications tailored towards their personal needs. Applications should be easy to add and remove. Sufficient information about available applications should be provided to enable users to select desired functionalities.
- **Simple navigation structure and clear instructions:** The system should provide a simple navigation structure to avoid confusion and frustration. For example, the system should allow users to complete health related tasks such as tracking their weight using a linear sequence of simple steps.

4 Evaluation

We evaluated the usability, acceptance and feasibility of the most important framework concepts in a user study involving 43 seniors (16 male, 27 female) aged 60 to 85. In the following, we describe the methodology used in the evaluation and present and discuss the results.

4.1 Methodology

For the study, a prototypical implementation of the framework, Healthcare for Life (HC4L), was created, based on the key design principles laid out in Section 3.2.

It was implemented using OpenSocial and Drupal. OpenSocial was preferred over Facebook and other single-platform APIs because it does not constrain developers with vendor restrictions and usage policies [Dhillon et al. 2012]. Developers have full control over their system and the freedom to integrate it with other OpenSocial containers. The Drupal CMS was leveraged for content management because of its support for OpenSocial.

HC4L supports both application users and application developers. Both groups are presented with distinctive functionalities based on their role in the system. System developers can add applications to HC4L, so that they become accessible to users via an application directory. Patients are provided with the following functionalities (see centre of Fig. 2): *Activities* to motivate friends with positive comments and to share their status; *Health Apps* to access applications added by developers; *Profile* to create a basic online health profile to enable others to locate them in the system; *Mail* to send mails to other users that are listed as “friends”; *Friends* to access the profile pages of “friends”, find and add new friends, and invite others to join HC4L; and *Settings* to make changes to general system settings such as privacy settings.

The following applications were created and made accessible via HC4L (see Fig. 2). A *Weight Tracker* allows users to enter their current weight at any time and track their weight over time using a configurable graph visualisation. A multiplayer *Memory Game* fosters interaction between users, allowing them to practice their memory together (in collaboration or as a competition) in a “Matching Pairs” game with a configurable level of difficulty. An *Exercise Tracker* allows users to track their physical activity over time by entering the time spent on different activities (e.g. walking, housework and swimming), measuring the total physical activity in terms of energy expenditure (i.e., calories burned). A *Vitals Tracker* allows users to enter vital sign measurements (systolic/diastolic blood pressure, resting heart rate and blood glucose) using a simple tabular interface, and track them over time using a configurable graph visualisation. The *Calorie Calculator* is an existing diet tracking application from LabPixies.com. Most of the tracking applications give visual feedback about the normal range of health parameters.

Participants were invited to use HC4L over a six-week period. They were asked to complete three questionnaires at different stages of the study: at the start of the study (initial questionnaire), at the end of the third week (interim questionnaire), and at the end of the sixth week (final questionnaire). Furthermore, a short interview was conducted with four selected participants at the end of the study to gain further insights into their experiences with and perceptions of HC4L.

The questionnaires incorporated existing established scales: the Multidimensional Health Locus of Control (MHLC) [Wallston et al. 1978], the Intrinsic Motivation Inventory (IMI) [Ryan 1982], and the System Usability Scale (SUS) [Bangor 2009]. MHLC was used to investigate whether HC4L can positively affect the users’ attitude towards managing their health. IMI was used to evaluate users’ subjective experience, i.e., their intrinsic motivation, in their interaction with HC4L. SUS was used to assess the overall usability of the system. In order to keep the questionnaires simple for seniors, shortened forms of these scales were used,

Fig. 2 Screenshots of health support applications embedded in HC4L

which were known to be equally effective. In order to identify usability issues and evaluate the usefulness of the different functionalities, we used additional, system-specific questions with responses rated on a 6-point Likert-scale.

4.2 Results and Discussion

The majority of the participants were active computer users (88%) using a computer almost every day. Less than half of them (44%) used social networking websites such as Facebook. Less than one third used self-care tools, e.g., blood pressure cuffs, glucometers or health websites.

Fig. 3 depicts the popularity of the various functionalities of HC4L over the trial period. The health applications were most popular among the participants (35% of all activities performed). This demonstrates the importance of having a plug-in feature for increasing the amount of available content. The Facebook-like comment page termed "Activities" was the second-most commonly used feature (22%), followed by the "Friends" page (17%). This indicates that users are interested in sharing experiences and making social connections. Among the applications, the Vital Tracker was most frequently used (29%), followed by the Exercise Tracker (28%), and the Weight Tracker (22%). The Calorie Calculator was least frequently used (8%) because of its comparably complex user interface, the time required to enter data, and because many food items were not relevant for a New Zealand context. This observation emphasizes the need for customising applications according to the users' location, culture and language.

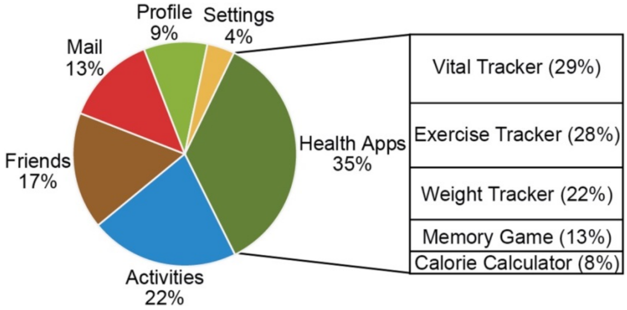


Fig. 3 Participants' activities in HC4L

Table 1 reports the mean score changes for those participants who completed both initial and interim MHLC questionnaires. Score changes for each MHLC subscale were calculated by subtracting baseline scores from follow-up scores. Paired t-tests were used to compare the MHLC mean scores. The findings show that there were some improvements on all the three subscales, although statistically not significant. There was a noticeable difference in the responses for "Powerful Others", which contained three statements to assess whether participants believed their health is controlled by others, e.g. health professionals. The responses to one of the statements, Health professionals control my health, showed a significant difference (t=2.343, p=0.03), indicating that users obtained a better understanding of their health and how it is affected by their own decisions. Although the results were overall not significant, they suggest that HC4L can affect its users positively. With a richer and more mature implementation, a larger sample and a longer study, we would expect stronger and more significant effects.

Table 1 Change in responses to the MHLC subscales within a three-week period (n = 23)

Subscale	Mean Change	Change SD	Change Range	t	p
Internal	.04	1.04	-4 to 2	-0.44	0.66
Powerful Others	-.29	1.27	-10 to 6	1.21	0.24
Chance	-.10	1.23	-6 to 5	0.51	0.61

Table 2 presents the mean values and standard deviations of five pre-selected subscales of the IMI (range 1 to 7), measured after week 3. It also illustrates the scores of two different senior age groups. Except for the "pressure/tension" scale, the results show mid scores in the range 4.11 to 4.40. The results suggest that the participants were fairly interested in the system, felt adequately competent, made a reasonable effort in using the system, and felt that the system has some value or utility for them. The "pressure/tension" subscale obtained a low score indicating that the participants did not experience stress while using the system. There were notable differences between age groups for "perceived competence" and "value / usefulness". Seniors aged 60-69 considered themselves more competent and found the system more valuable than older seniors, suggesting higher computer literacy.

Table 2 IMI results from 1 (strongly disagree) to 7 (strongly agree), mean ± std. deviation

Subscale	All (n = 24)	Age 60-69 (n = 12)	Age 70-85 (n = 12)
Interest/Enjoyment	4.40 ± 1.68	4.42 ± 1.73	4.39 ± 1.70
Perceived Competence	4.39 ± 1.78	4.89 ± 1.52	3.89 ± 1.94
Effort/Importance	4.11 ± 1.58	4.11 ± 1.57	4.11 ± 1.56
Pressure/Tension	2.61 ± 1.56	2.67 ± 1.45	2.56 ± 1.69
Value/Usefulness	4.25 ± 1.81	4.53 ± 1.83	3.97 ± 1.75

At the end of week 3, 24 participants rated the usability of the system on the SUS scale, with scores ranging from 35 to 100, a median of 65 and an average of 68.33. Only two responses were below 50. The adjective rating of the mean SUS score is "OK", which indicates HC4L is an acceptable system [Bangor 2009]. Participants' open-ended responses were useful in gaining insight into their perceptions of HC4L. Table 3 lists the most frequent responses.

Table 3 Most common positive and negative comments about HC4L (n=24)

Positive Responses	Frequency
I like the idea of it.	26%
It is easy to use.	23%
The health applications are a great help to keep track of one's health.	16%
Negative Responses	
Sorting out calories values for foods seems a lot of trouble (Calorie Calculator).	21%
I am not so keen on the social Facebook-like aspects of the system.	18%
The range of available applications is limited.	15%

In the final questionnaire 13 participants stated that HC4L encourages them to be better aware of their health (80% agreement), that the charts/graphs help to better understand health progress (80% agreement), and that the health applications reduce the need to use different websites for managing health (72% agreement). 65% agreed that HC4L has the potential to positively impact their life, and 56% found that HC4L's applications simplify cumbersome health monitoring tasks.

However, only 33% agreed that the social features motivated them to use the system, and only 31% agreed that the involvement of friends helped them to better manage their health. Four participants of the study expressed disappointment that their friend requests were not responded to. Most of the participants were not comfortable accepting strangers as "friends" in HC4L. A typical comment was: "*I would not share my medical details with someone I don't know*". One participant elaborated on this: "*I find the use of the word 'friends' for people I don't know and will never meet very inappropriate and off-putting. Also it's really important to learn more about the people in your circle so that you care enough about them and their goals to be able to offer support.*"

5 Conclusions and Future Work

We presented a framework for web-based telehealth systems targeted at seniors that are extendable, accessible and include social networking functionality. The framework was used to implement a prototype system, HC4L, which was evaluated in a study involving 43 seniors. The results indicate that most seniors are motivated to use such a system if sufficient and adequate health applications are provided. Furthermore, the results suggest that web-based telehealth systems have the potential to change the attitude of seniors towards their own health positively.

As a future work, systems such as HC4L should be evaluated with a wider and more mature range of health applications, which should ideally be designed in collaboration with health professionals. Furthermore, the long-term effects of such systems need to be explored in a long-term study with a larger sample of seniors.

References

[Bangor 2009] Bangor, A., Kortum, P., Miller, J.: Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. Journal of Usability Studies, 114–123 (2009)

[Borg 1982] Borg, G.: Psychophysical bases of perceived exertion. Medicine & Science in Sports & Exercise 14(5), 377–381 (1982)

[Botsis et al. 2008] Botsis, T., Demiris, G., Pedersen, S., Hartvigsen, G.: Home telecare technologies for the elderly. Journal of Telemedicine and Telecare 14(7), 333–337 (2008)

[Dhillon et al. 2011a] Dhillon, J.S., Wünsche, B.C., Lutteroth, C.: Leveraging Web 2.0 and Consumer Devices for Improving Elderlies' Health. In: Australasian Workshop on Health Informatics and Knowledge Management, Perth, Australia, pp. 17–24 (2011)

[Dhillon et al. 2011b] Dhillon, J.S., Ramos, C., Wunsche, B.C., Lutteroth, C.: Designing a web-based telehealth system for elderly people: An interview study in New Zealand. In: Proceedings of the 24th International Symposium on Computer-Based Medical Systems (CBMS), Washington, DC, USA, pp. 1–6 (2011)

[Dhillon et al. 2012] Dhillon, J.S., Ramos, C., Wunsche, B.C., Lutteroth, C.: Evaluation of web 2.0 technologies for developing online telehealth systems. In: Australasian Workshop on Health Informatics and Knowledge Management, Melbourne, Australia, vol. 129, pp. 21–30 (2012)

[Edelman 2009] Edelman: Health engagement barometer-health influence in the era of public engagement (2009), `http://healthbarometer.edelman.com/` (accessed October 16, 2013)

[Hibbard and Cunningham 2008] Hibbard, J.H., Cunningham, P.J.: How engaged are consumers in their health and health care, and why does it matter? (2008), `http://hschange.org/CONTENT/1019/` (accessed October 16, 2013)

[HRSA 2013] Health Resources and Services Administration, Telehealth. U.S. Department of Health and Human Services (2013), http://www.hrsa.gov/ruralhealth/about/telehealth/ (accessed October 4, 2013)

[Miller 2011] Miller, G.: Why loneliness is hazardous to your health. Science 331(6014), 138–140 (2011)

[Owens et al. 2011] Owens, S.G., Garner, J.C., Loftin, J.M., van Blerk, N., Ermin, K.: Changes in physical activity and fitness after 3 months of home Wii Fit use. Journal of Strength and Conditioning Research 25(11), 3191–3197 (2011)

[Ryan 1982] Ryan, R.M.: Control and information in the intrapersonal sphere: an extension of cognitive evaluation theory. Journal of Personality and Social Psychology 43(3), 450–461 (1982)

[Singh et al. 2010] Singh, J., Wünsche, B.C., Lutteroth, C.: Framework for Healthcare4Life: a ubiquitous patient-centric telehealth system. In: Proceedings of the 11th International Conference of the New Zealand Chapter of the ACM Special Interest Group on Human-Computer Interaction (CHINZ), pp. 41–48 (2010)

[Wallston et al. 1978] Wallston, K.A., Wallston, B.S., DeVellis, R.: Development of the Multidimensional Health Locus of Control (MHLC) Scales. Health Education & Behavior 6(2), 160–170 (1978)

[Zickuhr and Madden 2012] Zickuhr, K., Madden, M.: Older adults and Internet use. Pew Research Center, Washington (2012), http://www.pewinternet.org/Reports/2012/Older-adults-and-internet-use.aspx (accessed September 23, 2013)

Sleep Apnea Detection by Means of Analyzing Electrocardiographic Signal

P. Przystup, A. Bujnowski, A. Poliński, J. Rumiński, and J. Wtorek

Department of Biomedical Engineering, Faculty of Electronics,
Telecommunications and Informatics, Gdansk University of Technology, Gdańsk, Poland
piotr.przystup@gmail.com,
{bujnows,apoli,jwr,jaolel}@biomed.eti.pg.gda.pl

Abstract. Obstructive sleep apnea (OSA) is a condition of cyclic, periodic obstruction (stenosis) of the upper respiratory tract. OSA could be associated with serious cardiovascular problems, such as hypertension, arrhythmias, hearth failure or peripheral vascular disease. Understanding the way of connection between OSA and cardiovascular diseases is important to choose proper treatment strategy. In this paper, we present a method for integrated measurements of biosignals for automatic OSA detection. The proposed method was implemented using a portable device with the application of the Support Vector Machine (SVM) classifier. The specific objective of this work is to analyze the minimum set of features for the ECG signal that could produce acceptable classification results. Those features can be further expanded using other biosignals, measured by the portable SleAp device. Additionally, the influence of the body movements and positions on measurement results with SleAp system are presented. The proposed system could help to determine the influence of OSA on the state of the cardiovascular system.

1 Introduction

Obstructive sleep apnea is a ventilation disorder characterized by repetitive pauses in breathing during sleep, despite the effort to breath, caused by obstruction of the upper respiratory tract. The airflow could be either reduced (hypopnea), or completely stopped (apnea). Recent studies have shown that OSA could affect approximately 3 to 7 % for adult men and 2 to 5 % for adult women in the general population [Punjabi, 2008].

Patients with sleep disorders, especially OSA, are exposed on intermittent hypoxemia and increased CO2 level, which could change the response of the organism to sleep. Repetitive cardiovascular stressors, experienced by patients with OSA every night over the years, might be the reason of development of chronic cardiovascular diseases [Shahar et al. 2001]. However this mechanism has not been explained yet.

Currently the syndrome is diagnosed on the basis of a detailed medical history and polysomnographic examination [Jaimchariyatam et al 2013]. First of all, the possible symptoms (e.g. day-time sleepiness, snoring) and risk factors (e.g. obesity, abuse of alcohol) are described. In case of clinically significant suspicion of

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_15, © Springer International Publishing Switzerland 2014

disease a polysomnography is conducted. It is an all-night examination during which a set of biosignals is recorded, e.g. electrical activity of the heart (ECG), blood oxygen saturation (SpO2), air flow and many other. The examination can be performed only in sleep disorders clinics, under the supervision of medical staff. The collected data are analyzed by the trained technicians, using specialized software. Because of those requirements polysomnography is an expensive and a non-widely utilized examination.

Different methods have been proposed for alternative, automatic diagnosis of OSA, such as analysis of heart rate variability using wavelet transformation [Shrader et al. 2000], or based on analysis of snoring associated sounds [Silva et al. 2012]. Those solutions allow for preliminary classification of the patients, which could help to reduce cost and make examination available for more patients. However, it is still necessary to conduct traditional polysomnography, after preliminary classification, to confirm the diagnosis.

In this paper, we present a method for integrated measurements of biosignals for automatic OSA detection. The specific objective of this work is to analyze the minimum set of features for the ECG signal that could produce acceptable classification results. Those features can be further expanded using other biosignals, measured by the portable SleAp device. Additionally, the influence of the body movements and positions on measurement results with SleAp system are presented.

2 Methods and Materials

The portable system was designed to perform integrated measurements of data for needs of OSA detection. The multisensory recording device measures snoring sounds, larynx vibrations, respiration parameters, and ECG. In this paper, the influence of ECG on OSA detection is mainly investigated.

Data analysis was performed using ECG signals collected in the Physionet database [Physionet 2013]. Those well described signals were used to propose and to validate a vector of features for the classification process. The final results can be used as a ECG-based, OSA detection subsystem in the developed multisensory SleAp device.

2.1 SleAp – The Developed Recording Device

A simplified block diagram of measurement system consists of three main units: a power section, a logic circuit, and set of sensors (Fig. 1). The power section contains three highly efficient DC/DC converters and charging process controller for lithium-polymer and lithium-ion batteries. The logic circuit consists of a 16-bit microcontroller, a Bluetooth module, and a microSD memory card. The most important, from a measurement point of view, is the sensor section. It consists of an integrated circuit for multichannel ECG measurement with built in impedance measurement circuit. Microphone and 3-axis digital accelerometer also belong to this section. We decided to measure a set of biosignals during the night, such as three channel ECG, respiration, snoring sounds and larynx vibrations. Recorded data are stored on a microSD memory card and could be analyzed on an external computer. The device requires four electrodes to be attached on the patient's chest and an additional sensor, containing the microphone and the accelerometer, to be placed close to the larynx.

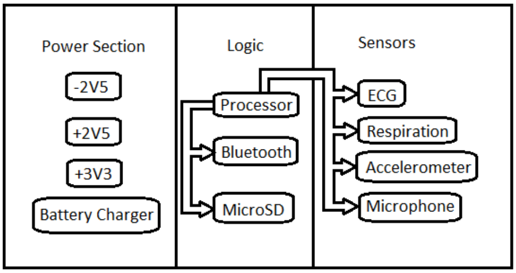


Fig. 1 A simplified block diagram of the developed device

Localizations of the electrodes and the sensor are shown in the Fig. 2. Typical electrode matrix allows for recording of three ECG leads (equivalent to Einthoven system). The same set of electrodes was used to measure impedance of the thorax using a two-electrode technique. The same electrodes were also used for delivering to the thorax an alternating current and for measuring of a resulting voltage.

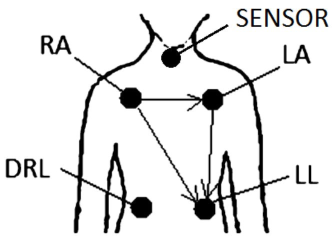


Fig. 2 Configuration of the electrodes' matrix and the localization of the sensor (RA- right arm electrode, LA- left arm electrode, LL- left leg electrode, DRL- driven right leg electrode SENSOR- microphone and accelerometer)

The ECG and the respiratory signals were sampled 250 times per second (250 Hz) using a 24 – bit analog to digital converter (ADC). The larynx vibrations were sampled at the same rate (250 Hz) however, using a 10 – bit converter. The acoustic signal (e.g. snoring sounds) was sampled at 1 kHz frequency with 10-bit ADC converter.

2.2 A Database of Signals

In development phase of the classification algorithm the apnea-ECG database has been used [Penzel et al. 2000]. It contains 35 all-night single channel ECG recordings. Half of them have apnea annotations for every minute of recording. The subjects of these recordings were patients between 27 and 60 years old, with weights between 53 and 153 kg. Twenty of them were diagnosed as severe OSA patients, five as mild OSA patients, and ten of them as normal (without OSA).

Additionally, the database was expanded to include data obtained from nocturnal examination of three subjects using the SleAp device. It allowed, due to recording of additional signals, calculation of additional parameters describing the respiration process. A thoracic impedance variation was examined in the same conditions as the ECG signals. The examined subjects, beside standard nocturnal examinations, were asked to simulate possible situations during the sleep.

2.3 Feature Extraction and the Classification Process

After data partitioning into 60-second blocks, a set of parameters was calculated for each block. First step was the detection of R waves, using algorithm proposed in [Pan and Tompkins 1985]. Next, Q and S waves were detected. They were defined as local minimum of the signal before and after R wave respectively. Based on this information, R-R interval and QRS complex duration were calculated. Our goal was to make this parameter independent of patient position changes during sleep. Finally, for each minute of the ECG signal, we have calculated the following parameters:

- Mean value of RR intervals,
- Median of RR intervals,
- Standard deviation of RR intervals,
- Inter quartile range of RR intervals,
- Mean value of QRS complex duration,
- Median of QRS complex duration,
- Standard deviation of QRS complex duration,
- Inter quartile range of QRS complex duration,
- Standard deviation of R wave amplitude,
- Inter quartile range of R wave amplitude,
- NN50,
- LF:HF ratio,
- Standard deviation of respiration signal amplitudes,
- Mean value of respiration signal amplitudes.

NN50 was defined as number of adjacent RR intervals, which differs more than 50 ms between each other. The LF:HF ratio – was calculated using power spectral density analysis of heart rate variability (HRV) and it was defined as low frequency (0.026 Hz to 0.06Hz) to high frequency (0.06 Hz to 0.25Hz) ratio [Otero et al. 2010].

Moreover, the information gain value for each of the parameters was calculated. The information gain evaluates the value of a parameter, by measuring the information entropy change from prior state to that one assuming the value of this parameter as given, and could be expressed by equation [Azhagusundari 2013]:

$$IG(X,a) = H(X) - H(X \mid a) \tag{1}$$

where $IG(X,a)$ – stands for information gain of parameter a, $H(X)$ and $H(X|a)$ denote entropy of variable X and entropy of variable X conditioned by parameter a respectively.

The classification process starts with sequencing the ECG signal into 60-seconds blocks. Each block was analyzed and features were extracted. Those features were divided into training set (70 %) and testing set (30 %) for support vector machine (SVM) classifier [Cortes and Vapnik 1995]. In the presented study, different kernel function has been tested. We have also tested a general approach, and a patient-specific approach (Fig. 3).

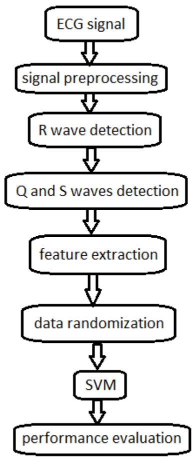


Fig. 3 Schematic diagram of the classification process

SVM constructs a hyperplane, or a set of hyperplanes in a high-dimensional space to separate the input data into two different classes (in this experiment: “OSA detected”, “OSA not detected”). A good separation is achieved by maximizing the margin between hyperplane and the input data from different classes.

If the input data are not linearly separable, it could be transformed into high-dimensional feature space using one of the kernel functions, such as quadratic, polynomial or Gaussian radial basis function [Vanschoenwinkel and Manderick 2005]. Such opportunity was also examined.

3 Results

The prototype of the proposed diagnostic device has been built (Fig. 4a). It contains of two 4-layers printed circuit boards (PCB) and plastic housing made in 3d printing technology. The developed device was designed as small as possible in order to minimize disturbance during a nocturnal examinations. The device and the sensor were attached to the body of the examined person using a parachute-like system of belts (Fig. 4b).

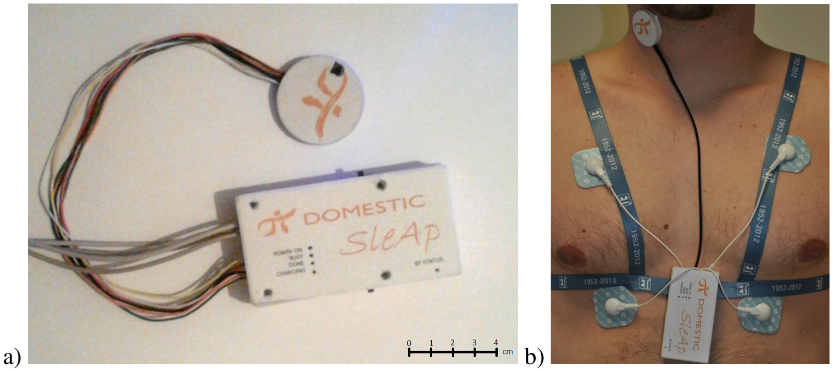


Fig. 4 a) The developed SleAp device. b) An example of the connection.

The recorded signals were dominantly of a very high quality and contained almost no interferences. Some examples are presented in Fig. 5. A three leads of electrocardiographic signals, electrical impedance of the thorax described as respiration, acoustic (microphone) and three components of accelerometric signal were measured and stored by the device SleAp.

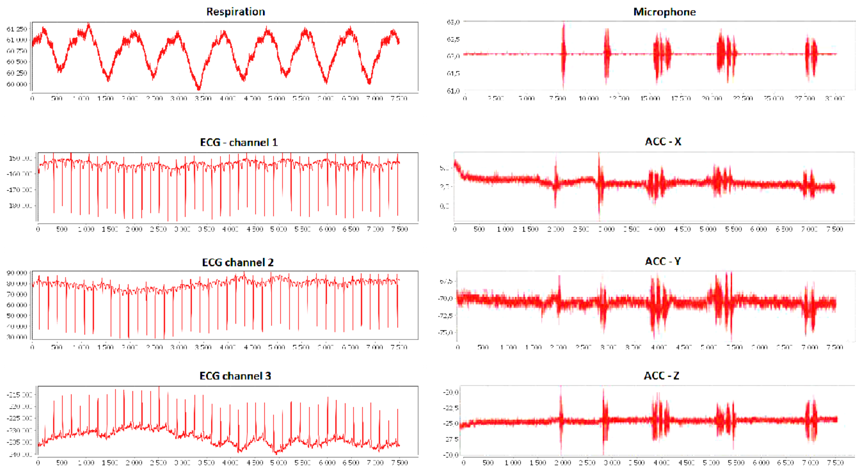


Fig. 5 An example of the recorded biosignals, acronym ACC stands for accelerometric signals

The most important signal changes, mainly in the ECG's amplitude, are associated with movements of the body and with different positions of the body. (Fig. 6).

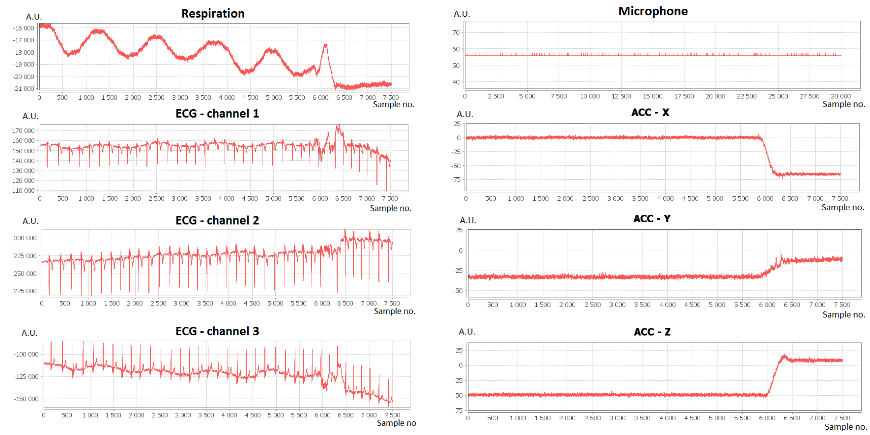


Fig. 6 An example of the recorded biosignals containing interferences evolved by body movement

Next, a body posture influence on the recorded ECG was examined. The signals were recorded for basic positions of a lying person (Fig. 7).

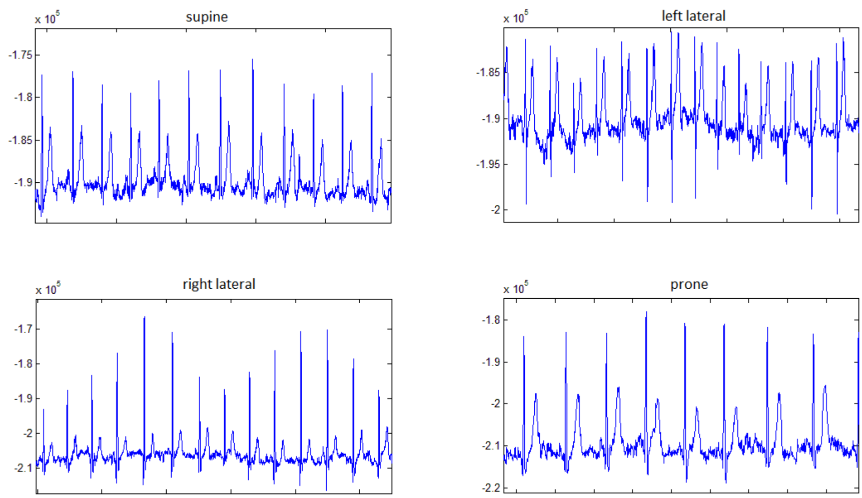


Fig. 7 ECG signal (a selected lead) registered during the different body positions

In each examination patient was asked to lye in different positions: supine, left lateral, right lateral and prone. Additionally, during this experiment the thoracic electrical impedance was measured and a low frequency time varying component recorded (Fig. 8).

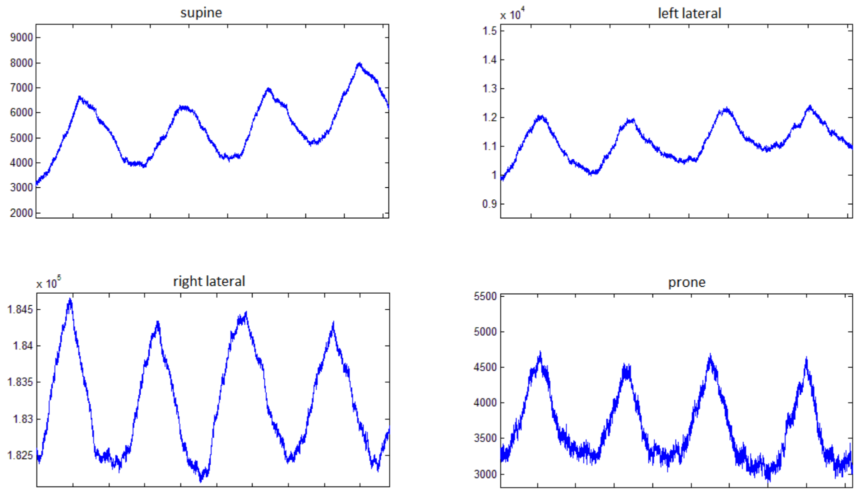


Fig. 8 Thoracic impedance measured in different body positions

In the following paragraph the results of classification are presented. However, in the presented results of the study only ECG signals recorded have been utilized. It was due to the fact that the number of examinations performed by means of the developed device was very small. Only three subjects passed a nocturnal examinations. Thus, a number of other than ECG signals was too small to perform any advanced analysis.

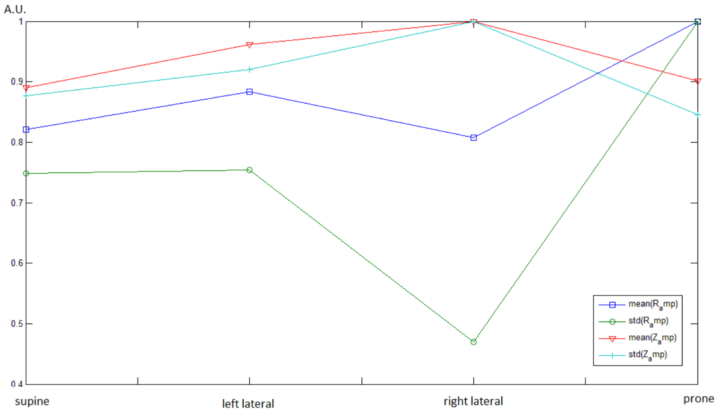


Fig. 9 Comparison of influence on body position changes on extracted features: mean value and standard deviation of R (ECG) and Z (electrical impedance - respiration) amplitudes are shown

First, as it is mentioned in the previous section the vector of features was obtained by appropriate data processing. After the extraction process the SVM classifier was trained. All data were randomly divided into training and testing sets. This step was repeated three times, to make sure that results are independent of the data randomization process. Four different kernel functions were analyzed. The classification accuracy for each iteration, together with an average result, are presented in the Table 1.

Table 1 Classification accuracy for the SVM classifier

Kernel function	Iteration 1	Iteration 2	Iteration 3	Average
Linear	76.14 %	75.68 %	76.59 %	76.13 %
Quadratic	93.48 %	92.05 %	91.44 %	92.32 %
Polynomial	92.88 %	93.94 %	93.18 %	93.33 %
RBF	94.51 %	94.15 %	94.73 %	94.46 %

Next feature selection based on information gain was used to find best features. Less features can reduce the complexity of the system (memory, computations, power). (Table 2). Performance of the classifier, which utilized the limited number of parameters, was compared to the classifier with full set of parameters (Table 3).

Finally, a patient specific approach has been tested. In this case, data for training and testing sets were obtained only from a single subject. The performance of this approach was tested for two cases: a patient with a mild OSA and a patient with a severe OSA (Table 4).

Table 2 Information gain values

Information gain	Parameter
0.7699	Std(R_amp)
0.765	Std(RR)
0.7609	LF:HF ratio
0.7508	Std(QS)
0.6311	Mean(QS)
0.6087	Mean(RR)
0.3919	Iqr(R_amp)
0.1798	Iqr(RR)
0.153	Median(RR)
0.1125	NN50p
0.1114	Iqr(QS)
0.0642	Median(QS)

Table 3 Performance evaluation of SVM classifier with reduced number of parameters (the highest values of information gain)

Number of parameters	Linear	Quadratic	Polynomial	RBF
2	74.46 %	78.13 %	78.26 %	81.93 %
4	81.52 %	87.50 %	91.17 %	91.03 %
6	82.20 %	90.49 %	92.12 %	92.53 %
All	82.07 %	92.12 %	92.93 %	94.16 %

Table 4 Performance evaluation of SVM classifier using patient-specific approach

Kernel function	Iteration 1	Iteration 2	Iteration 3	Average
Patient with mild OSA				
Linear	87.68 %	95.65 %	91.30 %	91.55 %
Quadratic	89.13 %	91.30 %	90.58 %	90.34 %
Polynomial	81.16 %	89.86 %	89.86 %	86.96 %
RBF	87.68 %	87.68 %	89.13 %	88.16 %
Patient with severe OSA				
Linear	93.44 %	86.89 %	85.25 %	88.53 %
Quadratic	86.89 %	83.61 %	83.61 %	84.70 %
Polynomial	86.89 %	86.89 %	86.89 %	86.89 %
RBF	90.16 %	86.89 %	86.89 %	87.98 %

4 Discussion and Conclusions

Respiration, body position and other factors, e.g. sleep disorders, influence on cardiac electrical and mechanical activity via different phenomena and mechanisms [Siebert et al 1999]. Many studies on relating the cardiac activity and the breathing process show a few mechanisms, which connect cardiovascular and respiratory systems. One of them is a respiratory sinus arrhythmia (RSA) [Otero et al 2010]. It is a physiological, natural variability in heart rate during the respiratory cycle. During inhalation, the heart rate increases, and during exhalation decreases. Sinus arrhythmia detection in electrocardiogram (e.g. using spectral analysis of RR intervals) allows estimating the respiration function [Pitzalis et al. 1998]). Heart rate is changing with successive phases of sleep and its natural variability may be affected by sleep disorders. Observation and analysis of RR intervals can be an effective way for preliminary classification of patients. Moreover, it is possible to estimate Apnea Hypopnea Index (AHI), by analyzing the ratio of low frequency to high frequency of heart rate variability, which corresponds to sympathetic and parasympathetic nervous system activity respectively. It has been proved that between AHI and LF/HF ratio exists a correlation. Number of adjacent RR intervals that differ more than 50 ms between each other are also correlated with AHI [Lado et al. 2011; Song et al. 2012].

Respiratory function disorders could be also associated with QRS duration changes. As was mentioned before, OSA may affect up to 7% of population. However this number significantly increases if we limit the group of interest to patients with coronary artery disease, and could be as high as 70%. Such high prevalence suggests that OSA may be a significant stressor in the development of cardiovascular events. Multiple studies have shown that OSA is associated with changes in cardiac structure, including left ventricular hypertrophy, which can widen the QRS complex [Drager et al. 2007]. QRS duration is an independent predictor of sudden cardiac death and mortality in conditions commonly associated with OSA, such as hypertension and heart failure. Clinical study on 221 participants reveals a weak, but significant and positive correlation between width of QRS complex and AHI. Furthermore this dependency increases with severity of OSA [Gupta et al. 2012].

Respiration process affects the conditions of ECG measurement. During the respiration the movement of electrodes with respect to each other, and to the heart could be observed. This effect generates slow amplitude changes in the electrocardiogram. Those changes could be extracted from the signal, and used to estimate respiration function. The amplitude of the R wave could be measured in reference to the baseline or to the QS waves [Widjaja et al. 2010]. Unfortunately, those approaches are exposed to errors caused by changing the body position during the sleep. As shown in the previous section the amplitudes of different ECG waves are modulated during the respiratory effort however, the depth of modulation depends on a position of the patient. While she/he is lying prone or supine, shapes of QRS complex are quite similar. Most visible difference is observed in amplitudes of T waves, which is not an object of this study. More interesting case is when the patient is lying lateral right or left. The asymmetry of QRS complex is clearly visible. Moreover, while lying lateral left, respiratory effort modulates S wave amplitude more significantly than R wave amplitude. To make R amplitude calculation independent of patient body position, those effects should be taken into consideration. Thus, we calculated the R wave amplitude in reference to minimum of Q and S waves amplitudes.

Unfortunately, information on respiration activity achieved by monitoring R wave amplitude is obtained non-uniformly in time. It is because the QRS complex is associated with depolarization process of ventricles. In turn, a variability of this process depends on many factors and it is a non-evenly distributed in time. It is equivalent to non-uniform sampling (in time) of respiration activity. This leads to a sophisticated approximation and re-sampling procedures. Basing on this knowledge, a set of features has been selected. Those are the parameters, which could be calculated from single-lead ECG, and contain information about respiratory function and breathing disorders. The goal was to acquire the best possible accuracy, with limited number of parameters to reduce a computational cost.

As shown in the previous section, we have obtained the best result for RBF kernel function in general approach. However, performance of quadratic and polynomial functions are almost as good as RBF. The worst result was obtained using linear kernel function. Slightly worse results were obtained with limited number of parameters. Thus, it is possible to lower the computational cost of the algorithm without significant reduction of the accuracy.

For the patient-specific approach, we got the best results for the linear kernel function. This leads to the conclusion that, in this case, we deal with linear problem in the classification process. The problem with this approach was a small set of data, which could be used in the training process. Database from physionet.org consists only of one-night recordings for each patient [Physionet 2013]. In effect, it limits the amount of input data. Most probable explanation of better performance of linear function is the presence of over fitting error for other kernel functions. Several papers describe sleep apnea detection using the SVM classifier. C. Varon [Varon 2013] proposed the algorithm using 4 features as an input for SVM classifier and obtained 85.07% accuracy. L. Almazaydeh with co-workers [Almazaydeh et al. 2012] described algorithm based on 10 features and resulting accuracy of 96.5%, however, they have limited the input data only to selected specific parts of ECG signals. It is also worth to mention that the best-achieved result during the contest on physionet.org was 92.62% accuracy [McNames et al. 2000]. This comparison shows, that presented algorithm has achieved a comparable or better performance for both general and patient-specific approach.

The presence of OSA events is visible not only in ECG. Therefore, the accuracy of the detection algorithm could be possibly increased using other biosignals (and associated parameters). One example could be measurement of respiration activity using another method than ECG. Such a measurement could be performed using an electrical impedance method [Wtorek 2003]. Variability of the thoracic impedance is a result of thoracic shape and volume changes, which are associated with the process of breathing. In addition, there are also changes in the internal structure of the thorax, such as organs movement, lungs expansion, etc. The changes in electrical conductivity of the lungs' tissue involved by changes in a volume of gas in the lungs during inhalation and exhalation processes are also important. The impedance measurement is carried out using a two-electrodes method. High-frequency current is supplied to the tissue via electrodes. The potential difference produced in this way is measured. Impedance is defined as the ratio of the measured voltage to the given current. The most significant advantage of thoracic impedance measurement is that the measurements are conducted continuously, not only at the moment of depolarization of ventricles termination. It is also important, that the recorded signal using this method is not strongly modified by changes of patient body position during the sleep. However, it is also susceptible to thorax deformations caused by other activity than respiration.

The developed SleAp system consists of two interconnected housings. The microphone with associated electronic circuit and accelerometer are placed in circular housing and connected to the device using a very flexible and pliable wires [Rendon et al. 2007]. It protects the accelerometer from attenuating vibrations of higher frequency and a very small power. As a result SleAp enables recording of the following, in addition to ECG (3 leads), set of signals: the accelerometric, the acoustic, and thoracic impedance changes (reflecting mainly the respiration activity) [Przystup et al. 2013]. In spite of enabling acquisition of eight different signals during a whole night its dimensions seems to be reasonably small. Moreover, the quality of the recorded signals (Fig. 5-8) is good enough to perform a reliable analysis.

Acknowledgment. This work has been partially supported by European Regional Development Fund concerning the project: UDA-POIG.01.03.01-22-139/09-00 -“Home assistance for elders and disabled – DOMESTIC”, Innovative Economy 2007–2013, National Cohesion Strategy and by Statutory Funds of Electronics, Telecommunications and Informatics Faculty, Gdansk University of Technology.

References

[Almazaydeh el at 2012] Almazaydeh, L., Elleithy, K., Faezipour, M.: Obstructive sleep apnea detection using SVM-based classification of ECG signal features. In: Engineering in Medicine and Biology Society, pp. 4938–4941 (2012)

[Azhagusundari 2013] Azhagusundari, B., Thanamani, A.S.: Feature selection based on information gain. International J. of Innovative Technology and Exploring Engineering 18, 2278–3075 (2013)

[Cortes and Vapnik 1995] Cortes, C., Vapnik, V.: Support-vector networks. Machine Learning, 273–297 (1995)

[Drager et al. 2007] Drager, L.F., Bortolotto, L.A., Figueiredo, A.C., Silva, B.C., Krieger, E.M., Lorenzi-Filho, G.: Obstructive sleep apnea, hypertension, and their interaction on arterial stiffness and heart remodeling. Chest 227, 1379–1386 (2007)

[Gupta et al. 2012] Gupta, S., Cepeda-Valery, B., Romero-Corral, A., Shamsuzzaman, A., Somers, V.K., Pressman, G.S.: Association between QRS duration and obstructive sleep apnea. J. of Clinical Sleep Medicine, 649–654 (2012)

[Jaimchariyatam et al. 2013] Jaimchariyatam, N., Dweik, R.A., Kaw, R., Aboussouan, L.S.: Polysomnographic determinants of nocturnal hypercapnia in patients with sleep apnea. Clinical Sleep Medicine 9(3), 209–215 (2013)

[Lado et al. 2011] Lado, M.J., Vila, X.A., Rodriguez-Linares, L., Mendez, A.J., Olivieri, D.N., Felix, P.: Detecting sleep apnea by heart rate variability anaysis: assesing the validity of databases and algorithms. J. Med. Syst. 35(4), 473–481 (2011)

[McNames et al. 2000] McNames, J., Fraser, A., Rechtsteiner, A.: Sleep apnea classification based on frequency of heart-rate variability. Computers in Cardiolog, 749–752 (2000)

[Otero et al. 2010] Otero, A., Vila, X., Palacios, F., Coves, F.J.: Detection of obstructive sleep apnea from the frequency analysis of heart rate variability. In: Proc. 3rd International Conference on Bio-inspired Systems and Signal Processing, Valencia, pp. 359–362 (2010)

[Pan and Tompkins 1985] Pan, J., Tompkins, W.J.: A real-time QRS detection algorithm. IEEE Trans. on Biomedical Engineering 32(3), 230–236 (1985)

[Pitzalis et al. 1998] Pitzalis, M.V., Mastropasqua, F., Massari, F., Passantino, A., Colombo, R., Mannarini, A., Forleo, C., Rizzon, P.: Effect of respiratory rate on the relationships between RR interval and systolic blood pressure fluctuations: a frequency-dependent phenomenon. Cardiovascular Research 38(2), 332–339 (1998)

[Punjabi, 2008] Punjabi, N.M.: The epidemiology of adult obstructive sleep apnea. American Thoracic Society 5(2), 136–143 (2008)

[Penzel et al. 2000] Penzel, T., Moody, G.B., Mark, R.G.: The apnea-ECG database. Computers in Cardiology, 255–258 (2000)

[Physionet 2013] Physionet (2013), `http://www.physionet.org` (accessed August 15, 2013)

[Przystup et al. 2013] Przystup, P., Bujnowski, A., Ruminski, J., Wtorek, J.: A multisensor detector of a sleep apnea for using at home. IEEE Xplore Digital Library, pp. 513–517 (2013)

[Rendon et al. 2007] Rendon, D.B., Rojas, J.L., Ojeda, F., Crespo, L.F., Morillo, D.S., Fernández, M.A.: Mapping the human body for vibrations using an accelerometer. In: Eng. in Med. and Biol. Society, pp. 1671–1674 (2007)

[Shahar et al. 2001] Shahar, E., Whitney, C.W., Redline, S., Lee, E.T., Newman, A.B., Nieto, F.J., O'Connor, G.T., Boland, L.L., Schwartz, J.E., Samet, J.: Sleep-disordered breathing and cardiovascular disease: cross-sectional results of the sleep heart health study. American J. of Respiratory and Critical Care Medicine 163(1), 19–25 (2011)

[Shrader et al.. 2000] Shrader, M., Zywietz, C., von Einem, V., Widiger, B., Joseph, G.: Detection of sleep apnea in single channel ECGs from the PhysioNet data base. Computers in Cardiology, 263–266 (2000)

[Siebert et al. 1999] Siebert, J., Wtorek, J., Rogowski, J.: Stroke volume variability - cardiovascular response to orthostatic maneuver in patients with coronary artery diseases. Annals of the New York Academy of Science 873, 182–190 (1999)

[Silva et al.. 2012] de Silva, S., Abeyratne, U.R., Hukins, C.: Impact of gender on snore-based obstructive sleep apnea screening. Physiol. Meas. 33(4), 587–601 (2012)

[Song et al. 2012] Song, M.K., Ha, J.H., Ryu, S.H., Yu, J., Park, D.H.: The effect of aging and severity of sleep apnea on heart rate variability indices in obstructive sleep apnea syndrome. Psychiatry Investigation, 65–72 (2012)

[Vanschoenwinkel and Manderick 2005] Vanschoenwinkel, B., Manderick, B.: Appropriate Kernel Functions for Support Vector Machine Learning with Sequences of Symbolic Data. In: Winkler, J.R., Niranjan, M., Lawrence, N.D. (eds.) Machine Learning Workshop. LNCS (LNAI), vol. 3635, pp. 256–280. Springer, Heidelberg (2005)

[Varon 2013] Varon, C.: Sleep apnea classification using least-squares support vector machines on single lead ECG. In: Engineering in Medicine and Biology Society, pp. 5029–5032 (2013)

[Widjaja et al. 2010] Widjaja, D., Taelman, J., Vandeput, S., Braeken, M.A.K.A., Otte, R.A., Van den Bergh, B.R.H., Van Huffel, S.: ECG-derived respiration: comparison and new measures for respiratory. Computing in Cardiology 37, 149–152 (2010)

[Wtorek 2003] Wtorek, J.: Electrical impedance technique in medicine. Series of Momographs-43. Publishing Office of Gdansk University of Technology (2003) (in Polish)

Application of Fourier Transforms in Classification of Medical Images

A. Sokołowski[1] and T. Pardela[2]

[1] Rzeszow University of Technology, Rzeszów, Poland
alex5@prz.edu.pl
[2] University of Information Technology and Management, Rzeszów, Poland
tpardela@wsiz.rzeszow.pl

Abstract. Selected methods of image processing have been applied to expert system supporting the process of identifying medical images. Fourier transform is well known tool for many applications in the processing of images in many fields of science and technology, also in medicine. In this case the image processing consists in spatial frequencies analysis of Fourier transforms of medical images. Skin lesions are studied on the base of their images and it seems that Fourier transformation is the right toll for such research. The distributions of selected colors in Fourier transform images are studied.

1 Introduction

The approach presented in the paper has been applied to knowledge base containing images of skin lesions. The descriptive method – ABCD formula [Friedman et al. 1985; Adrus et al. 2004] is well known method applied to routinely diagnosing melanoma. The above formula has been applied to databases being collections of cases describing melanoma data. The above knowledge base has been used for study the optimization of the ABCD formula with help of data mining system LERS [Grzymała-Busse 1992, Grzymała-Busse 1997]. The results of the classification of melanocytic skin lesions based on belief rules on the base of the same databases also have been presented [Grzymała-Busse et al. 2005].

Melanoma lesions studies have been used also methods of applying image processing algorithms [Schindewolf et al. 1993; Colot et al 1998]. The method epiluminescence microscopy (ELM) has been applied [Ganster et al. 2001] to melanoma image recognition. The wavelet transforms as another study methods and related to them coefficients created by wavelet decompositions have been used to classification benign and malignant melanoma skin lesions [Sikorski 2004]. To special studies of biological activity of cells, particularly to study early spectral changes accompanying malignant transformation of cells [Bogomolny et al. 2007] Fourier transform has been applied.

Fourier Transformation is well known as the research tool, but it seems that for the first time it is used as a tool for studying the melanoma skin lesions.

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_16, © Springer International Publishing Switzerland 2014

Fourier transformation has many features, which are useful in many fields of science. It seems to be very useful in image classification especially in melanoma image recognition. It is therefore important to explore algorithms that use this type of research tools. The expert systems with the addition of these tools will be more reliable.

2 Fourier Transform Properties

The Fourier transform is a mathematical transformation applied to transform signals or images between time (or spatial) domain and frequency domain, which has many applications in physics and engineering. The function *F*, which is the result of the transformation of the function *f* is called Fourier spectrum. In the general case Fourier transform converts function with complex values into functions with complex values. In the case of transformation images, Fourier transform converts real functions into complex functions. Then it needs to give a picture of complex function in the form of real function. One way to do this is to calculate the new real function as the square of absolute values of complex function *F* and then transform it in graphical form.

It applies various algorithms of fast Fourier transform (FFT). By far the most commonly used FFT is the Cooley-Tookey algorithm [Cooley et al. 1965].

To study the structure of images Fourier transformation is particularly useful, and this is a fact fully confirmed in scientific and technical literature. One-dimensional case is the easiest way to demonstrate how Fourier transform works. The Fourier transform for this case has the following form:

$$F(\nu)=\int_{-\infty}^{\infty} f(x)\exp(-i2\pi x\nu)dx, \tag{1}$$

where $f(x)$ is the examined function, ν is the spatial frequency, and $i=\sqrt{-1}$. If the simple example of sin function will be chosen:

$$f(x)=a\sin(2\pi\nu_1 x), \tag{2}$$

where ν_1 denotes fixed spatial frequency, and a is the amplitude, then

$$F(\nu)=\frac{a}{2\pi i}\left[\delta(\nu_1-\nu)+\delta(\nu_1+\nu)\right], \tag{3}$$

where δ denotes Dirac delta distribution. Equation (3) means that there are only two spatial frequencies $\pm\nu_1$. If the linear combination of two sin function is used:

$$f(x)=a_1\sin(2\pi\nu_1 x)+a_2\sin(2\pi\nu_2 x), \tag{4}$$

then

$$F(\nu)=\frac{1}{2\pi i}\left[a_1\delta(\nu_1 \pm \nu)+a_2\delta(\nu_2 \pm \nu)\right] \tag{5}$$

and the spectrum of function f consists of four spatial frequencies $\pm\nu_1, \pm\nu_2$. In more general case $f(x)$ can be the linear combination of many sin functions:

$$f(x)=\sum_{k=1}^{K} a_k \sin(2\pi\nu_k x), \tag{6}$$

and then

$$F(\nu)=\frac{1}{2\pi i}\sum_{k=1}^{K} a_k\delta(\nu_k \pm \nu). \tag{7}$$

The above equations can be easily generalized to 2 dimensions. Mechanism of action of Fourier transformation lies in the fact that it is easy to distinguish between structures of tested images trough the study of spatial frequencies distributions. It seems that analysis of the selected colors distributions should provide
relevant information from the point of view of medical diagnostics.

The equation describing the Fourier transform for two-dimensional functions has the following form:

$$F(\nu_x,\nu_y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{i2\pi(x\nu_x+y\nu_y)}dxdy\,, \tag{8}$$

where x, y denote the variables on the image plane, and ν_x, ν_y denote variables on Fourier plane. Because of the fact that studied images are sampled functions, equation (8) takes the following form:

$$F(m,n)=\sum_{k=0}^{K}\sum_{l=0}^{L} f(k,l)e^{-i2\pi(m\Delta\nu_x k\Delta x+n\Delta\nu_y l\Delta y)}\,, \tag{9}$$

where $\Delta\nu_x$, $\Delta\nu_y$ are the steps of the Fourier plane sampling, Δx, Δy are the steps of the image plane sampling, m=0, … M, and n=0, …, N. The values of the above steps should fulfill the so called sampling theorem [Marks 2009], i. e.

$$\Delta\nu_x \le \frac{1}{K\Delta x},\ \Delta\nu_y \le \frac{1}{L\Delta y}. \tag{10}$$

To apply the fast Fourier transform (FFT) algorithm the inequalities in equation (10) must to stay equalities. It is worth apply the FFT algorithm, because the calculation of the Fourier transform is done almost in real time.

Fourier transform as a research tool has many interesting properties, which can be used in many areas of science and technology. In the next parts of the paper a few ideas how to use selected properties of Fourier transform in image classification are presented.

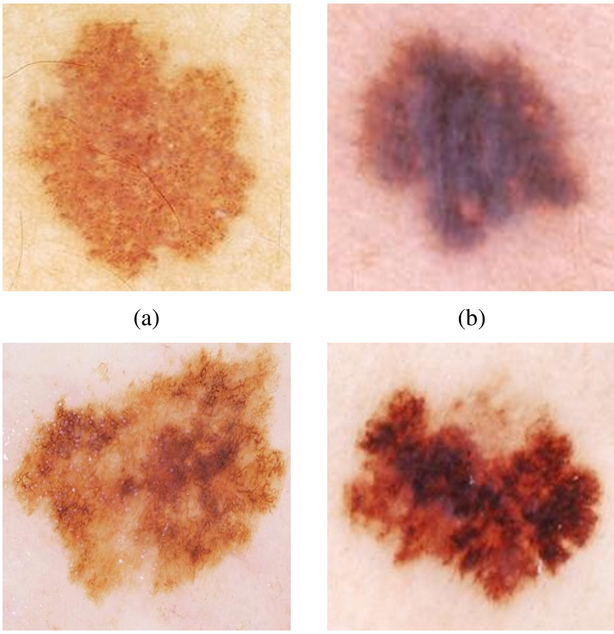
3 Numerical Calculations

In order to obtain processed images it has been applied FFT to images shown in Fig. 1. Fig. 1 presents four examples of melanoma disorders: (a) benign nevus, (b) blue nevus, (c) suspicious nevus, and (d) malignant nevus image.

Fig. 2 consists of Fourier transform images of the above input images for benign nevus, blue nevus, suspicious nevus, and malignant nevus, respectively.

A superficial analysis of the images in Fig. 2 shows that for such kinds of images their Fourier spectra in pure forms rather do not contribute to the image classification process. One can try to remove from Fourier spectra selected spatial frequencies. This situation is depicted in Fig. 3. Fourier spectra majority kinds of images have the common feature: the weight of the zero spatial frequency is in general much higher than the others. Because the zero and near zero spatial frequencies do not contribute the information about the image details, their removal is well-founded. Images on Fig. 3 also represent Fourier spectra, but here the spatial frequencies of radius of 5 pixels around the zero frequency have been removed. To explain shapes of Fourier spectra visible on the figures above, it should be mentioned that after every operations on pixels every Fourier spectrum has been scaled again.

The aim of this operation is to amplify the higher spatial frequencies because they are usually very weak compared to zero frequency.



(a) (b) (c) (d)

Fig. 1 The example of images before processing (a) – image of benign nevus, (b) – blue nevus, (c) – suspicious nevus, (d) – malignant nevus

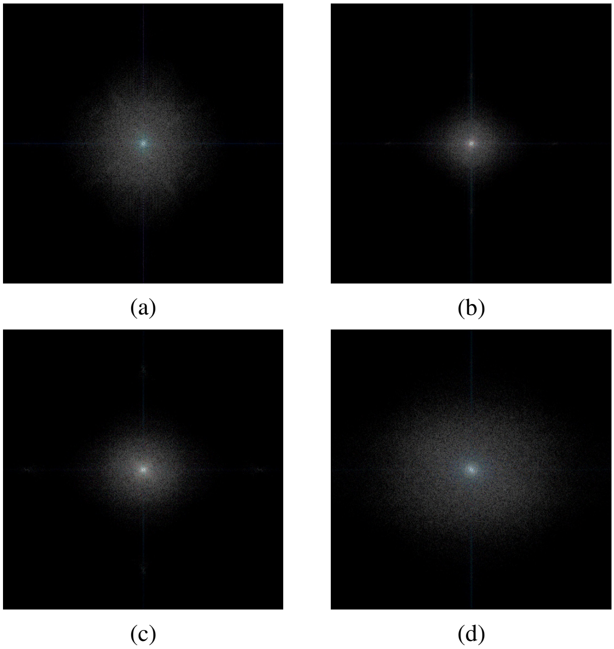


Fig. 2 The Fourier spectra of the images shown on Fig. 1, (a) – (d) denote benign nevus, blue nevus, suspicious nevus, and malignant nevus, respectively

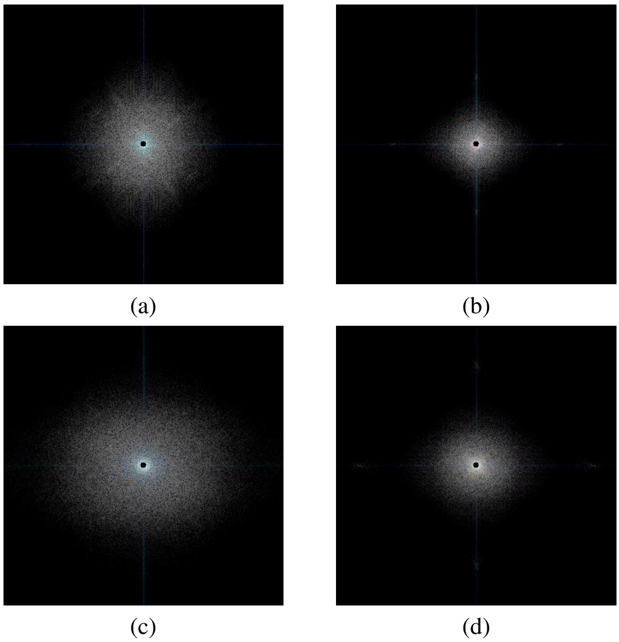


Fig. 3 The Fourier spectra from Fig 2 but with removed the main spatial frequencies

The aim of this operation is to amplify the higher spatial frequencies because they are usually very weak compared to zero frequency.

It seems that the distributions of colors in the Fourier transform images can be important to image classification process. As it can be seen on images from Fig. 4, it can be interesting to analyze for example the distributions of the grey color on the images. The distributions of the grey color in the Fourier images are presented here. The grey color in the RGB COLOR model is represented by three numbers 100, 100, 100 that mean the intensities of the red, green, and blue colors, respectively. To select any color from the image there should to determined the limits of the RGB components for this color. The grey color on Fig. 4 is displayed inside the border between 60 – 140 of base color of the RGB COLOR model. Fig. 5 shows distribution of color described by the coefficients in intervals 100 – 180, 20 – 100, 160 – 240 of the RGB COLOR model. These distributions have been obtained from the images presented in Fig. 3 in the same order ((a) – benign nevus, (b) – blue nevus, (c) – suspicious nevus, (d) – malignant nevus).

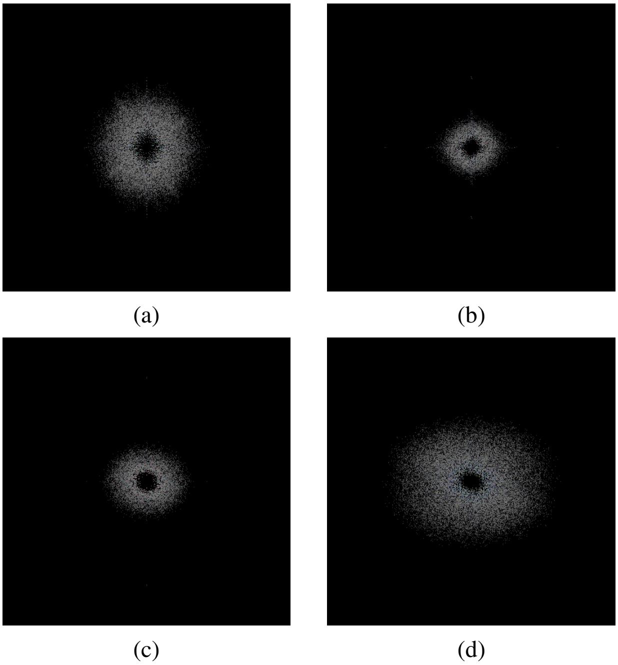


Fig. 4 The distribution of the grey color on Fourier spectra images from Fig. 3

The process of melanoma image classification will require the analysis of distributions of few carefully selected colors.

It is important to remark that all photographs of melanoma lesions should be taken under similar conditions. It concerns the distance from the object, object illumination, and the camera lens.

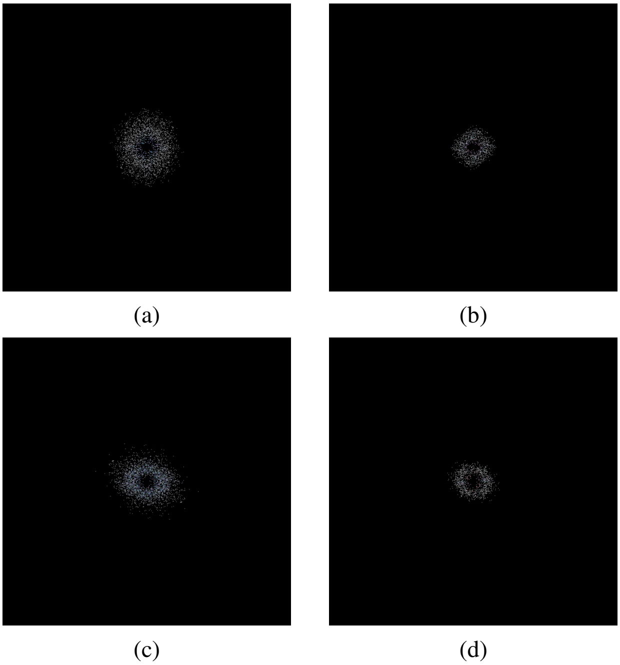


Fig. 5 The example of the distributions of yellow color in Fourier spectra images with removed zero spatial frequency

4 Conclusions

The research presented in the paper are preliminary studies about application of image processing tools to extraction of colors from Fourier spectra images. It can seen from the presented results, image processing tools can be very useful in the classification of the medical images. Adding the algorithms that extract some specific colors from examined images should be also useful. For a comparative analysis of processed images there can be successfully applied the selected statistical tools.

Acknowledgment. We would like to express big thanks for making available melanoma images to prof Z.S. Hippe, Chair of Expert Systems and Artificial Intelligence, University of Information Technology and Management in Rzeszow, Poland.

References

[Andrus et al. 2004] Andrews, R., Bajcar, S., Grzymała-Busse, J.W., Hippe, Z.S., Whiteley, C.: Optimization of the ABCD formula for Melanoma diagnosis using C4.5, a data mining system. In: Tsumoto, S., Słowiński, R., Komorowski, J., Grzymała-Busse, J.W. (eds.) RSCTC 2004. LNCS (LNAI), vol. 3066, pp. 630–636. Springer, Heidelberg (2004)

[Bogomolny et al. 2007] Bogomolny, R., Huleihel, M., Suproun, Y., Mordechai, S.: Early spectral changes of cellular malignant transformation using Fourier transform infrared microspectroscopy. Journal of Biomedical Optics 12, 024003 (2007)

[Colot et al. 1998] Colot, O., Devinoy, R., Sombo, A., de Brucq, D.: A colour image processing method for melanoma detection. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 562–569. Springer, Heidelberg (1998)

[Cooley et al. 1965] Cooley, J.W., Tookey, J.W.: An algorithm for the machine calculation of complex Fourier series. Mathematics of Computations 19, 297–301 (1965)

[Friedman et al. 1985] Friedman, R.J., Rigel, D.S., Kopf, A.W.: Early detection of malignant melanoma: the role of physician examination and self-examination of the skin. CA Cancer J. Clin. 35, 130–151 (1985)

[Ganster et al. 2001] Ganster, H., Pinz, P., Rohrer, R., Wildling, E., Binder, M., Kittler, H.: Automated melanoma recognition. IEEE Transactions on Medical Imaging 20, 233–239 (2001)

[Grzymała-Busse 1997] Grzymała-Busse, J.W.: A new version of the rule induction system LERS. Fundamenta Informaticae 31, 27–39 (1997)

[Grzymała-Busse 1992] Grzymała-Busse, J.W.: LERS A system for learning from examples based on rough sets. In: Słowiński, R. (ed.) Intelligent Decision Support. Handbook of Application and Advances of the Rough Sets Theory Rough Sets, pp. 3–18. Kluwer Academic Publishers, Dordrecht (1992)

[Grzymała-Busse et al. 2005] Grzymała-Busse, J.W., Hippe, Z.S., Mroczek, T.: Belief rules vs. decision rules: a preliminary appraisal of the problem. In: Kłopotek, M.A., et al. (eds.) Proc. of the International IIS, Poland, pp. 431–435 (2005)

[Marks 2009] Marks II, R.J.: Handbook of fourier analysis and its applications, ch. 5-8. Oxford University Press, Chapters (2009)

[Schindewolf et al. 1993] Schindewolf, T., Stolz, W., Albert, R., Abmayr, W., Harms, H.: Classification of melanocytic lesions with color and texture analysis using digital image processing. Anal. Quant. Cytol. Histol. 15, 1–11 (1993)

[Sikorski 2004] Sikorski, J.: Identification of malignant melanoma by wavelet analysis. In: Proc. of Student/Faculty Research Day, CSIS, Pace University, USA, pp. 10.1–10.9 (2004)

A Mobile Phone Application for Recognizing Objects as a Personal Aid for the Visually Impaired Users

K. Matusiak, P. Skulimowski, and P. Strumiłło

Lodz University of Technology, Łódź, Poland
karol.matusiak@dokt.p.lodz.pl,
{piotr.skulimowski,pawel.strumillo}@p.lodz.pl

Abstract. In this work we report applications developed for Android smartphones that were designed for the blind and visually impaired users. The main application is capable of matching photographed objects to a database of objects, e.g. food or medicine containers. The two other applications are aimed at detecting major colours and direction location of the maximum brightness regions in the captured scenes. The conducted tests that were run on the modern smartphones indicate that the database objects can be recognized with an accuracy reaching 89% with an average recognition time of approx. 30 sec. We conclude the paper with a short summary of the tests of the software dedicated to blind and partially sighted users.

1 Introduction

According to the European Blind Union (www.euroblind.org) the estimated population of the blind and partially sighted Europeans reaches 30 million. The conducted questionnaires conducted in the cooperation of the Polish Blind Union indicate that mobility and the so called activities of daily living are the major problems that are experienced by the persons with visual disabilities [Strumillo 2012]. Consequently blindness deprives the disabled of equal chances in access to education, skilled occupation and social activities.

Mobility of the visually impaired is diminished due to problems in spatial orientation and safe wayfinding and navigation in outdoor spaces. Whereas activities of daily living, according to Hersh and Johnson [Hersh and Johnson 2008], which the partially sighted point out to pose problems are the following:

- Health care, i.e. health care monitoring and the use of medicines,
- Alarm, alerting and timekeeping, e.g. solutions for controlling household appliances, smoke and monoxide detectors,
- Food preparation and consumption; utensils of special construction and talking (or sonified) kitchenware (e.g. liquid level indicators),
- Indoor environmental control and use of appliances, i.e. accessible systems providing feedback information about in a format suitable for a blind user,
- Money and finance; high- and low-tech solutions making shopping and money operations possible for the blind.

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_17,

In this chapter we concentrate our attention on the daily living activities related to personal care systems, in particular, systems that enable the visually impaired to identify objects, e.g. food packs, medicine containers and other items. Currently available solutions can be grouped into the two major groups:

- Low-tech labelling systems in which labels are attached to objects, e.g. with tactile signs or text messages in Braille [Gill 2008; Onishi and Ono 2011],
- High-tech systems that employ 1-D, 2-D barcodes, talking labels or radio-frequency identification devices (RFID) [Tinnakorn and Punyathep 2011].

These systems, however, require attaching special tags or visual signs to the objects. Consequently, they can be costly, since such systems need to be regularly maintained to keep them up to date.

We report a solution aimed at aiding the partially sighted persons in colour detection, light direction detection and recognition of objects. The system is based on a dedicated image recognition application running on an Android system smartphone. Image recognition results are communicated to the blind user by means of pre-recorded verbal messages.

2 Mobile Applications Dedicated to the Partially Sighted Users

2.1 Choosing a Target Mobile Platform

The currently observed trend is that phones with a physical keyboard are less and less popular. This tendency has initially posed a new usability barrier for the blind who had to go through difficulties in handling small touch-screen devices. However, new user interfaces based on the so called touch gestures has significantly improved user-friendliness of the touch-screen devices for the blind users. Currently, the most popular choice of smartphones among visually impaired users is either an iPhone with a very good tool called VoiceOver or a cheaper choice of one of the Android-based device. Note here that, the non-operating system phones, are not commonly used by blind users, because they usually do not offer any special tools aiding this group of users.

For our image analysis application we have decided to choose the Android platform, because of wide popularity of the Android based devices. According to Gartner's analysis in the second quarter of 2013 this mobile platform has gained 79% share of the smartphone market [WWW-1 2013]. Moreover it is equipped with speech synthesis and accessibility software.

The Android system is similarly well adapted to the needs of the blind as the iOS system. It is equipped with the TalkBack accessibility service. This tool allows to verbally comment and read GUI (Graphical User Interface) components.

Choosing a modern smart phone has also another advantage: they are equipped with good quality on-board units like digital cameras, GPS receivers, audio recorders etc. These units can be used in a dedicated software for the blind user. On the other hand, there are numerous dedicated devices available for the blind such as navigation devices, colour detector readers etc. However, they feature the two main disadvantages. Firstly, they are expensive (their cost is much higher than that of mobile applications) and secondly, the blind user is forced to carry several devices, each for a specific application, e.g. a phone and a navigation device.

2.2 Examples of Existing Solutions

There are a number of image analysis applications for mobile devices currently available on the market. The software termed Recognizer developed by LookTel [WWW-2 2013] is an application dedicated for iPhones, that is supposed to recognize an object within the camera field of view that was previously stored in a local database of objects' images. The application is intended to help visually impaired people to recognize household objects. For the best results, object templates stored in the database should be captured by a sighted person in a predefined orientation. The authors do not show recognition results for arbitrary orientation of the scanned objects.

The EyeRing project [Nanayakkara et al. 2012] is a finger-worn device that communicates with an Android mobile phone. The device houses a VGA mini-camera, a 16MHz AVR microcontroller, a Bluetooth module and control buttons. The task of the mobile device is running speech processing algorithms and all computer vision algorithms. The currently implemented functionality of the device is as follows: detection of banknotes, recognition of colours and distance calculation which is supposed to work as a "virtual walking cane". This solution, however, is costly and requires an additional device to be worn by a blind user.

3 Image Processing Applications

We propose the three following image processing modules that we have developed for Android smartphones: a colour detection module, an object recognition module and a light source detector. An advantage of the algorithm we proposed for object recognition model is its scale and orientation invariance.

3.1 The Colour Detection Module

The colour detection algorithm works on images taken with an automatic flash with the smallest resolution possible (320x480 pixels). The RGB colour images are converted into the HSI (Hue Saturation Intensity) colour images. This colour space enables to represent the colour in a single parameter.

We tested two different approaches. In the first approach, in the HSI colour space, the average value of the colour (the H component) is determined for the photo taken. The average colour is compared with a predefined reference table of colours and the colour of the photographed object is determined. If the image is too dark or too bright a special warning message is generated.

In the second approach a special colour histogram is computed. Each image pixel is represented in the HSI space and allocated to a predefined colour histogram bin. If there is a significant disproportion between the first and the second most frequent image colour the most frequent one is communicated to the user only. If the occurrence frequency of more than one colour is similar, the application informs the user about mixture of colours, e.g. by voicing the message "yellow-red colour".

An example of the colour detection module in use is shown in Figure 1. In the first version of the software we have used more than 50 different colours. However, the tests with participation of blind users revealed that just 16 colours are more suitable to be used in practice. This number was limited due to problems in recognising subtle changes of colours in varying lightning conditions.

We have tested the colour detection algorithm on reference colours (blue, green, yellow, red, orange, gray, brown) for verifying its robustness under varying lighting conditions. Two test runs were carried out. One with the camera flash switched on and the second for the flash switched off. For the flash on, for all lighting conditions the tested reference colours were recognized correctly except for the brown which was always recognised as orange. For the flash off and good lighting conditions all except the orange colour (which was recognised as red) were recognised correctly. Finally, for the flash off and poor lighting conditions the wrongly recognised colours were yellow (recognised as orange) and orange (recognised as red).

3.2 The Light Detector Module

The idea of light detector module was proposed by one of the blind users. The application operates in real time and is based on the content of the camera's preview image. In the first step, the average brightness of the centre part of the image is calculated. Next, an audio signal with a frequency dependent on the average brightness is generated. The brighter the image the higher is the frequency of the generated sound. The application proved suitable in robust localization of light sources, for example streetlights or a lamp in the room.

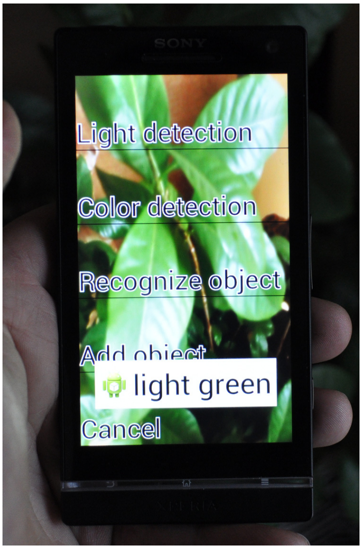


Fig. 1 Colour detection module of the application running on a Sony Xperia S smartphone. Detected colour is read using the TextToSpeech module

3.3 Image Recognition Algorithm

The goal was to design an application which would allow to recognize objects from images recorded by the camera of a mobile device. The object recognition algorithm should be insensitive to image registration parameters, i.e. scale, rotation and lighting conditions. Moreover, the recognised object should be robustly detected and localised in the image context (e.g. among other similar objects). The SIFT (Scale-Invariant Feature Transform) proposed in [Lowe 1999] was applied in the developed application. The SIFT is considered as a very powerful computer vision algorithm for detecting and describing local image features. SIFT allows to compute feature descriptors strongly independent on the image registration conditions. These descriptors are further used to recognize objects in the proposed application.

The main steps of the algorithm used in the application are shown in a block diagram in Figure 2. Firstly, an RGB image is captured by a mobile phone camera. Then, the image is converted into greyscale (in our approach colour information is not used for the object recognition procedure). Further, a key point detection procedure is performed. Originally, Difference of Gaussian (DoG) based algorithm was used, which builds a special image structure called the Gaussian pyramid [Lowe 1999].

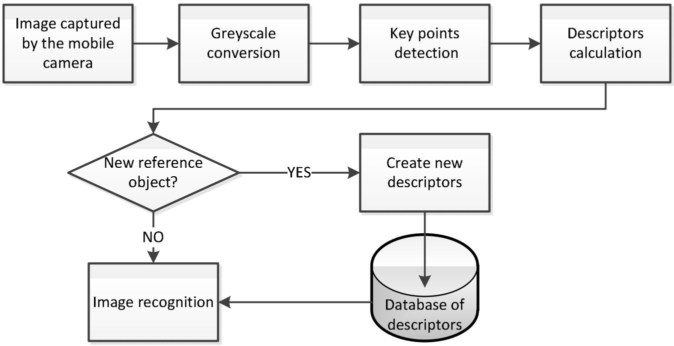


Fig. 2 Block diagram illustrating main steps of the object recognition algorithm

To improve performance of the application the features from Accelerated Segment Test (FAST) [Rosten and Drummond 2006] algorithm was implemented. It is one of the fastest corner detection algorithm. The initial procedure was modified to achieve insensitivity to object scaling and pixel noise. We used a Gaussian filter for removing high frequency, random image contents. Key point detection is performed for an original size photo and scaled down by a factor of two. Finally, a descriptor is calculated for each of the detected key point.

The key-point descriptors are calculated on the basis of local image gradient magnitudes and orientations. Key points' descriptor is based on gradient magnitudes computed for 16 or 4 pixels adjacent to a keypoint. These values are used to form the so called edge orientation histogram. A gradient magnitude m and a gradient orientation θ of each pixel is given by the formula:

$$m(x,y) = \sqrt{(I(x+1,y) - I(x-1,y))^2 + (I(x,y+1) - I(x,y-1))^2} \quad (1)$$

$$\theta(x,y) = \text{atan2}\,(I(x,y+1) - I(x,y-1), I(x+1,y) - I(x-1,y)) \quad (2)$$

$$\text{atan2}\,(x,y) = \begin{cases} \arctan(x/y) & x > 0 \\ \arctan(x/y) + \pi & y \geq 0, x < 0 \\ \arctan(x/y) - \pi & y < 0, x < 0 \\ +\frac{\pi}{2} & y > 0, x = 0 \\ -\frac{\pi}{2} & y < 0, x = 0 \\ \textit{undefined} & y = 0, x = 0 \end{cases} \quad (3)$$

where: $I(x,y)$ – brightness value of a pixel at x,y coordinates,
atan2(x,y) - function defined in Java language library Java.lang.Math.

The mode of the histogram indicates the main edge orientation in the tested image fragment. To achieve insensitivity to rotation these values are shifted by the orientation value. Example descriptor for 2x2 regions for the selected keypoint from Fig. 3 is shown in Fig. 4.

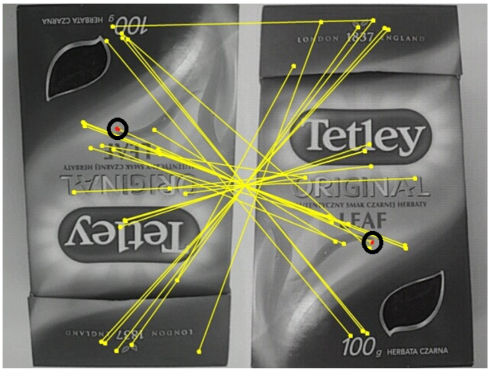


Fig. 3 Example of a key point descriptor calculated for 2x2 regions around the keypoint (marked by circles), the Y axis represents gradient magnitude, the X axis represents orientation bins

The detected keypoints and their parameters are used either as descriptors of a new object in a database or can be matched to earlier defined database pattern objects. A single object can feature more than one database entries that contain parameters representing different projection views of the same object, e.g. different textures, colours or shapes.

We propose a new method of keypoint orientation assignment in comparison to earlier work [Matusiak and Skulimowski 2012]. For each key point, a neighbourhood is defined by the use of a Gaussian weighting function. Values on the edge orientation axis of the histogram are calculated with unit degree precision, this is a more precise approach than using histogram bins, as was originally proposed in [Lowe 1999]. However, this method implies more complicated procedure of finding the main orientation of the key points.

A search with a predefined range is performed, in this step of the algorithm, to determine the highest total sum of the gradient magnitude. This range is set to 15 degrees and, moreover, the magnitude values are multiplied with the Gaussian weighting function. This function assigns the highest weighting factor to the orientation located in the centre of the analyzed range and is decreasing away from the centre. This type of weighting has proved to be efficient in a more precise detection of the major edge orientation for the analysed key point. The described procedure rotates the search window from 0 to 359 degrees.

A search with a predefined range is performed, in this step of the algorithm, to determine the highest total sum of the gradient magnitude. This range is set to 15 degrees and, moreover, the magnitude values are multiplied with the Gaussian weighting function. This function assigns the highest weighting factor to the orientation located in the centre of the analyzed range and is decreasing away from the centre. This type of weighting has proved to be efficient in a more precise detection of the major edge orientation for the analysed key point. The described procedure rotates the search window from 0 to 359 degrees.

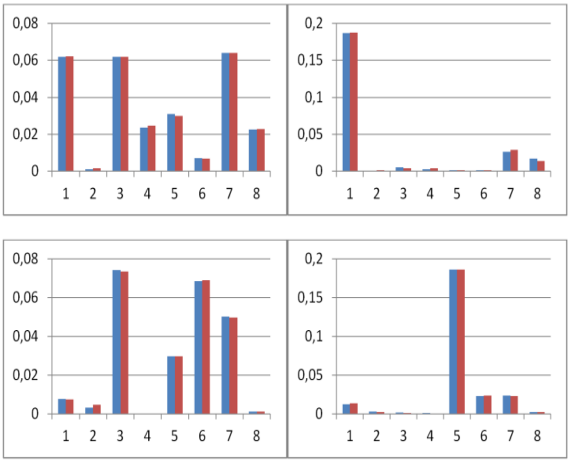


Fig. 4 Comparison of matched key points descriptors from Fig. 3, the Y axis represents gradient magnitude, the X axis represents orientation bins

3.4 Classification and Matching the Keypoints

In our previous work we show [Matusiak and Skulimowski 2012], that it is possible to find pairs of corresponding descriptors in the analyzed image, with high similarity measure, but they turned out to correspond to different object's points. That is why estimation of object pose (rotation) was added to the proposed classification procedure. Consequently, pairing the keypoints describing objects from the database and the keypoints describing newly scanned objects can produce many false positive recognition results.

In the algorithm proposed here, for each pair of the descriptors of the keypoints that were paired using a similarity measure, a difference of main orientations with precision to unit degree is calculated. On the basis of this data the application builds a histogram showing frequency of occurrence of all rotations. By using this histogram the most frequent difference is defined. As a result the most probable change in rotation of the object in a scene can be determined. Pairs of the classified key points that feature different rotation are treated as invalid and excluded from the recognition algorithm (we assume the object is stiff). Moreover, in the classification procedure information gathered from the statistics of keypoint matching procedure is used to decrease influence of wrongly detected keypoints on the final object recognition result. This procedure is implemented by setting a threshold to the number of matched keypoint pairs on which recognition can be based. If the algorithm cannot find sufficiently high number of keypoints' pairs for the given template, that template is excluded from the recognition algorithm. Another threshold that we use is a number of key points from the test image for which a particular keypoint is the nearest neighbour in the descriptor feature space. In an ideal situation one key point from a database should be the nearest neighbour for only one key point in the test image (one-to-one mapping of the keypoints).

Fig. 5 shows examples of classification results. Red lines (or darker lines in a greyscale image version) join pairs of the descriptors that were rejected after the proposed verification steps. Yellow lines (or brighter lines in a greyscale image version) describe pairs used in the classification procedure.

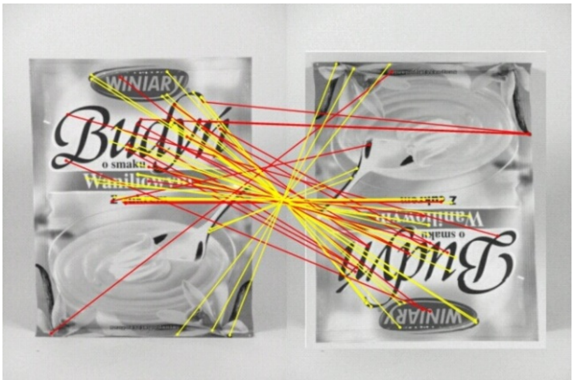


Fig. 5 Results of correct classification of the test object (right) to the object stored in a database (left); note rotation of the test object by 178 degrees

3.5 Auditory Presentation

Standard graphical user interface (GUI) elements of Android operating systems make application too complex for the blind. What is more, screen reading software usually sonify all standard GUI elements by generating unnecessary sound information. We have proposed custom user interface with built-in sonification, which is implemented using Android libraries. Application screen is divided into five sections, which correspond to different functionalities. When the user touches a section or moves between them a proper audio information is sounded. The functions are activated when the user releases his touch on a screen.

Such an approach allows to generate dedicated messages which are freed from information overload as complained by the blind users of ordinary screen readers. The view of the proposed user interface is shown in Fig. 1.

4 Tests of the Applications

The application was implemented in Java programming language for Android OS with the use of Eclipse IDE. The object recognition algorithm was tested in terms of performance and quality of results. We used HTC Desire HD, Samsung Galaxy S3 and Sony Xperia S smartphones and Motorola Xoom M602 tablet with good quality camera and computational capabilities, but also a low cost model HTC Explorer. Comparison of these mobile phones' hardware is shown in Tab. 1. Object recognition test was divided into two phases. For the test purposes of the first phase a database of 30 photos of food products' packages was collected and saved in a mobile phone memory. Each photo contains a front side of products' package and was taken at daylight conditions. During the test we took also 19 photos at different rotations and positions of one of the objects that is pre-stored in a database. Results of the preliminary tests of the object recognition application are as follows (on a HTC Desire HD smartphone): correct recognition rate: 89%, false recognition rate: 5,5%, no object recognized: 5,5%.

Table 1 Comparison of hardware architecture of mobile phones used in tests

	HTC Desire HD	Sony Xperia S	HTC Explorer	Samsung Galaxy S3	Motorola Xoom M602
CPU model	Qualcomm MSM8255	Qualcomm MSM8260	Qualcomm MSM7225A	Samsung Exynos 4412	Nvidia Tegra 250 T20
CPU frequency	1 GHz	1,5 GHz Dual core	600 MHz	1,4 GHz Quad core	1,0 GHz Dual core
CPU architecture	Qualcomm Scorpion	Qualcomm Scorpion	ARM Cortex-A5	ARM Cortex-A9	ARM Cortex-A9
RAM size	768 MB	1 GB	512 MB	1 GB	1GB

The second phase of tests was carried out for a larger database comprising 40 objects. We took 15 test photos in poor, artificial light conditions and not obvious test objects. Some of them had visual elements similar to more than one object from the database, like brand logo or name. Also the objects had additional information, e.g. price reduction stickers, modified design or different version of the same product. For this more challenging test the results are as follows: correct recognition rate: 74%, false recognition rate: 6%, no object recognized: 20%.

The main reason for poorer classification performance in this test run was the appearance of the same brand logo on several products in the database because the detected keypoints in the test photo were within the region of the brand logo.

During tests we also gathered data for measurement of algorithm performance on different mobile devices. We calculated two values: average time of class template creation and comparison time of the classification process. The results are summarised in Tab. 2.

Performance of the application highly depends on hardware capabilities, especially CPU unit. It can be noticed that the number of cores has more impact than clock frequency. Overall performance disproportion between the outermost results is noticeable by user.

The main problem in a practical use of the colour detection module was caused by varying lightning conditions, i.e. when the camera flash must be used during photo capture. In good lighting conditions colour detection performance was voted adequate by blind users of the application. Also the usability of the light direction detector has been positively appraised by visually impaired users. The application allows to locate the light source after a short time delay needed by the built-in automatic procedure of image capture settings of the camera.

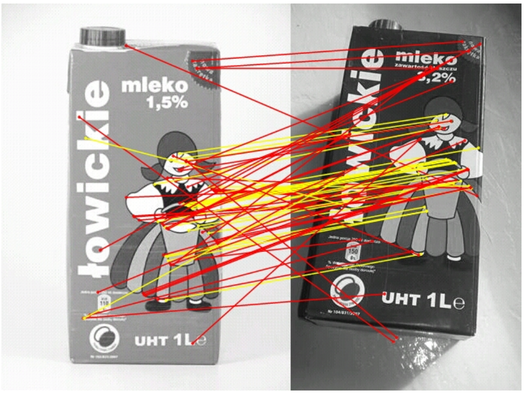


Fig. 6 Example of proper classification of an object with a modified coverage design

Table 2 Comparison of time performance of the image recognition application

	HTC Desire HD	Sony Xperia S	HTC Explorer	Samsung Galaxy S3	Motorola Xoom M602
Template calculation time [s]	3,72	2,76	5,51	1,75	2,81
Comparison time [s]	40,31	30,20	61,3	21,36	30,51
Comparison time per class [s]	1,01	0,75	1,53	0,53	0,76

5 Summary

The colour detector, the light direction detector and the object recognition applications were developed for the visually impaired users. The software tools were implemented and tested on the following mobile devices: HTC Desire HD, HTC Explorer, Sony Xperia S, Samsung Galaxy S3, Motorola XOOM MZ601. The conducted tests of the applications have revealed that the performance of the algorithms strongly depends on the quality of the built-in camera and image acquisition lighting conditions. Correct classification rate varies between 75-89% and depends on the photo quality. The object recognition procedure run on the fastest smartphone took about 20s for a database of 40 templates. The application is equipped with a custom sonified user interface. Future work will focused on the performance improvement by using a native code and comparison to similar applications based on cloud computing.

Acknowledgments. The project is financed in part by the National Centre of Research and Development of Poland in years 2010–2013 under the grant NR02-0083-10.

References

[Gill 2008] Gill, J.: Assistive devices for people with visual impairments. In: Helal, S., Mokhtari, M., Abdulrazak, B. (eds.) The Engineering Handbook of Smart Technology for Aging, Disability, and Independence, pp. 163–190. J. Wiley and Sons (2008)

[Hersh and Johnson 2008] Hersh, M., Johnson, M.: Assistive technology for visually impaired and blind people. Springer, London (2008)

[Lowe 1999] Lowe, D.: Object recognition from local scale-invariant features. In: Proc. Inter. Conf. on Computer Vision, pp. 1150–1157 (1999)

[Matusiak and Skulimowski 2012] Matusiak, K., Skulimowski, P.: Comparison of key point detectors in SIFT implementation for mobile devices. In: Bolc, L., Tadeusiewicz, R., Chmielewski, L.J., Wojciechowski, K. (eds.) ICCVG 2012. LNCS, vol. 7594, pp. 509–516. Springer, Heidelberg (2012)

[Nanayakkara et al. 2012] Nanayakkara, S.C., Shilkrot, R., Maes, P.: EyeRing: an eye on a finger. In: Proc. Inter. Conf. Human Factors in Computing, pp. 1048–1050 (2012)

[Onishi and Ono 2011] Onishi, J., Ono, T.: Contour pattern recognition through auditory labels of Freeman chain codes for people with visual impairments. In: Proc. International IEEE Systems, Man, and Cybernetics Conference, Alaska, USA, pp. 1088–1093 (2011)

[Rosten and Drummond 2006] Rosten, E., Drummond, T.: Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part I. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006)

[Strumillo 2012] Strumillo, P.: Electronic navigation systems for the blind and the visually impaired. Lodz University of Technology Publishing House (2012) (in Polish)

[Tinnakorn and Punyathep 2011] Tinnakorn, K., Punyathep, P.: A voice system, reading medicament label for visually impaired people. In: Proc. of RFID SysTech, 7th European Workshop on Smart Objects: Systems, Technologies and Applications, Dresden, Germany (2011)

[WWW-1 2013] Gartner says worldwide sales of mobile phones, `http://gartner.com/newsroom/id/2573415` (accessed September 11, 2013)

[WWW-2 2013] LookTel recognizer, `http://www.looktel.com/recognizer` (accessed February 25, 2013)

Mobile Melanoma Diagnosing System – A Preliminary Attempt

T. Mroczek

University of Information Technology and Management, Rzeszów, Poland
tmroczek@wsiz.rzeszow.pl

Abstract. The paper presents a new strategy for noninvasive diagnosis of melanocytic skin lesions which differs from European research focused mostly on methodology of classification of tumors, description of some symptoms and pigment changes in the phase of incurable disease or in the disease requiring a surgical intervention. The new method is based on applying visual methods presenting the parameters describing the disease as well as on methods of automatic visual analysis in the process of extraction of lesion features. The proposed system utilizes two new algorithms: hybrid synthesis of medical images and automatic analysis of digital images.

1 Introduction

During the last fifty years more and more people are affected by melanoma. The largest numbers of melanoma cases are observed in Australia (between 13.6 and 28.9 percent of men and 18.0 and 25.3 percent of women) and in the United States (12.4 percent of population). In Europe most cases of melanoma are in Norway (10.5 percent of men and 13.5 percent of women). Additionally, as follows from European statistics, the number of melanoma cases grows about ten percent yearly in Europe [WWW_1 2013].

Unexpectedly, melanoma diagnosticians observe that melanoma is fatal most frequently for people who do not consider themselves to be in the risk class. However, melanoma diagnosed early is curable, in spite of numerous statistical reports to the contrary [WWW_1 2013; WWW_2 2013]. Australia is an example of a country where 90% of cases with a Clark level I and II melanoma are diagnosed due to a high awareness of the society and attention of physicians for early diagnosis and treatment of the skin pigment changes [Breslow 1970].

It seems that the societal awareness for consequences of melanoma early diagnosis is still not sufficient, especially in European countries. Insufficient melanoma awareness in Poland was confirmed by a test conducted on more than a thousand adults in April of 2011. One quarter of adults was aware of a lesion on their skin that may be dangerous. But the question “Have you been going to a physician only to test the lesion whether it is dangerous?” was negatively answered by three quarters of people from the group that observed worrisome lesions.

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_18,

Melanoma is very dangerous in spite of so many sources of information about symptoms and an access to the methods of computing parameters describing the degree of risk. These methods are based on atypical changes of skin pigment. Description of such changes is based, in turn, on difficult concepts, so that such materials are not easily comprehensible for an average person. Moreover, parameters describing lesions are familiar only to researchers involved in oncology.

Taking all of that into account, it is clear that we should increase the level of societal awareness through popularization of methods of noninvasive lesion diagnosis. These methods should be user-friendly and comprehensible for all. In order to reach this goal a new strategy is proposed for explaining the parameters typical for melanocytic skin lesions. This strategy applies mostly mobile multimedia visual technologies including handsets, smartphones and tablets. The concept presented in this paper is: (**i**) an evolution of the previous research on methods of skin lesion diagnosis conducted by an extended team [Hippe and Mroczek 2003; Blajdo et al. 2004; Grzymala-Busse et al. 2007] and (**ii**) developing of a new approach to noninvasive diagnosis of melanocytic skin lesions. Thus in the next section initial assumptions of the mobile system are presented. Whereas the general description of the system is presented in Sec. 3.

2 Initial Assumptions

The proposed methodology is based on applying visual methods presenting parameters describing the disease as well as on methods of automatic visual analysis in the process of extraction of lesion features. It is extremely important that in this approach it is possible to acquire and store images of melanocytic skin lesions directly from the particular users of the system. It is exceptionally significant since from some time an increasing research interest is observed in images presenting early stages of melanoma. At the same time there is lack of professional computer data bases containing images of such lesions. Well known centers of medical informatics, offering until recently an access to such data, discontinued (maybe temporarily) an access to this service [WWW_3 2013]. Other centers offer standard, always the same set of lesion images, with commercial use in mind [WWW_4 2013]. Standing law on protection of private information causes lack of access to specialized data sets containing such images and difficulties in creating such new databases. The current interpretation of that law imposes the necessity of a patent’s consent on creating and storing of an image. In this approach, in which creation of the skin lesion image by a mobile devices and then automatic analysis of the image, is performed. The user is also asked for the permission to store image for the purpose of forming a reliable database on the melanocytic skin lesions. The resulted database can contain not only simulated images but also digital images of actual lesions. The approach improves the quality of classification process, provides better explanation of properties of classified objects, and it makes possible to do new research.

Additionally the implementation of a mobile version of the system through a specialized store of mobile applications is proposed. Thus society, especially young people, excessively using tanning, will be more aware on a risk of being affected by melanoma. One of deliveries is a free, mobile system called **MMDS** (Mobile Melanoma Diagnosing System) and designed for noninvasive diagnosis of the level of the danger that the skin lesion is affected by melanoma.

To implement the strategy of explaining the parameters characterizing melanocytic skin lesions the creation of a new methodology of a computer based system supporting medical diagnosis is needed. This system is based on two new algorithms: hybrid synthesis of medical images and automatic analysis of digital images. The MMDS contains a few computer program modules. The most important features of these modules are explained in the next section.

3 Description of the MMDS System

3.1 Server Part

The system uses the client-server architecture. The server part, presented on Fig. 1, is designed to cooperate with the client of the system (a mobile application). Its central part is the Control Module, whose the main task is to control of the information flow between remaining elements of the system.

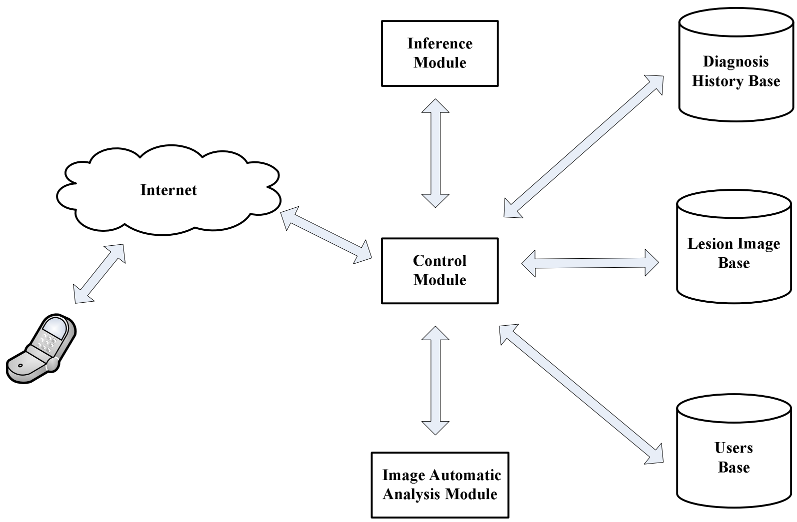


Fig. 1 A server part of the MMDS system

The classification of melanocytic skin lesions and the evaluation of the degree of disease progress is conducted in the **Inference Module**. In previous research, conducted by an extended team, five methods of skin lesion diagnosis have been developed: (i) the standard **ABCD** rule (based on the **TDS** parameter), (ii) an optimized ABCD rule (based on own parameter New TDS), (iii) a decision tree (based on the ID3 algorithm [Grzymala-Busse et al. 2004], (iv) a genetic dichotomization (based on linear learning machine in genetic search of the most important attributes) and (v) belief networks (based on the K2 algorithm) [Hippe and Mroczek 2003]. All of these classification methods were contained in the additional algorithm of evaluation and voting implemented in the **IMDLS** (Internet Melanoma Diagnosing and Learning System) [Hippe et al. 2004, Hippe et al. 2005]. The Inference Module includes such reliable and robust classification methods.

Another important module of the system is the **Image Automatic Analysis Module** in which a new approach to extraction of important features of melanocytic skin lesions is implemented. In this innovative module in the first step of analysis the acquired images are cleaned (removal of noise and normalization) and then image segmentation is conducted. The image is partitioned into regions that are homogenous with respect to selected properties.

The classification of the melanocytic skin lesions is based on the identification of some characteristic features or symptoms. Among methods used most frequently by dermatologists [Menzies 1997; Argenziano et al. 1998; Stolz et al. 2006] is the Stolz's method, based on the ABCD rule. According to this rule, four parameters are evaluated: Asymmetry (A), Border (B) where the lesion border is evaluated, Color (C) denotes the number of different colors, out of 6, present in the lesion and Diversity of structures (D), one or up to 5. The main purpose of the extraction process is an application of the corresponding algorithms for determining these four parameters. The preliminary research contributed to developing new, effective methods for extraction of Asymmetry, Border and Color [Cudek et al. 2011]. Currently research for extraction of Diversity is conducted. New, unique methods of verifying the lesion diversity are being tested. All discussed features of the lesion are transferred to the Inference Module.

Additionally, the server part is equipped with three data bases: **Diagnosis History Base**, **Lesion Image Base**, and **User Base**, where the lesion features, inputted from the client application or acquired in the process of image analysis, together with the corresponding results of diagnosis, the lesion images inputted from the client application, and the accounts of users of the MMDS system are stored, respectively.

3.2 Client Part

The next main part of the system (the client part) is designed as an application for mobile devices. The structure of this part of the MMDS is presented on Fig. 2.

The method of data acquisition presented by the **Data Acquisition Module** is based on semantic conversion of text data describing four main features (*Asymmetry*, *Border*, *Color* and *Diversity*) into digital images. The main principle used here is a mixture of (i) procedures of synthesis used in vector graphics, (ii) procedures of raster graphics and (iii) elements of supervised machine learning. In a nutshell: synthesis procedures of medical images enable so called *controllable insertion* of templates into synthetic lesion images. Such templates (so called *pre-defined textures*) that contain fragments of real-life lesion images are prepared a *priori* by applying supervised machine learning. Such a synthesis algorithm was implemented in own information technology tool (called ***Image*Synthesizer**) [Hippe et al. 2009]. This system synthesizes static melanocytic lesion images that are the most dangerous, i.e., from the groups Nevus and Melanoma, including all special subgroups of these lesions. The efficiency of that algorithm was confirmed by a conducted test. In this test (i) a set of real-life melanocytic skin lesions created by using a digital camera was compared with (ii) a set of synthetic images created by this algorithm. Independent experts representing dermatology and medical imaging evaluated the algorithm quality. Experts were trying to guess which images were real-life (created by a digital camera). A detailed description of the algorithms and results of conducted experiments were published in [Hippe and Piatek 2007, Hippe et al. 2008].

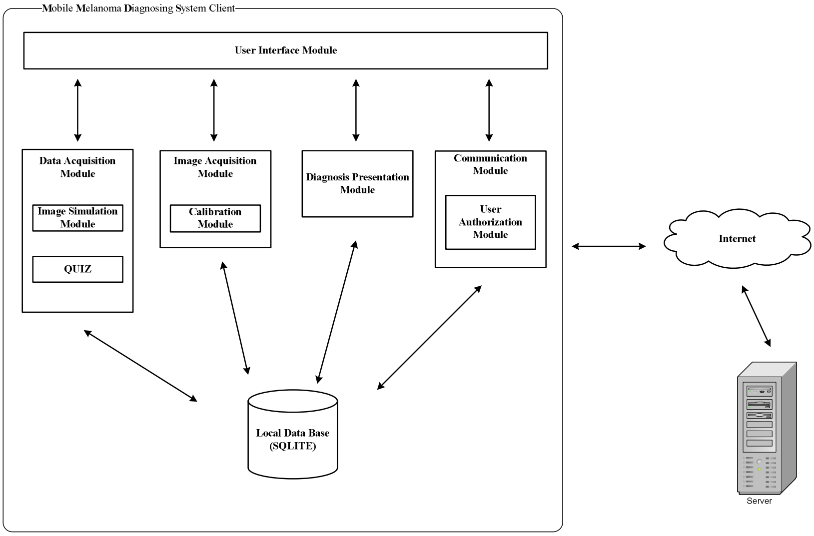


Fig. 2 A client part of the MMDS system

The input data, in the form of digital images, are acquired using mobile devices. There are some restrictions since the image is stored in a small device. Hence such a camera must be miniature, so the image is of poor quality. To improve resolution of small photo-sensitive matrices their sensors are small as well, the restriction here is light diffraction. The smaller the sensor is, the less energy is coming to it and the more noise is present in the signal, as it is visible when the picture is enlarged. Taking this into account, the Data Acquisition Module is an alternative to the **Image Acquisition Module**. On the other hand, among many available mobile devices, some of them are of a sufficient standard, being functionally close to digital cameras. These digital images are quite adequate for a corresponding analysis.

Taking into account the heterogeneity of images created by mobile devices and the way a picture was taken, the **Calibration Module** has been designed. This module contains information about objects, methods of measurement, and results of measurement. Additionally, this module is able to transform results of measurement by applying some mathematical formulas. Thus the final output is determined by the system that automatically calibrates the results and the outputted parameters of the images of melanocytic skin lesions are standardized.

The results of diagnosis are presented in the form of a text on the monitor of the mobile device. To make the system more efficient, especially when many users use the system at the same time, it is assumed that the process starting from the request for diagnosis and ending with the result of diagnosis is asynchronous, i.e., the response of the server may be delayed. Therefore, the **Diagnosis Presentation Module** is equipped with a notification system. Additionally, the user receives an e-mail with detailed information about results of diagnosis.

The **Communication Module** passes the information between a mobile client application and the MMDS server, in order to (i) transfer images and features of melanocytic skin lesions stored in the **Local Data Base**, (ii) transfer results of diagnosis and (iii) conduct user authorization. Universal Internet services and secured transfer of information is used for this purpose. Thus, the Internet, independently of localization and details of implementation, connects all program components. Due to such solution the system is able to cooperate with all kinds of clients regardless of the target software platform.

The user is an integrated part of the information process. It is crucial that the user accepts the system and that the system is user-friendly. The success of the system depends on the way user interacts with the system. It is important that the system is equipped with the **User Interface Module** that presents results graphically and is a source of dynamic access to data. Such interface has to be intuitive and simple, with the assumption that the user does not need any technical preparation. The system works with a client that uses a touchable screen as well. Screens of mobile devices are usually small, so only essential information is displayed. The **Image Simulation Module** is designed for users who may input graphical images in a very simple and intuitive way. Obviously, the user does not need to know anything about the corresponding medical area. There is also no need for a sophisticated equipment.

4 Conclusion

The solution described in this paper is based on a novel strategy in which the user receives information about melanocytic skin lesions through a sophisticated system, acting behind the scene and using specialized algorithms and a well designed structure.

As it was motioned in previous sections, the system utilizes two new algorithms: hybrid synthesis of medical images and automatic analysis of digital images. The first algorithm was implemented and tested. The results showed that the synthetic images, generated by means of the developed algorithms, can be successfully used as an alternative source of knowledge in relation to digital medical images. Whereas the results of the application of the primary version of the algorithm for automatic analysis of digital images are promising, the final implementation still requires additional work to be done.

It is worth to mention that the concept of the system briefly described here can be applied for the construction of other diagnostic systems preventing dangerous diseases and also in various domains of science and technology.

References

[Argenziano et al. 1998] Argenziano, G., Fabbrocini, G., Carli, P., De Giorgi, V., Sammarco, E., Delfino, M.: Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions. Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. Arch Dermatol. 134, 1563–1570 (1998)

[Breslow 1970] Breslow, A.: Thickness, cross-sectional areas and depth of invasion in the prognosis of cutaneous melanoma. Ann. Surg. 172(5), 902–908 (1970)

[Blajdo et al. 2004] Blajdo, P., Grzymala-Busse, J.W., Hippe, Z.S., Knap, M., Marek, T., Mroczek, T., Wrzesień, M.: A suite of machine learning programs for data mining: chemical application. In: Debska, B., Fic, G. (eds.) Information Systems in Chemistry, vol. 2, pp. 7–14. University of Technology Editorial Office (2004)

[Cudek et al. 2011] Cudek, P., Paja, W., Wrzesień, M.: Automatic System for Classification of Melanocytic Skin Lesions Based on Images Recognition. In: Czachórski, T., Kozielski, S., Stańczyk, U. (eds.) Man-Machine Interactions 2. AISC, vol. 103, pp. 189–196. Springer, Heidelberg (2011)

[Grzymala-Busse et al. 2004] Grzymala-Busse, J.W., Hippe, Z.S., Knap, M., Mroczek, T.: A New algorithm for generation of decision trees. TASK Quarterly 8(2), 243–247 (2004)

[Grzymala-Busse et al. 2007] Grzymała-Busse, J.W., Hippe, Z.S., Mroczek, T.: Deriving Belief Networks and Belief Rules from Data: A Progress Report. In: Peters, J.F., Skowron, A., Marek, V.W., Orłowska, E., Słowiński, R., Ziarko, W.P. (eds.) Transactions on Rough Sets VII. LNCS, vol. 4400, pp. 53–69. Springer, Heidelberg (2007)

[Hippe and Mroczek 2003] Hippe, Z.S., Mroczek, T.: Melanoma classification and prediction using belief networks. In: Kurzyński, M., Puchała, E., Woźniak, M. (eds.) Computer Recognition Systems, pp. 337–342. Univ. of Technology Edit. Office (2003)

[Hippe and Piatek 2007] Hippe, Z.S., Piatek, L.: Synthesis of Static Medical Images – an Example of Melanocytic Skin Lesions. In: Computer Recognition System 2. Advances in Soft Computing, vol. 45, pp. 503–509 (2007)

[Hippe et al. 2005] Hippe, Z.S., Grzymala-Busse, J.W., Blajdo, P., Knap, M., Mroczek, T., Paja, W., Wrzesien, M.: Classification of medical images in the domain of melanoma skin lesions. Advances in Soft Computing 30, 519–526 (2005)

[Hippe et al. 2008] Hippe, Z.S., Grzymala-Busse, J.W., Piatek, L.: Synthesis of medical images in the domain of melanocytic skin lesions. In: Information Technologies in Biomedicine. Advances in Soft Computing, vol. 47, pp. 225–231 (2008)

[Hippe et al. 2009] Hippe, Z.S., Grzymała-Busse, J.W., Piątek, Ł.: Synthesis of static medical images with an active shape model. In: Cyran, K.A., Kozielski, S., Peters, J.F., Stańczyk, U., Wakulicz-Deja, A. (eds.) Man-Machine Interactions. AISC, vol. 59, pp. 429–434. Springer, Heidelberg (2009)

[Menzies 1997] Menzies, S.W.: Surface microscopy of pigmented skin tumors. Australas. J. Dermatol. 38, 40–43 (1997)

[Stolz 2006] Stolz, W., Braun-Falco, O., Bilek, P., Landthaler, M., Burgdorf, W., Cognetta, A.: Atlas of Dermatoscopy. Czelej Edit. Office, Lublin (2006)

[WWW_1 2013] `http://www.cancer.org/Cancer/CancerCauses/SunandUVExposure/skin-cancer-facts` (accessed October 4, 2013)

[WWW_2 2013] `http://skincancer.about.com/od/skincancerbasics/a/Melanoma.htm` (accessed October 4, 2013)

[WWW_3 2013] `http://www.dermoncology.com/` (accessed October 6, 2013)

[WWW_4 2013] `http://www.pless.fr/dermatoscopie/` (accessed October 6, 2013)

User-Centered Design and Older People: Introductory Considerations for Engaging Users

C.J. Soraghan[1,2,3], S. Hermann[4], and G. Boyle[1,2]

[1] Department of Medical Physics & Bioengineering, St. James's Hospital, Dublin, Ireland
[2] Mercer's Institute for Successful Ageing, St. James's Hospital, Dublin, Ireland
[3] Department of Medical Gerontology, School of Medicine, Trinity College Dublin, Ireland
[4] Trinity Centre for Bioengineering, School of Engineering, Trinity College Dublin, Ireland
{Csoraghan,gboyle}@stjames.ie, hermanns@tcd.ie

Abstract. Industry and researchers alike are becoming more aware of the need to put the user at the centre of their design models. At each stage the user should be included and consulted in order to design more acceptable, appropriate, and useful technologies. As engineers, we tend to develop technology in a lab to a specification that we assume is useful and perhaps even adhere to standards and guidelines. In this chapter, the lead engineer reports on experiences from a study assessing 'universal design' of technology in the homes of older people. These findings are largely experiential and are presented in order to foster an appreciation of the context of older people's use of technology in the home; to encourage other designers and engineers to come outside of their traditionally prescribed scope and enter the realities of the older person's home to assess technology use and germinate empathic design; and to outline practical methods and approaches as to how to implement home interviews and technology assessments.

1 Introduction

The role of researchers can be subdivided into many disparate fields. However, in practice, research projects may have limitations in the extent of the multidisciplinary nature of the group due to funding limitations and personnel shortages where researchers will need to acquire and exercise skills from different fields. This is especially the case of the engineer. Moreover, in order to appreciate the real needs of older adults and to design better products, engineers and designers need to empathize with older adults to understand the context of use of any potential product, and to ensure that requirements are being met. This could be done by observing and including the user (such as in user-centered design). As such, the aim of this research was to propel the engineer from the comfort of the lab to the realities of the older user's home. By doing this, the lead author, a clinical engineer working at a hospital, gained a number of insights and experiences in how to engage the older user. It is hoped that this chapter can provide encouragement for other engineers to carry out their own in-home assessments of technology and to assist and enable other engineers/designers in doing this by way of the experiences and guidelines outlined by the authors.

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_19,

In the next section the authors report on some background of user-centered design for older people and of the importance of in-home assessments. Section 3 describes the research study context and methodology out of which the experiences (listed in Section 4) of the engineer were borne, with some discussions and final conclusions given in Section 5 and Section 6, respectively.

2 Background: User-Centered Design for Older People

User-centered design (UCD) considers the perspective of the user throughout all stages of the design of a technology. Many technologies designed for older people have been produced which did not necessarily include the point of view of the end-user of the product at all - and as a result many technologies are unusable for older people. There has been a move towards “inclusive design”, “universal design”, “participatory design” and other methods of empathizing with the older user in product design. However, the literature has noted that methods of engagement with older people in this process are not directly amenable to standard user-centered design techniques [Eisma et al. 2004, Dickinson et al. 2007]. That is, the process of using certain UCD techniques with older people as explored by the same research group was said to have a confounding affect on their ability to complete the assigned task, particularly in those with issues in cognition, processing capacity and functional limitations as a result of ageing (vision, hearing, dexterity). They [Dickinson et al. 2007] demonstrated how one could engage older users using adaptations of UCD but these have noted that still, the most efficient methods for gaining requirements from the older users have been one-to-one conversations with them [Dickinson et al. 2007]. However, these one-to-one encounters may not be suitable for large cohort studies and there is still merit in self-reporting independent questionnaires as a means to gain an understanding of user-perspectives [Dickinson et al. 2007]. Nevertheless, to understand older people’s real experiences with technology and to get true requirements from the older user group in user-centered design there is perhaps a need to discern between self-report and what actually happens when a user uses a technology. This in-situ method was the approach taken in this work. It also serves a purpose of immersing the research group into real-world complexities surrounding context of use of technology and develops a much needed rapport with the community for long-term relationships for enabling further research into the future.

In this chapter, we wish to report on the introductory considerations that should be made for engaging with older people in their homes in user-centered design and/or universal design research. There are some guidelines in the literature as to how researchers can design the field work, e.g. [Dickinson et al. 2007], and so this current work adds to the knowledgebase. Firstly, the authors describe the methodology and the context for the research that was carried out, so that the reader can appreciate the setting of where the experiences came from, which are explored in Section 4.

3 Research Study Context and Methodology

The Question

The authors wished to explore the real technology needs of older adults in their homes and to gain an appreciation of their home/community social and environmental contexts which tends to influence the appropriateness and effectiveness of certain technologies for them. For this we took a universal design approach to investigate the level to which technologies in their homes were 'universally designed'. A series of in-home interviews were used to explore this theme. The results of the study and questionnaire design (adapted from a number of existing questionnaires [Beecher et al. 2005; Lenker et al. 2011; The Centre for Universal Design 2000; The Centre for Universal Design 2002] will be presented elsewhere. In this current chapter, the authors wish to discuss the experiences of the lead author (engineer) doing this assessing of technology in the homes of older people, again, as an encouragement and resource for other engineers doing similar in-home technology assessments.

The Participants and Their Community

33 older people (age range 65-92 years old) were recruited for the study: 30 female and 3 male. All participants were based in the Liberties Area of Dublin 8 (inner city). This area of Dublin is very old and known to be socioeconomically deprived [WWW_1 2011]. Housing was typically small to medium red brick terraced houses with some low rise local authority flat complexes. Some examples of the area assessed are given in Fig. 1.

The author[1] was not previously familiar with the Liberties Area. He found that there was a 'local' friendly atmosphere with vibrant selling in street markets. Although there were known drug abuse and related activities in some of the areas visited, generalizations were challenged by the apparent neighborly relationships between local residents and local achievements in the area - e.g. a prize winning garden. It was found to be important to walk the area and talk to people in the regions near the participant's homes. This provided useful insight and empathy for older people, seeing them as people living in social and environmental contexts which may negate the use of certain technologies or approaches. Rather than seeing them as generic "older technology users", other relevant considerations were aroused which may affect technology use, such as security or the insufficiency of access facilities, for example.

Most participants lived at home alone and worked in manual labour before retiring or worked in the home (housewives). In general, they had limited exposure to technology except common domestic devices such as radio, CD/DVD players, TV, set top boxes, mobile phones, laptops (in some cases), smartphones (in a limited number of cases), and kitchen and domestic appliances.

[1] In this chapter, when the singular term "author" is used it refers to the 1st author of this text.

Fig. 1 Participants community environment context: Liberties Area in Dublin 8, Ireland

The Researchers

A team of 3 researchers took part in the interviews which were carried out in the homes of each of the older people recruited. These consisted of an engineer (1^{st} author) working at a hospital, a local community development worker, and an ergonomics specialist (2^{nd} author). The experiences described later in this chapter are principally those from the point of view of the lead researcher (1^{st} author) who led all interviews. He has a background in Electronic Engineering with a PhD in Biomedical Engineering Design and now operates as a postdoctoral research fellow in technologies for older people. This role includes the review and design of technology for older people.

The Interviews

The research team visited the homes of 33 older participants, and interviews were carried out in each home using a bespoke 'Universal Design' survey tool with questionnaires, along with observational studies of the older users using different technologies (for example, see Fig. 2) in their homes.

Fig. 2 Example of technologies assessed during the interviews

The recruitment was such that each participant had a social alarm (aka pendant alarm) which could be used to request help by pressing a button on the device. When the button is pressed it transmits a radio-frequency signal picked up by a telephony system connected to a phone-line which calls a monitoring center who then calls the user back to see what is wrong - or calls other predetermined contacts if there is no answer and finally emergency services if all other avenues fail or if the situation requires it. All participants had received the pendant alarm from a government funded scheme in Ireland. The interviews assessed this pendant alarm technology and one other technology in the user's home that they reported was difficult-to-use (either directly reported or discovered from getting them to use a number of devices in their home). The questionnaire was used in combination with open conversation to elicit other contextual information and experiences of the older user. The user was asked to carry out a number of familiar/common tasks with the devices (send a text message on a mobile phone; defrost something in the microwave; tune in a TV station; change the temperature on an electric radiator) while concurrently the interviewer asked questions and made observations of the user's interaction and use of the technology. This was facilitated and supported by the survey tool that was developed for this study.

4 Considerations for Engaging with Older Users In-Situ

Table 1 below outlines the main considerations that the engineer found from in-home assessments. Many of these quashed preconceptions and assumptions that the engineer had about older people and provide rich context of use of technology.

They are listed in chronological epochs - from recruitment stage, to the pre-interview stage, right through to the post-visit stage. These are discussed in more detail in [Soraghan et al. 2013].

Table 1 Main considerations for engaging older adults in interviews in their homes

Consideration	Description
RECRUITMENT	
Sensitive recruitment	- Cold calling older people can be insensitive
	- Use an empathetic local community group
Avoid Alarm	- Fronting as a health centre can be alarming
	- Have a local community group front the project
Living alone, but not at home	- Many over 65 have very active lives
	- Don't expect it to be easy to contact people when trying to recruit them
Dynamic recruitment	- Meeting recruitment numbers may be difficult
	- Use dynamic means: word of mouth and post interview requests (ask the interviewee if they know anyone eligible)
	- Important to leave a positive sentiment for future recruitment and dynamic recruitment
Conveying the message	- It can be difficult to convey the research purpose to participants, even with detailed instructions
	- Work with a community worker to rephrase materials; evolve terminology to be user-friendly; blend details into the vernacular and plain English. People may still accept participation even if the message is not understood, so trust is key
Time to visit	- Choose a time to suit participants
	- For our cohort, between 10am-3pm was the best
PRE-VISIT	
Confirm the appointment	- It's not enough to have your time set in stone.
	- Consider mistakes by participants and plan to re-confirm the appointment with them
Finding your way	- Cities like Dublin are geographically intricate
	- Use hand-held technology and maps to plan a route
Preconceptions and trust	- Participants can be naturally wary
	- Conversational structure and empathy help gain trust and make feedback more context rich
Identification	- Social responsibility: bring a form of identification as a courtesy and display before entering the home
VISIT INTRODUCTION	
Informing the participant	- Participants can be eager for answers at the outset
	- Put their mind at ease and instruct them as to your leading role in the interview; give them a tutorial at the end

Table 1 (*continued*)

Out of sequence data: Be prepared	- Participants can give multiple answers all at once and ask questions at the onset of the interview - Allow onset questions to avoid cognitive loading implications; Use flexible recording materials; set-up quickly to minimize disruption to flow
Refreshments	- Participants homes are not labs. It can be seen as a social visit. Accept refreshments and hospitality so as not to affect the interview
Recording instruments: Out of sight	- Having recording instruments in line of sight can cause some anxiety - in some participants more than others. Keep recording instruments out of direct line between the researcher and the participant
Body language & demeanor	- Initial atmosphere can be reserved, formal, arms folded, and non-relaxed - Important to facilitate maximum participation by allowing open discussion; be receptive, supportive, affirmative; take time to listen and laugh
Type of Visit	- Not all interactions are the same. Different types of visit - short, long, proxy, and long-response visit. -Prepare strategies to cater for different types of visit
DURING THE VISIT	
Discerning reality	- Direct questioning about having difficulty using technology seldom works. This needs exploration & observation of the participant using the technologies - User's can report being satisfied with a device while not understanding how to use it. Satisfaction can be biased from opinions of others. So investigate and exercise discernment - Researcher's presence can bias the behavior. Some user's use of technology was only as a result of the researcher's presence. Classic measurement error
Challenging assumptions	- Engineers and others have assumptions and preconceptions about older people which can affect what they perceive older people's needs are - For example, people with pendant alarms are not all frail; and contrary to what the author believed not all older people like photographs! (e.g. it reminds some of bad memories, or reminds them that they have gotten a lot older since then)
Unexpected visitors	- People/family may arrive at the home during interviews - Create strategies for controlling loss of focus due to unexpected visitors, or take a break
Adapting the approach	- Speak louder if needed. Adapt terminology to a more basic level as needed. Don't assume they know basic technology vocab such as rewind, etc. Prepare to use abstractions and analogies to explain terms/ideas
Terminology misinterpretations	- Don't confuse cognitive issues with terminology issues. Reword again and again as needed, and don't accept your first inclination from an answer. Be clear

Table 1 (*continued*)

A device cannot be too simple	- Even a single button device can seem deceptively easy to use and pose a usability risk
	- No technology should be considered too obvious or straightforward. Life experiences and exposure to technology can shape this
Repetition and speed	- Speak slower; Repeat yourself often but in different ways including scenarios. Give the user time to answer - allowing 10's of seconds of silence or utterances if need be
	- Reassure them to take their time when answering
What is an older person	- Over 50? 60? 65? It's an unfixed, large age range
	- Don't label the participant as old; adjust project and occupation titles with alias's where appropriate
Dealing with concerns	- Adverse occurrences should be planned for during interviews. A community worker can be instrumental in dealing with concerns
Older people do read manuals	- Expect a range of technology skills in older people. Some do read manuals
	- Conversely, don't expect all to be able to grasp written material - owing to perhaps, eye sight loss and lack of fundamental literacy skills
DIY older people	- Don't assume all older people cannot do DIY. Older people are a diverse group. Check each one
Attitudes to technology	- Technology that supports independent living can be positively personified; others find it takes away independence. So it varies with the user
	- Technology can be used for unintended purposes - e.g. a stairlift as a laundry elevator! If it works as they want it to then it is fulfilling its purpose to them
POST-VISIT	
Fix Something!	- Lasting relationships are important when working with communities
	- At the end of interviews, fix some simple technology (that takes <20mins) - tune in TV stations, show how to use a device, etc. It can add additional insight into the needs and wants of the older person
	- This creates additional appreciation and trust and lasting relationships; word can also spread in communities of older adults - and quickly too
A simple thank you	- Gratitude is important, especially to the socially mature; this creates positive sentiment and lasting relationships with the community
	- Give a simple thank you card; and if possible a small monetary voucher/gratuity is appreciated
Making it last: Creating an experience	- Do something different - create an event and invite the participants to explore the results of the study
	- Dramatic readings and guest speakers can be used to convey the message and get interactive feedback with the study participants/users

5 Discussion

A number of researchers reported on the challenges in engaging older people in research studies on technology and these are discussed in light of the current study.

Diversity and Differentiation: In terms of recruitment, it has been reported that one issue engaging with older people is the sheer diversity of the 65+ group making representative samples difficult to generate [Gregor et al. 2001]. Eisma reported [Eisma et al. 2004] that the diversity is larger in older rather than younger groups, and such ideas tend to challenge pre-conceptions about "older people" as seen in this chapter's findings. With such a range there also follows a range of abilities. Interestingly, Eisma accommodated for the diversity in the group by using different MS Powerpoint® presentations with different groups in focus group interactions. A shorter presentation was used with Stroke group; and for technology in the home for relatively inactive older adults, non-relevant (and perhaps inappropriate) technology was not included in the presentations. Therefore tailoring the approach to different "groups" of older people based on their abilities would be wise. Such approaches were not necessary for the current study, since the survey tool used was flexible so as to allow for such a range of users; and the technologies addressed were devices the user's were already familiar with in their own homes. Such tailored approaches however, would potentially be needed if introducing new technologies or concepts for new technologies to this user group.

Approach - Including the Users: There are many ways of engaging a user in user-centered design. Dickinson suggested that in-home interviews were very effective eliciting more stories and a wider range of data from the users [Dickinson et al. 2007]. Furthermore, user-centered design techniques such as "think aloud" were found not to work well with this group and can interfere with the process of completion of the experimental task - as discussed in [Dickinson et al. 2007]. Therefore, it has been suggested that one-to-one sessions may be the best method although it poses limitations as to the scale of the study with limited resources. Going into older people's home has, however, an added benefit as found by the authors of this study: the frailer, underrepresented older person who ordinarily is not fit to come to a lab or focus group for design discussions can be included when doing in-home assessments since the effort required is minimised. Moreover, the user would perhaps tend to be more comfortable and natural in their own home and this complex interaction could have added benefit.

Assumptions: To avoid assumptions about what older people know, the experimenters should make sure to be very explicit in every expectation they have of the interviewee in the interview as they may most likely be unfamiliar with the experimental techniques used [Dickinson et al. 2007]. There can be a tendency during interviews to assume the rules of engagement for an interview to be known - but these should be formally specified, e.g. “we want to see you use the device, rather than us talking you through it, but we can show you at the end of the interview if you like”. Socially skilled older people can tend to attempt to lead by trying to include the facilitator in the experiment [Dickinson et al. 2007]. Therefore, in this current study, the authors found it best to make sure to specify their expectations of the user’s participation and feedback.

User Perceptions: It was a key theme of the current study that older users tended to belittle themselves and their ability to use new technology. This confusion is shared by other researchers [Eisma et al. 2004] where it was found that a particular older person said they could use television-based email (using a keyboard and remote control) but thought that they would never be able to use a PC. This agrees well with suggestions that perhaps older users would be more accepting of technology that is similar to what they are already used to such as the TV [Coleman et al. 2010]. Nevertheless, some users in the current study went as far as to call themselves “thick” and other derogatory terms. The author’s have an interest in investigating and addressing ‘stereotype threat’ in older people in Dublin to see if they can be shown that these perceptions are false. Dickinson also reported on the significance of ‘time’ in learning to use technology, which should be borne in mind [Dickinson et al. 2007]. When learning, users may tend to fair well but many after a few weeks tend to have a “Crisis Week” where they seem not to be able to do what they could previously. In all cases however, they reported that they had a breakthrough week between 3-6 weeks from the onset of the learning sessions. Such reports are important so as not to generalize that older users cannot learn to use new technology and it is ok if the learning curve is not a smooth upward one. Other possible myths and half truths are discussed by [Wandke et al. 2012] challenging assumptions about older people and technology that are mooted, e.g. that if we wait for 20 years or so, then the new generation of older users will be ‘tech savvy’ and so the problem will solve itself; whereas in fact, if technology continues to pace as it has done in the past 30 years, the technology landscape will look much different than today and so the same technology use problems will exist - just in a new form.

6 Conclusion

Older users of technology should be included in technology design for older people so as to meet their requirements, needs, and wants. One method of inclusion has been explored in this chapter - in-home interviews and technology assessments. In future, the authors wish to conduct further community research and explore methods of counteracting their negative opinion of their capabilities of using technology.

Acknowledgment. The authors would like to acknowledge the financial contribution of the Centre of Excellence of Universal Design/National Development Authority (CEUD/NDA) and for their commissioning of this research project. The authors would also like to thank the extensive aid and support of the South Inner City Community Development Agency (SICCDA) and especially Kay King.

References

[Beecher et al. 2005] Beecher, V., Paquet, V.: Survey instrument for the universal design of consumer products. Applied Ergonomics 36(3), 363–372 (2005)

[Coleman et al. 2010] Coleman, G.W., Gibson, L., Hanson, V.L., Bobrowicz, A., McKay, A.: Engaging the disengaged: how do we design technology for digitally excluded older adults? In: Proc 8th ACM Conf. on Designing Interactive Systems, pp. 175–178 (2010)

[Dickinson et al. 2007] Dickinson, A., Arnott, J., Prior, S.: Methods for human–computer interaction research with older people. Behaviour & Information Technology 26(4), 343–352 (2007)

[Eisma et al. 2004] Eisma, R., Dickinson, A., Goodman, J., Syme, A., Tiwari, L., Newell, A.F.: Early user involvement in the development of information technology-related products for older people. Universal Access in the Information Society 3(2), 131–140 (2004)

[Gregor et al. 2001] Gregor, P., Newell, A.F.: Designing for dynamic diversity—making accessible interfaces for older people. In: Proc of Workshop on Universal Accessibility of Ubiquitous Computing: Providing for the Elderly, Alcácer do Sal, Portugal, pp. 22–25 (2001)

[Lenker et al. 2011] Lenker, J.A., Nasarwanji, M., Paquet, V., Feathers, D.: A tool for rapid assessment of product us-ability and universal design: development and preliminary psychometric testing. Journal of Prevention, Assessment and Rehabilitation 39(2), 141–150 (2011)

[Soraghan et al. 2013] Soraghan, C., Hermann, S., Boyle, G.: Design and older people's real issues: Experiences of an engineer assessing technology in the user's home. In: Proc. 6th Inter. Conf. on Human System Interaction, pp. 600–607 (2013)

[The Centre for Universal Design 2002] The Centre for Universal Design (NCSU-USA) Universal Design: Product Evaluation Count-down

[The Centre for Universal Design 2000] The Centre for Universal Design (NCSU-USA), A Guide for Evaluating the Universal Design Performance of Products (2000)

[Wandke et al. 2012] Wandke, H., Sengpiel, M., Sönksen, M.: Myths About older people's use of information and communication technology. Gerontology 58(6), 564–570 (2012)

[WWW_1 2011] Census, Small Area Population Statistics (SAPS) (2011), http://www.cso.ie/en/census/census2011smallareapopulationstatisticssaps/ (assessed November 14, 2013)

The Assistive Walker for Visually Impaired Using Hand Haptic

S. Yokota[1], H. Hashimoto[2], D. Chugo[3], and K. Kawabata[4]

[1] Faculty of Science and Eng., Setsunan Univ., Osaka, Japan
yokota@mec.setsunan.ac.jp
[2] Advanced Inst. of Industrial Tech., Tokyo, Japan
hashimoto@aiit.ac.jp
[3] School of Science and Eng., Kwansei Gakuin Univ., Hyogo Japan
chugo@kwansei.ac.jp
[4] RIKEN-XJTU Joint Research Unit, RIKEN, Saitama, Japan
kuniakik@riken.jp

Abstract. The purpose of this research is to develop the sight assistive walker using hand haptic. The feature of the system is user centered system. The main important tasks while walking such as environmental recognition, path planning and locomotion are made by the user. This system is not involved in such important tasks. It provides the users only the environmental information through hand haptic in order to create user's mental map. The system configurations are as follows. This system is something like a cane, and it has a small passive mobile platform on its tip. The LRF (Laser Range Finder) is used for environmental sensing. The powder brakes are used for generating passive brake forces. In order to let the mobile platform move both lateral and back-forward, the four omni-directional wheels are adopted. By generating brake force based on the distance to the obstacles in the moving direction, user can recognize the environment through the hand haptic information. The experimental result showed that the proposed system can assist the walking and creating mental map by presenting environmental information through the hand haptic.

1 Introduction

Many supporting systems for visually impaired have been developed up to the present. For example, a guide dog system [Iwatsuka et al. 2004] adopts an autonomous mobile robot as its mobile platform, and navigates users. An indoor navigation system was proposed for the visually impaired [Kulyukin et al. 2004]. This system also navigates users by using small mobile robot and embedded small sensors into environment. The cylindrical guide robot was designed which has three wheels driven by motors [Tsuda et al. 2009]. In many researches, the system is able to navigate user. The common features among these researches are that system can recognizes environmental information and conducts path planning, and navigate users by using active actuators. That means the user cannot take positively part in environmental recognition and path planning. To describe in an extreme

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_20, © Springer International Publishing Switzerland 2014

manner, the user can reach the destination without user's intention. If user cannot recognize environment by him/herself and user's intention is not involved, there will be a possibility to increase user's sense of insecurity. Therefore, the problem of such autonomous navigation type assistive systems is that the user's intention is not involved actively.

For this problem, The assistive walker was proposed which consists of a cart with no motors, a force feedback joystick and a Laser Range Finder (LRF) [Yokota et al. 2007]. This system transmits environmental information to user through the force feedback joystick with hand haptic. User can walk by its own and recognize environmental information while pushing cart and rotating joystick. The advantage of this system is that the user twists actively own wrist for sensing environment, namely this is a user oriented system. In other words, this system entrusts the user all decisions including environmental recognition, path planning and walking. However some improvements remain. The cart of this system has non-holonomic constraint, namely moving direction is restricted. The environmental recognition and walking are not done simultaneously, because user has to stop in order to scan environment by LRF and force feedback joystick.

Therefore, this research develops the assistive system which provides environmental information through hand haptic. The proposed system can move omni directional and enables the user to sense environment and to walk simultaneously.

2 Design Concept

The important point of above mentioned the assistive walker is: The role of the system is to provide the environmental information, the decisions and actions such as path planning and environmental recognition should be entrusted to user.

In other hand, the mental map is important factor when the user walks while making path planning with own intention. The mental map is the environmental information which user has in own brain, and is called cognitive map.

As we mentioned above, this research develops the assistive system for visually impaired. From the above, the purpose of this research is to develop the system which assists for creating mental map by providing environmental information through the hand haptic. Providing the environmental information contributes to creating mental map for user. The created mental map helps user to make the path planning and walking. The features of proposed system are to provide the wide range of environmental information through the hand haptic and entrust the user the decision and action for walking. In order to realize this feature, the system doesn't use any active actuators. The system uses passive actuators for enabling the user to get force feedback and to walk simultaneously. Here, this environmental information should be wide surrounding area of users. This system is small apparatus which has environmental sensing device and passive actuators. The user can sense the force feedback based on environmental information through the grip of the system while user walks. The concept is shown in Fig.1.

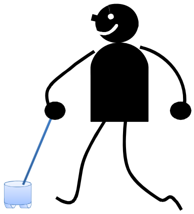


Fig. 1 The Concept of Assistive System

In order to provide force feedback to user, the system should have a certain device which generates forces. In addition, our system concept is not to use any active actuators, therefore this system is based on the concept ”Passive Robotics[Hirata et al. 2006]”. Thus, proposed system adopts some wheels and brakes.

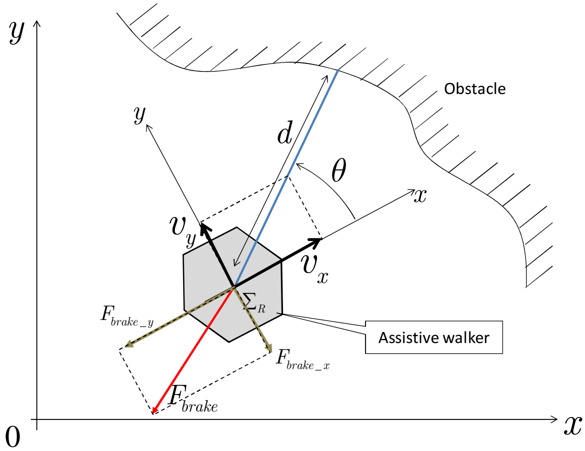


Fig. 2 The Coordinate of the Assistive walker

The system generates braking forces based on the distance to the obstacle in the moving direction of the system while user is walking and sliding the system. By this manner, the user can recognize environment information and create mental map. The scheme for generating braking force is: As Fig. 2 shows, the velocities of x axis and y axis of the system are respectively v_x [m/s] and v_y [m/s] on the local coordinate of the syste . The moving direction of the system θ [rad] is derived by these two velocities as following.

$$\theta = \tan^{-1}\left(v_y / v_x\right) \tag{1}$$

The distance to the obstacle in the direction θ is d [m]. The braking force F_{break} [N] is generated in inverse proportion to the distance d as following.

$$F = \alpha \frac{1}{d} \quad (2)$$

Here, α is conversion coefficient from distance to force. This force is transmitted to the user through the grip, then user can recognize environmental information by hand haptic.

3 System Configuration

The braking force F_{break} is generated in the opposite of moving direction of the system θ. The moving direction θ is varied by user's action (e.g. sliding the system), thus θ is not constant. Therefore the braking force should be generated in the omni-directional. For this requirement, the omni-wheels and powder brakes (Fig. 3) are adopted.

The powder brakes easily generate braking torque in the proportion to the current inputting the brakes. The powder brakes generate F_{break_x} , F_{break_y}, then the resultant force of these becomes F_{break}. In order to generate F_{break_x} F_{break_y} the powder brakes are attached to the omni-wheels and control the viscous resistance of each wheel.

Fig. 3 Omin-wheel and Powder Brake

3.1 Deployment of Omni-Wheels

Here, let us consider the deployment of the omni-wheels. This system is moved in right/ left and forward/backward by user through the grip attached to the system. This grip is attached to the origin of the local coordinate of the system, thus the system orientation is not varied. As we described before, this system generate F_{break_x} and F_{break_y}. Therefore, the controlled force direction is 2 translation D.o.F. (x and y direction), the orientation is not controlled.

Meanwhile, conventional omnidirectional mobile mechanism uses three or four omni-wheels with actuators, and controls translation 2 D.o.F. and orientation 1 D.o.F.. In case of adopting conventional three wheels type for our system (Fig. 4(a)), the three brakes are needed to the every axis for controlling translation 2 D.o.F.. On the other hand, in case of adopting conventional four wheels type (Fig. 4(b)), we need to attach brakes on all four wheels, or 3 brakes by using mechanical decoupling [Kawabata et al. 2002]. From above, our system adopts the wheel deployment as shown in Fig. 5. By this deployment, we can control 2 translation forces by two powder brakes without redundancy.

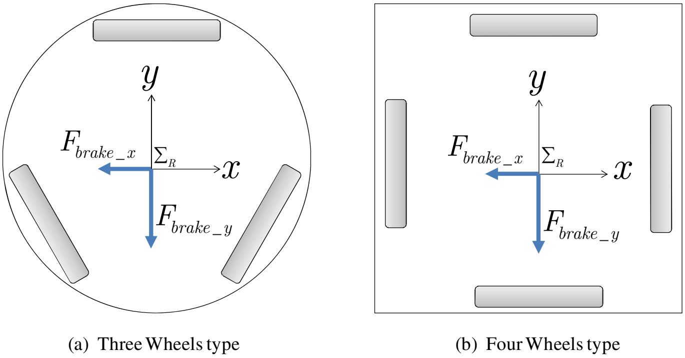


Fig. 4 Conventional deployment of omni-wheels

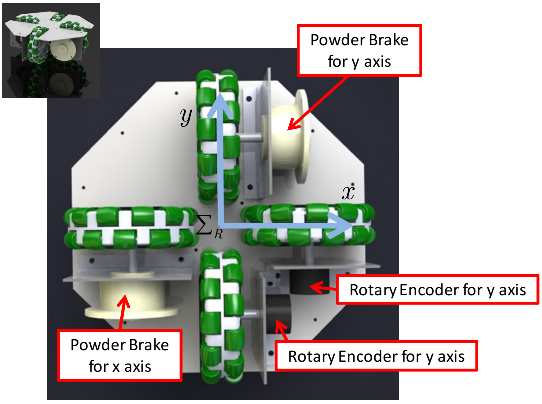


Fig. 5 Deployment of Omni-wheels

3.2 Components and Configuration

As other components besides powder brakes, the system adopted two rotary encoders for measuring ,and LRF(Laser Range Finder, URG-04LX, Hokuyo Automatic Co., Ltd) for measuring the distance to the obstacles. The system configuration of these components is shown in Fig. 6. The microcontroller counts the pulses from rotary encoders, and send it to the embedded Linux which is host controller. The host controller calculates the moving direction using (1) after receiving v_x , v_y then obtain the distance data d in the direction of θ from LRF. The braking force is calculated by (2), and is divided into x , y direction. Divided two forces are sent to microcontroller, then the microcontroller controls the powder brakes. Armadillo-420 (core is ARM 9) is used for embedded Linux, and PIC18F2550 is used for micro controller.

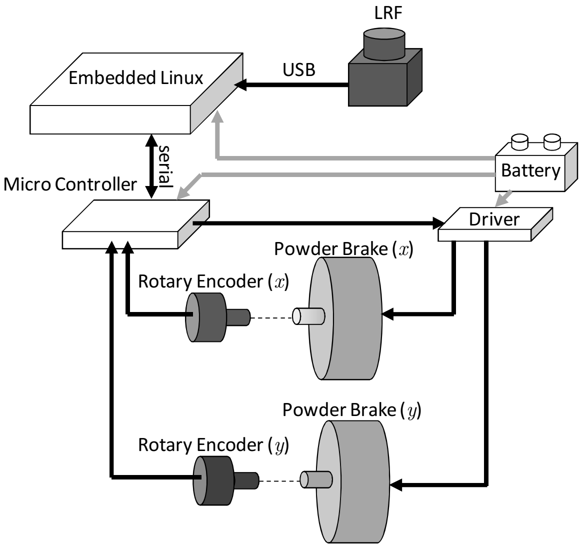


Fig. 6 System Configuration

3.3 Powder Brake Driver

Powder brake is ZKG-10YN(Mitsubishi electric) in our system. Generally, the special amplifier is needed for controlling the powder brake. The special amplifier is too big for our system. Therefore we design the driver for powder brake in this subsection.

The magnetism powder is enclosed between output shaft and case of powder brake. The coil is mounted on the case. When inputting the current into the coil, the magnetic field is generated. And magnetism powder become the chain form, the friction is generated between case and output shaft. By this manner, we can generate the brake torque which is in proportion to the inputting current. From this driving principle, since this driving method is same as conventional brush DC motor in principle, we consider the PWM driving using FET. Driving circuit is conventional FET source grounded circuit. We measure the relation between PWM duty ratio and magnetic excitation current. The result is shown in Fig. 7. It turned out that the relation between duty ration and current is not linear relation.

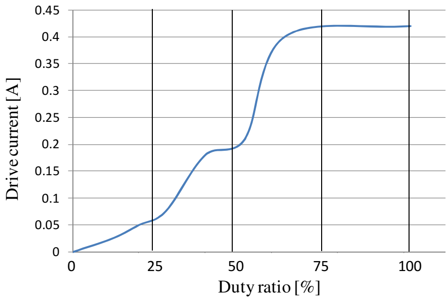


Fig. 7 Drive current VS PWM duty ratio

The other driving method is considered because the nonlinearity is not preferable. The special amplifier controls the brake torque by using continuous current control. From this fact, we consider analog control not digital control. In order to realize analog control, we use the power operational amplifier. Microcontroller does not have DA functions. Thus we make analog signal by CR filter and PWM signal (Fig .8). In this circuit, we measure the relation between PWM duty ratio and current. The result is shown in Fig. 9.

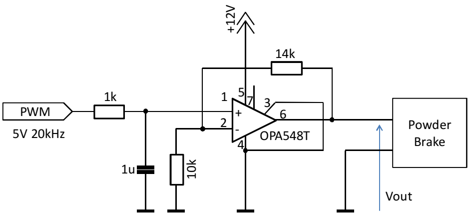


Fig. 8 Schematic of Driver

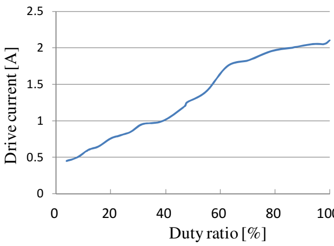
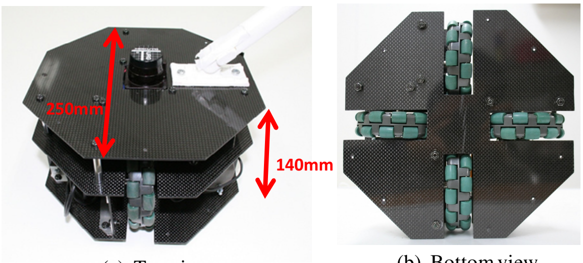


Fig. 9 Vout VS PWM duty ratio

From this figure, the relation between duty ratio and current has a linearity compared to the PWM +FET drive circuit. Therefore, this system adopts the circuit depicted in Fig. 8 as a driver of powder brake.

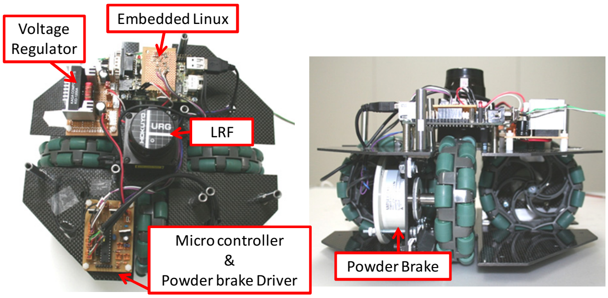
Based on the above discussion, we made a prototype. Since the body should be lightweight as much as possible, we use the CFRP as the body material of the system. The power source is lithium polymer battery (3 cells 700mAh). The prototype is shown in Fig .10, 11 and 12.

Fig. 10 Prototype



(a) Top view (b) Bottom view

Fig. 11 Dimension of the Prototype



(a) Top view (b) Side view

Fig. 12 Inside of the Prototype

4 Experiment

In order to verify the function and motion of the system, the basic experiment was conducted. The crank shaped corridor (Fig. 13) was prepared. The subject eyes were covered, and the subject did not know the shape of corridor in advance.

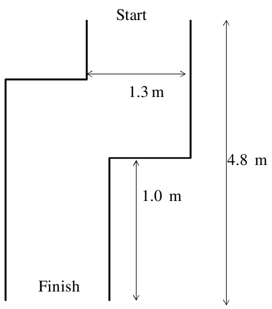


Fig. 13 Experimental environment

As a result, the subject could walk through this corridor. Fig. 14 shows the experimental scene. The interview to the subject shows that the user can briefly know the corridor shape by using the force transmitted through the grip. Therefore it turned out that proposed system can assist the walking by presenting environmental information through the hand haptic.

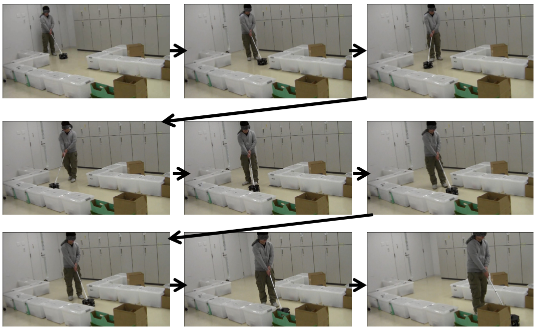


Fig. 14 Experimental scene

In addition, we did another experiment to verify the creation of mental map. We arranged two plastic boxes as Fig. 15 shows, and the subjects with the walker was in the center of two boxes. The subjects didn't move and scanned the environment using the walker at the center of two boxes. The subject received the environmental information using hand haptic.

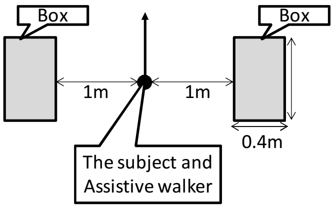


Fig. 15 Arrangement of boxes

The subjects did not know the environment in advance of the experiment. After the experiment subject were moved to another room with eye mask. In the another room, the subjects were asked to sketch the image of the environment. This image is the mental map. The examples of user's mental map are shown in Fig. 16.

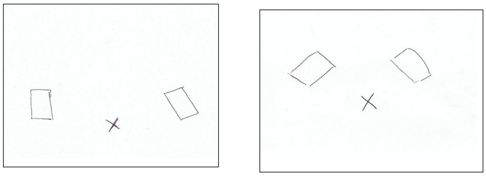


Fig. 16 Imagined Boxes Deployment by Subjects

In this figure, "x" indicates the position of the subjects with the walker. In this figure it turned out that the mental map was correctly created by using hand haptic.

5 Conclusions

This research develops the assistive system which provides environmental information through the hand haptic for visually impaired. The proposed system can move omni-directional and enables the user to sense environment and to walk simultaneously. The role of the system was to provide the environmental information, the decisions and actions such as path planning and environmental recognition should be entrusted to user.

This system is small apparatus which has LRF, powder brakes as actuators and omni-wheels. The user can sense the force feedback based on environmental information through the grip of the system while user walks and moving the system. This grip is attached to the origin of the local coordinate of the system, thus the system orientation is not varied. Therefore, the controlled force direction is 2 translation D.o.F. (x and y direction), the orientation is not controlled. From above, our system adopt the wheel deployment as shown in Fig .5. By this deployment, we can control 2 translation forces by two powder brakes without redundancy.

The system configuration is: The microcontroller counts the pulses from rotary encoders, and send it to the embedded Linux which is host controller. The host controller calculates the moving direction using (1) after receiving v_x , v_y, then obtain the distance data d in the direction of θ from LRF. The braking force is calculated by (2), and is divided in to x, y direction. Divided two forces are sent to the microcontroller, then the microcontroller controls the powder brakes.

The basic experiment was conducted using crank shaped corridor. As a result, the subject could walk through this corridor, and create the mental map. It turned out that proposed system can assist the walking and creating mental map by presenting environmental information through the hand haptic.

References

[Hirata et al. 2006] Hirata, Y., Hara, A., Kosuge, K.: Development of intelligent walker based on passive robotics concept. Journal of Robotics Society of Japan 24(2), 270–276 (2006) (in Japanese)

[Iwatsuka et al. 2004] Iwatsuka, K., Yamamoto, K., Kato, K.: Development of a guide dog system for the blind people with character recognition ability. In: Proc 17th Inter. Conf. on Pattern Recognition, pp. 453–456 (2004)

[Kawabata et al. 2002] Kawabata, K., Suzuki, T., Kaetsu, H., Asama, H.: Ominidirectional mobile platform for reserach and development. Journal of Robotics and Mechatronics 14(2), 105–111 (2002)

[Kulyukin et al. 2004] Kulyukin, V., Gharpure, C., De Graw, N.: Human-robot interaction in a robotic guide for the visually impaired. Technical Report SS-04-03, pp. 1–X. The AAAI Press, Menlo Park (2004)

[Tsuda et al. 2009] Tsuda, D., Wang, S., Miura, N.: Developmet and safty validation of lead for a guidance robot. Biomedical Fuzzy Systems Association 11(1), 43–48 (2009) (in Japanese)

[Yokota et al. 2007] Yokota, S., Hashimoto, H., Sasaki, A., Takeda, D., Ohyama, Y.: Supporting technologies for weak-eyesight persons - automatic color conversion and training vision. In: Proc. SICE Annual Conference, pp. 3064–3068 (2007)

Concept and Design of SEES (Smart Environment Explorer Stick) for Visually Impaired Person Mobility Assistance

M. Yusro[1,3], K.-M. Hou[1], E. Pissaloux[2], K. Ramli[3], D. Sudiana[3], L.-Z. Zhang[1], and H.-L. Shi[1]

[1] Laboratoire d'Informatique de Modélisation et d'Optimisation des Systèmes, Université Blaise Pascal, France
{Yusro,kun-mean.hou,hongling.shi}@isima.fr

[2] Institut des Systèmes Intelligents et de Robotique, Université Pierre et Marie Curie (UPMC, PARIS 6, Paris-Sorbonne), France
Edwige.Pissaloux@upmc.fr

[3] Department of Electrical Engineering, Faculty of Engineering, Universitas Indonesia, Indonesia
{k.ramli,dodi}@ee.ui.ac.id

Abstract. With the advent of robot companion concept, the independent mobility became one of the hottest problem in general, and allows make progress on the independent mobility of the Visually Impaired Person (VIP). Today, this latter is based mainly on two elements: the white stick and VIPs memory. However, the feedback provided by a white cane, even an enhanced one (a smart stick), is not adequate to the principle of autonomy and to the growing complexity of our cities. Therefore, new capabilities such as orientation and space awareness should be acquired by VIP. The first prototype of SEES (SEE-stick and SEE-phone) is presented in this paper. This paper: (1) introduces concepts of orientation and space awareness, (2) outlines the architecture of the SEES (Smart Environment Explorer Stick), an enhanced white cane, which assists the VIP's to acquire orientation and space awareness capabilities, and (3) discusses the SEES possible implementation (SEES hardware and software first prototype).

1 Introduction

Data from [WHO 2011] indicates that there are 285 millions visually impaired people (VIP) currently in the world, 39 millions of them are blind and 246 millions have the low vision. Approximately 90% of people with visual impairment live in developing countries. About 65% of all people who are visually impaired are aged 50 or more, while this age group comprises about 20% of the world's population. Information from European Blind Union [EBU] estimates that more than 30 millions of blind and partially sighted persons live in Europe, and, in the average, 1 of 30 European experiences sight loss, i.e. there are four times more partially sighted than totally blind persons. In Indonesia, 1.5% of Indonesia's

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_21, © Springer International Publishing Switzerland 2014

population (approximately 3.6 millions) is estimated to be blind what indicates that the blindness is one of the most important epidemiological problems [MHRI 2013]. According to the survey in France [Lafuma et al 2006] prevalence of visual impairment increases exponentially with age as shown in Table 1 (data given in thousands of people).

Table 1 Prevalence of visual impairment at older ages in France

Age/ Type of Visual Impairment	60-69	70-79	80-89	90-99	100+
Low vision	3.06	5.92	14.10	23.13	33.71
Total blindness	0.21	0.09	0.91	4.73	3.27
Visual impairment	3.27	6.01	15.01	27.86	36.88

The above statistics show that it is necessary to find the most appropriate solutions for VIP integration in the modern society and to assist their independent life and interaction. Blindness and visual impairments are a major hindrance in daily life such as access to information or interaction with the environment.

The *mobility* is a one possible interaction with the external world. The human *autonomous mobility* is a capability to reach a place B from the current place A without assistance of another person. However, the autonomous mobility raises several problems [Pissaloux 2009; 2011]; two of them, namely *orientation* and *space awareness*, are keys elements for a true independent mobility.

The concept of (self) *orientation* means the knowledge of the relative spatial position between our current position A and targeted spatial location B. Two basic questions which subtend the orientation are: *"where I am now?"* and *"how to reach the destination from my current position?"*. Consequently, the orientation allows planning globally a specific route (path) to reach the targeted location from the current point.

Space awareness is the capability to know about urban and social data in our peri-personal space ("space around our body") and our navigational space [Tversky 2005]. Some questions which subtend the space awareness might be: *"what are the nearest streets located one with respect to the other and with respect to me? what is the traffic light status at the street I like to cross? where is the library I like to get in? "*. Consequently, the space awareness permits to understand how our nearest space is organized and how we can interact with.

Existing mobility systems to assist the VIP's navigation in the known/unknown and indoor/outdoor environments do not usually support the orientation and space awareness concepts.

This paper proposes to fill this gap by defining a new assistive device for VIP and named SEES 'Environment Explorer Smart Stick'; this device aims to assist some basic sub-functions of the orientation and space awareness.

Therefore, the paper is organized as follows: Section 2 provides a state-of-the-art on the available devices for VIP mobility and focus on technological solution to support *space awareness* concept; Section 3 outlines the SEES architecture; Section 4 propose very preliminary technical evaluation of the SEES prototype some functions. Section 5 indicates some future works toward a full operational SEES prototype.

2 White Cane Evolution: A State of the Art

The proposed state-of-art is composed of two synergetic parts: a review of some existing cane bases solutions for VIP mobility (2.1) and an introduction of the concept of context-aware human-computer interface (HCI) for VIP mobility (2.2).

2.1 Some Existing Cane Bases Solutions for VIP Mobility

Several technological solutions have been proposed in order to assist the VIP mobility. This paper investigates only the evolution of assistive devices based on a cane: white cane, smart cane, robotised smart cane and intelligent cane.

The white cane (or white stick) is the first device invented by a British photographer, James Briggs, in 1921 that became blind after an accident, and it is still largely used to assist VIP mobility. The white cane assists the VIP's walking by providing the tactile feedback to their hand on the status of their near navigation space; its permits to determine on the ground in a distance around 1m ahead of the VIP, obstacle-free zones and existing obstacles. Once an obstacle detected, the VIP should elaborate an appropriated avoidance procedure by the cognitive integration (synthesis) of the discrete tactile feedbacks. Figure 1 [Willough, Monthei 1998] allows to imagine how complex is the global information integration (the whole obstacle shape recognition) using a white stick.

Fig. 1 Complex procedure of obstacle detection using a white stick

The cane/stick has several drawbacks [Palleja et al. 2010]. Indeed, it does not support the orientation and space awareness: these both functions strongly depend on the VIP memory of his/her surroundings established during the previous mobility experiment (learning) and his/her kinesthesia sense.

The white stick turns into a *smart cane* by adding several sensors to it. These sensors are usually dedicated to detect some obstacles not sensed by the white cane but being located in the reach of the cane (such as overhanging obstacles or upward stairs) or to extend the reach of the cane. These additional obstacles should be localized in the solid angle subtended by (ultrasound or laser) sensors attached to the white cane. A (tactile/hand or audio) feedback to VIP exploits the reflected by the obstacle intensity of the sensor generated energy and captured by the cane. The ”whole” obstacle shape should be ”reconstructed” (cognitively integrated) by the end-user from cane scanned points. The distance to the obstacle is estimated thanks to TOF (Time-of-Fly) between sensor and the reached nearest obstacle. Smart canes provide neither orientation nor space awareness assistance. Several smart canes have been realized; they differ by the feedback sent to the VIP.

The *Teletact* (figure 2) is a handheld laser telemeter device which detects an (overhanging) obstacle with 1% of error at the distance from 10cm to 10m at the rate of 40 measurements per second [Farcy et a 2006]. The distance is provided through vibrating feedback on the user’s palm hand.

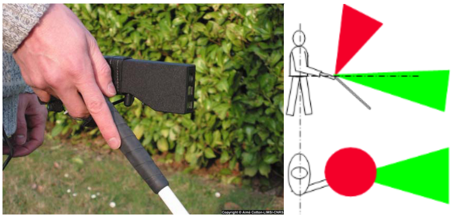


Fig. 2 Teletact Cane, University of Paris 11

The *UltraCane* [Hoyle 2006], a *smart stick* designed at the University of Leeds (figure 3), uses also the ultrasonic sensor to detect obstacles and transforms this information into vibrations which stimulating a thumb keep on the handle button.

Fig. 3 UltraCane, University of Leads

The smart stick *K-Sonar* (figure 4), built at Canterbury University, New Zealand, uses also ultrasonic sensor for obstacle point-wised detection but it provides an audio feedback (space point-wise sonification) [Terlau et al. 2006].

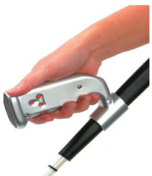


Fig. 4 K-Sonar, Canterbury University

The *GuideCane* (figure 5), a smart (robotised) stick, is designed to assist the VIP to detect ground located and specific obstacles such as upward stairs. The set of ultrasonic sensors located on a distal end of the smart stick transform acquired data into a vibrating feedback stimulating the end-user palm [Borenstein, Ulrich 2001].

Fig. 5 GuideCane, Michigan University

The *SmartCane* of the Indian Institute of Technology (figure 6) [Rohan et al 2007] detects knee-above obstacles using ultrasons.

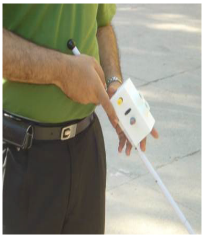


Fig. 6 SmartCane, Indian Institute of Technology

The *Intelligent Cane* "iCane" built by the National Taiwan University (figure 7) partly assists the VIP walking (obstacle detection) and basic space awareness. iCane orientation assistance is supported by the RFID tags: data are exchanged between the white cane RFID, and RFIDs embedded in the environment [Chang, Tsung-Hsiang et al 2005]. The data are transmitted to the end-user via PDA headphones connected with Bluetooth to the iCane. The provided information useful for walking is related to the space urbane organisation (e.g. street intersection, elevator, stairs, nearby shops/market, etc.). The main drawback of this system is that the iCane requires RFID instrumented environment (high implementation and maintenance costs, not sustainable solution).

Fig. 7 iCane, National Taiwan University

It should be observed that all current smart sticks do not assist neither orientation nor space awareness. In order to provide the VIPs with such assistances the smart cane/stick should be transformed into a space explorer smart stick (SESS). The SEES system, presented in the section 3, proposes a solution to overcome these limits.

2.2 Context-Aware Human-Computer Interface for VIP Mobility

A task executed by an agent (human, biological, virtual, etc.) is performed in a certain context. A *context* is a set of all necessary and useful for the considered task execution information (and only this information). For example, as far as the mobility is considered the current status of the environment (user's near and navigation spaces) should be taken into account. Indeed, if we want to cross a road, there is necessary to check the status of the traffic light which says what is effectively possible to do; we speak about *space awareness.* For VIPs, the environmental noise (such as the car's noise) constitutes the useful space awareness cue on the status of their near space. If the computer is used as a media for interactions with the external world, the contextual data should be provided by the *Human-Computer-Interaction* (HCI) systems.

Currently, one of the most fundamental tasks when building intelligent HCI systems is multimodal sensing and collecting space awareness data (environmental information) pertinent for a given task [Stillman, Essa 2001]. The sensors involved in space awareness data collection can be passive or active. The passive sensors do not disturb the observed environment, contrary to the active sensors. A visible spectrum vision sensor (such as a vision camera) is a typical example of passive sensors; an ultrasonic sensor is an example of active sensors.

Some of existing mobility assistance provides very limited space awareness to VIP. "Navigational beacons" for example [Loomis et al 2005], provide a repeating, directionally selective voice message during the indoor navigation; however, they are limited to this collaborative environment which became intelligent by embedding specific information system and passive sensors into it.

Providing maximal awareness on the environment, without requiring any modification to the existing infrastructures, will significantly enhance the experience of mobility to the VIP in any kind of environments.

However, all (active and passive) sensors should be integrated in a wearable device what empowers VIP with new sensory capabilities.

Moreover, the HCI system behavior should be organized as a set of states; a transition from one state to the other will be driven by a data provided by the sensors (after their fusion). A feedback to user (a guidance message) will accompany each state transition.

The SEES system presented in the section 3 proposes an integration of the space awareness concept into a wearable unit using several active and passive sensors.

3 The SEES System

This section provides an overview of the targeted space smart explorer stick (SSES) system named SEES (3.1), details of sensors to be integrated in its first prototype (3.2) and SEES system operational modes useful for VIP independent mobility (3.3).

3.1 SEES Overall Architecture

The SEES system contains three main components (cf. figure 8):

- A global remote server (iSEE),
- An embedded local server (SEE-phone), and
- A smart stick (SEE-stick).

The iSEE is a global server providing the web services for the VIP such as remote real-time hint and help and remote monitoring (trace the VIP location). The SEE-phone is based on a commercial smart phone. It is used as an embedded local server and provides the local services for the SEE stick such as route vector and internet access. SEE-phone is the key device for orientation and space awareness; indeed, the SEE-phone communicates with the GPS, through the web server accessing to the map database, and with the others mobile devices. The SEE-phone is always connected to SEE-stick through Wi-Fi.

The SEE-stick is a white cane with a battery of active and passive sensors attached to it such as classic camera, GPS receiver, wheel encoder, ultra son, compass and accelerometer. These sensors will be used for environmental and VIP mobility cues collection and for high level mobility function implementation.

It should be stressed that 6LoWPAN (IPv6 over Low Power Personal Area Network and) and RPL routing protocol are adopted to implement the SEE-stick; therefore, according to the context the SEE-stick can connect P2P with the iSEE and future transportation system [Yibo et al 2011; Shi et al 2011].

Consequently, the SEES system can be considered as an implementation of ITS '*Intelligent Transportation System*' concept.

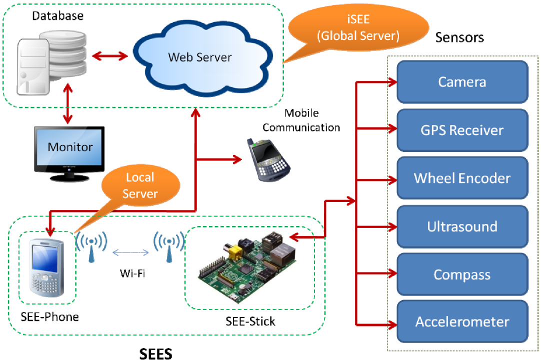


Fig. 8 The global concept of SEES system

The SEES possesses the following independent mobility characteristics:

- The VIP can know continuously his/her current location;
- The others person can monitor the VIP's journey and provide a remote assistance, if necessary;
- SEES could be used in the near future by VIP to access urban transportation system (such as car, bus and train) and to GIS (geographic and information system and local detailed maps).

The mobility cues collected by SEES sensors will be transformed into high level space awareness knowledge in order to:

- Estimate or predict the status on an object (for example: status of the traffic light or status of the walking surface);
- Obtain the accurate location data (what allows to track the user and check if (s)he is on his/her correct way);
- Send quickly error/alerts messages to VIP when orientation mistake occurs.

All feedbacks sent to VIP (as a result of multimodal data fusion) are passed via voice or touch stimulation channels.

3.2 SEES Support of Mobility Functions: Sensors, Considered Mobility Cues and Running Models

SEES targets to assist some walking, orientation and space awareness functions. The considered functions are the most required one by the end-users. For the first SEES prototype, the targeted *space awareness function* is the traffic light status detection; the targeted *walking* functions are the following: obstacle detection, walked distance estimation and surface roughness estimation. The targeted *orientation* functions are: direction to the targeted location estimation and end-user current location/position estimation.

The targeted functions have a direct impact on selected sensors. The status of the traffic lights – a *space awareness function* – is detected by using smart phone *camera* of the SEE-phone (a commercial smart phone).

The walking function is supported by a combination of *ultrasonic, camera, wheel encoder, accelerometer* and *compass.*

The *ultrasonic* sensor is used to detect the obstacles in front of the VIP (e.g. tree, wall, etc.) up to 2m ahead.

The *wheel encoder* sensor is mounted on the SEE-stick wheels; it is then used to estimate VIP travel distance from the starting place. The distance is estimated by combining the data from the wheel encoder, the route vector and the GPS. The travel distance is periodically sent to the local server (SEE-phone) and global server (iSEE) for higher level mobility functions.

The *camera* of SEE-stick is used to detect and recognize the geometry and urban organization of the space (such as road junctions, map of cross section or traffic lights localization).

In association with camera information, *accelerometer* sensor data is used to estimate walking surface roughness, so the VIP can adapt their posture. The *compass* sensor is used to detect a moving direction of the VIP. Data from the compass will be integrated with the GPS data and wheel encoder data to enhance the precision of VIP location and distance estimation.

The *camera* of SEE-stick is used to identify and recognize the geometry of the space such as the type of road junctions or map of cross section. Moreover, this camera is also used to detect the traffic lights position.

The *orientation function* will use two main SEE-phone sensors: camera and GPS (includes a Google map). The GPS is used to determine the current location of the VIP, while the Google map allows to plan the VIP's journey in order to reach his/her spatial target. Periodically the GPS data will be sent to the database server for remote monitoring the VIP, displaying and checking his/her itinerary in real-time. Table 2 summarizes the different sensors and the mobility functions supported by the SEES system.

The mobility cues will be the inputs to the SEES layer which will transform them into high level knowledge useful for mobility; this knowledge will be directly conveyed to the VIP (a feedback).

Table 2 Type and function of sensors

Sensor	**SEES System**		**Mobility Cues**
	SEE-stick	**SEE-phone**	**(Output Data)**
Ultrasonic	V		Obstacle detection
Wheel Encoder	V		Travel distance
Accelerometer	V		Surface roughness
Compass	V	V	Direction
Camera	V	V	Obstacle/object status /space geometry
GPS	V	V	Location/position

3.3 SEES Running Models

SEES is a 3-module system (SEE-stick, SEE-phone and iSEE) with WIFI data transmission between the SEE-phone and SEE-stick. SEES has several working modes in function of the SEES hardware configuration. Table 3 shows all SEES modes. It is possible to switch from one model (assistance level) to another during the operation if the appropriate hardware is available.

Table 3 SEES operational modes

Mode	SEES System			Description
	SEE-stick	**SEE-phone**	**iSEE**	
Mode 0	V			Basic mode
Mode 1		V	V	Phone mode
Mode 2	V	V		Local mode connection
Mode 3	V	V	V	Complete mode

The SEES interface to VIP offers four assistance levels (operational models):

1. *Mode 0* is always active as it is the SEES minimal hardware configuration; only the SEE-stick is active - the SEES system becomes a SEE-stick.

   This mode will allow the VIP to walk without a smartphone. The walking will be assisted by stick sensors provided data (GPS, ultrasonic, camera, encoder, accelerometer and compass sensor).
   The main drawback of this mode is that the VIP can not receive any confirmation/information (or calls his/her friend/family) in relation to his/her journey progresses (correct or wrong). However, it is possible that SEE-stick is connected to the iSEE through Wi-Fi access point, when needed (opportunistic routing);

2. *In* the *Mode 1* -phone mode- the SEES works by using the SEE-phone and iSEE. This mode will allow a user to move without a smart. The mobility will be assisted using smartphone sensors (GPS, camera and compass) and iSEE server.
   The drawback of this mode is that the VIP can not detect neither obstacles nor the roughness of surface;
3. *Mode 2* -local mode- makes SEES working with both its subparts, SEE-phone and SEE-stick. The SEE-stick cooperating with the iSEE server and the SEE-phone is used only as a router (Wi-Fi access point) to overcome the mobility difficulties. The VIP can get a remote help from persons who monitor the server. In this mode, the SEES provides the selected travel specific hints and advices;

4. *Mode 3* – the complete mode – makes SEES working with all its sub parts (SEE-phone, SEE-stick and iSEE). The SEE-stick cooperates with the SEE-phone and iSEE to overcome the mobility difficulties. The VIP can get a remote help (from relatives using the telephone (mobile communication) or the others person who monitor the server. In this mode, the SEES provides the selected travel specific hints and assistance.

4 Experimental Preliminary Evaluations of SEES System Functions

As the SEES system is an association of two subsystems, SEE-Stick and SEE-Phone, preliminary experiments have been realized in order to estimate their technical performances: basic navigation with SEE-stick prototype (4.1) and basic space awareness function –traffic light status recognition – for SEE-phone/smart phone (4.2).

4.1 Experimental Technical Preliminary Evaluation of SEE-Stick Prototype

Basic walking assistances were implemented in SEE-stick prototype: it aims to track the VIP walking on a standard route, i.e. route, which consists of sidewalks and cross-streets (figure 9). The ideal navigation (reference) route started in one yellow point and continues through yellow path to another yellow point.

Three methods for track generation have been evaluated and compared the generated track accuracy:

- The RISS-GPS-LKF based on RISS (Reduced Inertial Sensor System)/GPS integration which we improved with the Linearized Kalman Filter (LKF) [Zhang et al 2013], (which generated the optimal track),
- The RISS/GPS integration method with GPS only (GPS track), and
- The RISS only (*Dead Reckoning*, DK track).

The tracking lines obtained by the three methods are shown in Figure 9. The yellow line is the reference navigation route; the red line (the optimal track) is given by RISS-GPS-LKF integration method; the green line corresponds to DK estimated navigation route and blue track is a GPS track.

The RISS-GPS-LKF integration method, comparing with GPS and DK methods, provides the best estimate of VIP's track.

Fig. 9 Experiment 1 of navigation accuracy validation

The first experiment with the VIP was done under a clear sky condition without GPS outage and without the VIP knowing the route in advance. The system provided an audio feedback to the VIP in order to keep him/her walking on the track estimated by the RISS-GPS-LKF algorithm.

During the walking process, the system checks if VIP's trajectories and headings agree with navigation route. The angle between VIP's heading and the sub-route's direction should be less than 45°. If the angle is great than 45° for a short period (e.g. 3s), the system tells the VIP the angle between his/her heading and the route direction and asks the VIP to adjust his or her heading.

4.2 Experimental Technical Preliminary Evaluation of SEE-Phone

Two functions have been implemented in SEE-phone which are useful for assistance for the VIP space awareness assistance: color detection using smart phone (for traffic light status recognition) and remote route tracking (for continuous check of the VIP itinerary).

Experiments have been implemented on Android emulator system (running on personal computer) and on smart phone [Yusro et al. 2013]. For traffic light recognition – a camera of smart phone Samsung SIII Model GT-I9300. The whole process can be summarized as follows:

- Capture the color image of the traffic light;
- Calculate the average of HSV (a reference data) of the captured image;
- When the value of the colors (traffic lights) captured by the SEE-phone camera is in the range of reference data, the voice application TTS (*Text To Speech*) generates an audio message : "*red stop here*", "*yellow slowly run*" or "*green please run away*".

The results of the traffic light status detection are displayed in figure 10. The upper images (left to right) are signs of traffic lights status for pedestrian. The lower images (left to right) show results of color detection by SEE-phone. In this experiment, android application program detects three colors: red, yellow and green.

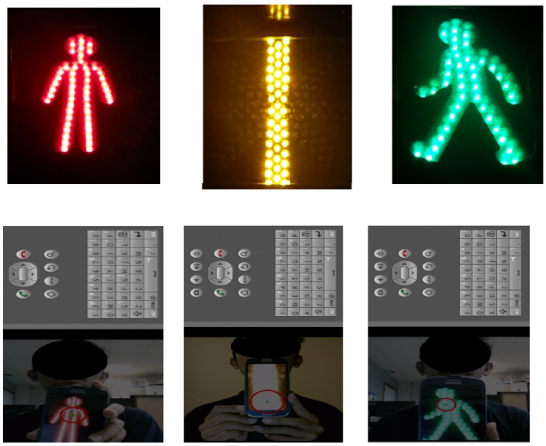


Fig. 10 The color detection on android emulator

5 Conclusion

This paper has introduced the concept of the SESS, a space explorer smart stick, for the visually impaired mobility assistance, and its first prototype design. This concept is based on – the space awareness and orientation capabilities – key elements for (the VIPs) independent mobility.

The architecture of the SEES system, an example of SESS, has been proposed; it integrates three sub-systems which complement each other: iSEE, SEE-phone (based on a commercial smart phone) and SEE-stick (based on the white cane).

The realization of a SEES involves various sensors which work and cooperate simultaneously such as ultrasonic sensors, visible spectrum cameras, compass sensor, accelerometer sensor, wheel encoder sensor and GPS receiver. The SEES system preliminary evaluations for the VIP tracking while walking in simple scenarios and for traffic light status recognition have been successful; the real track provided by the RISS-GPS-LKF algorithm was close to the reference track and the traffic light status was correctly recognized.

Future works will tune the proposed SEES overall design, will add and modify assistances, will evaluate its technical performance and will design a pilot studies which will involve the targeted population with different degrees of visual impairment (participative design).

Acknowledgment. This work has been sponsored by the Indonesian and French governments, and French government research program "*Investissements d'avenir*" through the IMobS3 Laboratory of Excellence (ANR-10-LABX-16-01), by the European Union through the program Regional Competitiveness and Employment 2007-2013 (ERDF–Auvergne region), and by the Auvergne region.

References

[Borenstein, Ulrich 2001] Borenstein, J., Ulrich, I.: The guidecane - applying mobile robot technologies to assist the visually impaired. IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans 31(2), 131–136 (2001)

[Chang et al 2005] Chang, T.-H., Ho, C.-J., Hsu, D.C., Lee, Y.-H., Tsai, M.-S., Wang, M.-C., Hsu, J.: *iCane* – A partner for the visually impaired. In: Enokido, T., Yan, L., Xiao, B., Kim, D.Y., Dai, Y.-S., Yang, L.T. (eds.) EUC-WS 2005. LNCS, vol. 3823, pp. 393–402. Springer, Heidelberg (2005)

[EBU 215] European Blind Union, `http://www.euroblind.org/resources/information/nr/215` (accessed February 28, 2014)

[Farcy et al 2006] Farcy, R., Leroux, R., Jucha, A., Damaschini, R., Grégoire, C., Zogaghi, A.: Electronic travel aids and electronic orientation aids for blind people: technical, rehabilitation and everyday life points of view. In: Proc. Conf. and Workshop on Assistive Technologies for People with Vision and Hearing Impairments Technology for Inclusion (2006)

[Hoyle 2006] Hoyle, B.S.: The UltraCane mobility aid at work - from training programmes to in-depth use case studies. In: Proc. Euro-Assist Conference and Workshop on Assistive Technologies for Vision and Hearing Impairment (2006)

[Lafuma et al 2006] Lafuma, A.J., Brézin, A.P., Fagnani, F.L., Mesbah, M., Berdeaux, G.H.: Prevalence of visual impairment in relation to the number of ophthalmologists in a given area: a nationwide approach. Journal of Health and Quality of Life Outcomes 4, 34 (2006)

[Loomis et al 2005] Loomis, J., Marston, J.R., Golledge, R.G., Klatzky, R.L.: Personal guidance system for people with visual impairment: a comparison of spatial displays for route guidance. J. Vis. Impair Blind 99(4), 219–232 (2005)

[MHRI 2013] Ministry of Health Republic of Indonesia (2013), `http://www.depkes.go.id/index.php/berita/press-release/2191` (accessed February 28, 2014)

[Palleja et al. 2010] Pallejà, T., Tresanchez, M., Teixidó, M., Palacin, J.: Bioinspired Electronic white cane implementation based on a lidar, a tri-axial accelerometer and a tactile belt. J. of Sensors 10, 11322–11339 (2010)

[Pissaloux 2009] Pissaloux, E.: Assistive technologies to support independent mobility of visually impaired. In: Hersh, M.A. (ed.) Proc. Conf. and Workshop on Assistive Technologies for People with Vision and Hearing Impairments, Past Successes and Futures Challenges, CVHI 2009 (2009)

[Pissaloux 2011] Pissaloux, E.: Mobilité de déficients visuels: Concepts, modèles computationnels et assistances technologiques. In: Proc. 2e Int. Conf. Sur l'Accessibilité et les Systèmes de Suppléance aux personnes en situations de Handicap, pp. 306–315 (2011)

[Rohan et al 2007] Rohan, P., Ankush, G., Vaibhav, S., Dheeraj, M., Balakrishnan, M., Kolin, P., Dipendra, M.: 'Smart' cane for the visually impaired: technological solutions for detecting knee-above obstacles and accessing public buses. In: Proc. 11th International Conference on Mobility and Transport for Elderly and Disabled Persons (2007)

[Stillman, Essa 2001] Stillman, S., Essa, I.: Toward reliable multimodal sensing in aware environments, perceptual user interfaces. In: Workshop held in conjunction with ACM UIST 2001 Conference, Orlando, Florida, pp. 15–16 (November 2001)

[Shi et al 2011] Shi, H.L., Hou, K.M., ZhouH, L.X.: Energy Efficient and fault tolerant multicore wireless sensor network: E2MWSN. In: Proc. 7th Inter. Conf. on Wireless Communications, Networking and Mobile Computing, pp. 1–4 (2011)

[Terlau et al. 2006] Terlau, T., Penrod, W.M.: 'K' sonar curriculum handbook. American Printing House for the Blind (2008)

[Tversky 2005] Tversky, B.: Functional significance of visuospatial representations. In: Shah, P., Miyake, A. (eds.) Handbook of Higher-Level Visuospatial Thinking. Cambridge University Press (2005)

[Willough, Monthei 1998] Willough, D., Monthei, S.: Modular Instruction for Independent Travel for Students Who are Blind or Visually Impaired: Prweschool through High School (1998), `https://nfb.org/images/nfb/publications/fr/fr27/2/fr270213.htm` (accessed February 28, 2014)

[WHO 2011] WHO, Visual Impairment and Blindness. Fact sheet 282 (2011), `http://www.who.int/mediacentre/factsheets/fs282/en/index.html` (accessed February 28, 2014)

[Yusro et al 2013] Yusro, M., Hou, K.M., Pissaloux, E., Ramli, K., Sudiana, D., Lizhong, Z., Shi, H.L.: Design and implementation of SEE-Phone in SEES (Smart Environment Explorer Stick). In: Proc. Inter. Workshop of New Information Communication, Science and Technology for Sustainable Development (2013)

[Yibo et al 2011] Yibo, C., Hou, K., Zhou, H., Shi, H.L., Liu, X., Diao, X., Ding, H., Li, J., de Vaulx, C.: 6LoWPAN stacks: A survey. In: Proc. 7th Inter. Conf. on Wireless Communications, Networking and Mobile Computing, pp. 1–4 (2011)

[Zhang et al 2013] Zhang, L.Z., et al.: A smart environment explorer stick (SEES): concept and design of its orientation and walking sub-system. In: Proc. Inter. Workshop of New Information Communication, Science and Technology for Sustainable Development (2013)

GiraffPlus: A System for Monitoring Activities and Physiological Parameters and Promoting Social Interaction for Elderly

S. Coradeschi[1], A. Cesta[2], G. Cortellessa[2], L. Coraci[2], C. Galindo[3], J. Gonzalez[3], L. Karlsson[1], A. Forsberg[8] , S. Frennert[7], F. Furfari[4], A. Loutfi[1], A. Orlandini[2], F. Palumbo[4], F. Pecora[1], S. von Rump[5], A. Štimec[6], J. Ullberg[1], and B. Ötslund[7]

[1] Örebro University, AASS, Sweden
name.surname@oru.se
[2] CNR-ISTC, Rome, Italy
name.surname@istc.cnr.it
[3] Malaga University, Spain
jgonzalez@ctima.uma.es, cgalindo@ctima.uma.es
[4] CNR-ISTI, Pisa, Italy
name.surname@isti.cnr.it
[5] Giraff AB, Sweden
stephen.vonrump@giraff.org
[6] XRPOISD, Slovenia
ales.stimec@xlab.si
[7] Lund University, Sweden
susanne.frennert@certec.lth.se, britt.ostlund@design.lth.se
[8] Örebro Landstinget, Sweden
anette.forsberg@orebroll.se

Abstract. This chapter presents a telehealth system called GiraffPlus supporting independent living of elderly in their own home. GiraffPlus system is a complex system which monitors activities and physiological parameters in the home using a network of sensors. The elaborated information is presented to the primary user, the elderly, and to secondary users like health care and home care providers and possibly to family members as a help to assess possible health and wellbeing deterioration, provide acute alarms, and support health procedure. The secondary users can also visit the elderly via the Giraff, a teleoperated robot that can communicate and move in the home under the control of the secondary user. The chapter focusses in particular on the deployment of the system in six real homes in Sweden, Italy and Spain. The chapter outlines the technological various components used, the expectations of the users and the evaluation method.

1 Introduction

An emerging application area for intelligent systems is the use of such system in the home and for domestic use. In particular, applications which concern individuals

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_22, © Springer International Publishing Switzerland 2014

with special needs, disability or in general promote an improved life quality are particularly interesting for these new technologies. This application area poses a number of key challenges from a human systems interaction perspective. Such a system should be in operation for long periods of time and should interact and be used by non-experts which may even suffer from cognitive decline. The system should also be able to cope with unexpected situations or events for which it may not have been pre-programmed. Therefore, there are not only technological challenges to deploying such systems at home, but also the ability to thoroughly evaluate such system is essential in order to ensure that the system can fulfill the needs of the intended users. In this chapter we present a system developed with an EU project called GiraffPlus. GiraffPlus is a complex system which can monitor activities and physiological parameters in the home using a network of sensors distributed in the home. The sensors can measure e.g. blood pressure or detect e.g. whether somebody falls. They can also assess where the person is, if is sitting in the couch or lying in bed, if he is using electrical appliances or opening cupboards or the fridge. Different services, depending on the individual's needs, can be pre-selected and tailored to the requirements of both the older adults and health care professionals. At the heart of the system is a unique telepresence robot, Giraff, which lends its name to the project. The robot uses a Skype-like interface to allow e.g. relatives or caregivers to virtually visit an elderly person in the home (see Fig.1). The GiraffPlus system monitors daily activities of the elderly and provides a visual interface which allows care givers to see a summary of the activities. The physiological data can also be viewed in the context of the daily activity of the users. The monitoring is complemented with a social interaction component where health professionals and elderly can discuss the collected data during visits made via the Giraff robot.

Fig. 1 The GiraffPlus system deployment in a home

In the GiraffPlus system, the interaction between humans and the system occurs with several modalities. Firstly, there is an interface that shows the data collected by the system to health and homecare professionals and can be used both offline or during the visits with the Giraff. Secondly, family and friends are able to connect to the system and see parts of the information according to their level of authorization to the data. Thirdly, the elderly has the possibility to see his or her own data which has been collected by the system. Finally, all the information is stored in a database and can be retrieved for any specific time window enabling therefore long term analysis of the data.

Special emphasis in the project is given to evaluations and input from the users so that the system can have an empathetic user interaction and address the actual needs and capabilities of the users. The GiraffPlus system is already installed in 6 homes of elderly people distributed in Sweden, Italy and Spain and will be installed and evaluated in a total of 15 homes. These evaluations will drive the development of the system. This chapter presents the user expectations, the preliminary evaluations of the system and the deployment of the first version of the system in the six homes. The chapter also provides the initial evaluation results after a first period of use.

The chapter is organized as follows: A general description of the technological components of the GiraffPlus system is given in Section 2. Section 3 provides a description of the selected test sites, the evaluation methodology and the deployment of the system. Finally a discussion, conclusion and outline of planned work in the project are given in Section 4.

2 The Deployed System

In this section we briefly present the GiraffPlus system. A more detailed description of the system is in [Coradeschi et al. 2013]. The **Giraff telepresence robot** uses a Skype-like interface to allow caregivers to virtually visit an elderly person in the home. The Giraff robot is enhanced with semi-autonomy in order to increase safety and ease-of-use. The GiraffPlus system also includes **a network of sensors.** Data from these sensors are processed by an **advanced context recognition system** based on constraint-based reasoning which both detects events on-line and can perform inference about long term behaviors and trends. **Personalized interfaces for primary and secondary users** are developed to access and analyze the information from the context recognition system for different purposes and over different time scales. An important feature of the system is an infrastructure for adding and removing new sensors seamlessly, and to automatically **configure** the system for different services given the available sensors. This is done using planning techniques. These features provide an adaptive support which facilitates timely involvement of caregivers and allows monitoring relevant parameters only when needed.

The specific instance of the GiraffPlus system that has been deployed at the test sites includes the components described below.

2.1 Hardware Components and Middleware

The kind of sensors deployed in the test sites are described below. Fig. 2 shows the sensors. For each test site a specific group of sensors is selected depending on the home topology.

Passive Infrared Detectors (PIR). The passive infrared detectors are wireless movement detectors that can be used to detect the presence of the user in a particular room. Once a movement is detected, the PIR sensor generates a signal to the GiraffPlus middleware. This information is essential to register and monitor the activity of the user.

Fig. 2 Presence sensor, Electrical power consumption, Occupancy sensor, universal sensor, door contact

Electrical Usage Sensor. This sensor sends a radio signal to the GiraffPlus middleware, when a connected electrical appliance is switched on/off. It is intended to indicate when primary users have used a particular appliance, such as a kettle, microwave, etc.

Occupancy Sensor (bed/sofa/chair). Occupancy sensors are pressure pads that can be placed on mattresses, sofas, chairs, etc. to check the time the user is resting. They need to be connected to a universal sensor to transmit the information to the Giraffplus middleware.

Door Contact. This sensor detects when two objects are close. It is typically used for doors, windows, cabinets, etc. to detect when they are opened/closed. It needs to be linked to a universal sensor in order to transmit the information to the middleware.

Universal Sensor. The Universal sensor enables wired devices and other equipment to raise wireless alarm calls and appropriate radio messages to home units using Plug and Play functionality. When the attached device is triggered, the Universal sensor sends the appropriate message, permitting the elements connected to the GiraffPlus middleware, e.g. the Context Recognition module, to be aware about the current situation.

Physiological Sensors. The physiological sensors include a kit composed by an android tablet and assessment kit that measures weight, blood pressure and glycaemia.

Android-Based Mobile Phone and a Pulse Oximeter. The main purpose of the sensor is a continuous monitoring of physiological parameters (pulse rate and oxygen saturation) and fall detection using the accelerometer in the mobile phone.

Giraff Telepresence Robot. The Giraff robot is a telepresence device that allows remote users to connect and make a virtual visit. It includes a zoom camera with a wide-angle lens, microphone, speaker and a 13.3" LCD screen mounted on top of a base that the visitor can control remotely and move around in the remote location. The 13.3" LCD can turn 180 degrees to let the user see who is calling and to interact with the Giraff. The device is specifically designed to work in home settings and provides visual/social cues for the user, such as lowering to a sitting position when at a table or bedside. The Giraff robot can be seen in Fig.1.

The user interface on the Giraff robot contains an answer button, hang-up button, volume knob and a touchscreen so the user can interact with the system; everything else is controlled by the remote user to simplify the robot usage, and in fact it does require any interaction from the user to operate. The Giraff robot also includes a charging station where it charges its battery when not in a call.

The **middleware** provides communication functionalities between software components deployed at home and on secondary user's personal computer.

Heterogeneity and distribution of both hardware and software resources is hidden by the middleware. Instead of providing different components at different levels of the system to access storage and context recognition functionalities, the current middleware architecture includes dedicated modules at the same level. Similarly, any device configured at home can be accessed in a homogeneous way by local and remote services.

The buses used to announce and discover resources installed at home (service bus) and the buses to access sensors' readings (context bus) need to be managed by a message broker that is installed on the same home server and it is accessed through the VPN by the remote clients.

In addition to communication capabilities, the current middleware layer presents APIs to retrieve historical and real-time sensor data and to query for activities in the monitored home environment.

A **remote Database** is used where all the sensor data and the elaborated information is stored and easily accessed by all components of the system.

2.2 Context Recognition and Configuration Planner

Context recognition and the configuration planner are important components of the system in order to give meaning to the sensor data. The former recognizes the activities performed in the home (e.g. eating-dinner, sleeping, food-cooking, and moving) on the basis of the sensors data. An activity has an extension in time, and it typically involves changes in observable state variables. The task of the configuration planner is to configure the sensor network in terms of subscriptions (and possibly also giving parameters to functionalities) so that the state variables requested by the context recognition system are monitored.

Given a task, the **configuration planning** is the problem of generating a functional configuration of a networked robot system consisting both of mobile robots and sensors and actuators distributed in the environment that solves that task. In a functional configuration, sensory, computational and motoric functionalities belonging to the different devices are connected with communication channels.

The work by [Lundh et al. 2008a], [Lundh et al. 2008b], [Lundh et al. 2009] offers a centralized solution to the problem, which has been successfully employed in an intelligent home test bed. A related approach for multi-robot systems is AsyM TRe-D [Tang et al. 2005], [Parker et al. 2006]. ASyMRTe-D works by connecting functionalities, belonging to different robots, into coalitions. The GiraffPlus configuration planner may produce configurations that change over time, in particular when actuation (pre- and post-conditions of actions) is involved. The configuration planner makes the sensor network providing the information (state variable values) that the context recognition system needs, and may also change state variables through actuation. A particular important capability of the GiraffPlus configuration planner is to handle many requests at the same time.

Context recognition is a fundamental capability of the proposed system. Context recognition addresses the general goal of providing the system with the capability of inferring context about the human user(s) and of the environment, including its ubiquitous sensors and actuators. In the particular sub-case of human context recognition, this capability often goes under the name of activity recognition. The context recognition system deals with activities. An activity has an extension in time, and it typically involves changes in state variables (e.g. eating-dinner, sleeping, food-cooking, and robot-moving). Similar constraint-based approaches to inference in the domain of domestic activity monitoring include [Cesta et al. 2010], [Pollack et al. 2003].

Given the requests from services, the context recognition system determines what state variables need to be monitored and requests these variables from the configuration planning system. It then continuously receives data about these state variables from the sensor network, enters them into the time line and derives activities from them. The context recognition system may also change its request for state variables to the configuration planner, depending on what activities comes up or are expected to come up. And it may request changes of state variables (actuation) when certain activities are triggered.

2.3 Interacting with the User

The GiraffPlus system uses different modalities to interact with the users. The secondary users receive reports of the activities and physiological parameters and can see these information on-line via an interactive program (Fig 3 A and B). The elderly can see the information in the Giraffe robot touch screen (Fig 3 C and D).

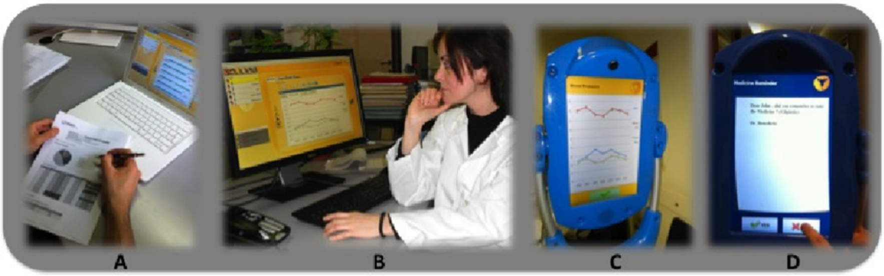


Fig. 3 Interaction modalities of the GiraffPlus system

The secondary users can also connect to the Giraff robot using the interactive tool (Fig. 4 left). Fig. 3 B and C shows a graphical representation of the physiological data, while Fig. 4 right shows a representation of the activities performed in a period of time.

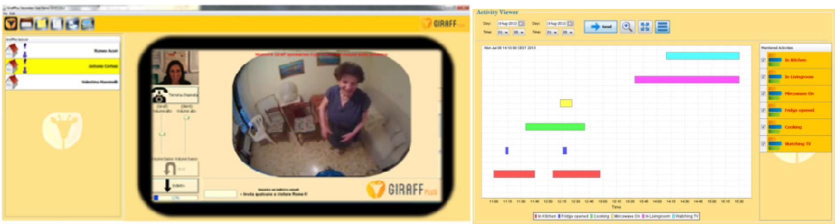


Fig. 4 (left) tool to connect to the Giraff robot; (right) representation of activities

3 Long Term Evaluation

3.1 User Requirements

The GiraffPlus system is being developed on the basis of user requirements from both primary (elderly living in their apartment), and secondary (health care professional or family members and friends) users. Overall a **total of 325 users** (among primary and secondary) have been involved in this initial phase. Qualitative and quantitative research has been carried out in the Sweden, Spain and Italy, to elicit user requirements and expectations in terms of type of services as well as system design and appearance. The study started with a literature review, followed by focus groups and questionnaires. Specifically, the results of the focus groups have been validated via questionnaires and workshop with users. Some qualitative cross-cultural analysis has also been performed in order to highlight differences emerged during the studies in the three countries. Result of this effort is a list of user requirements with an associated level of *priority* and a set of preferences on different mockups of a component of the system that can be both used to influence the future architecture definition and functional specification of the GiraffPlus system. This comprehensive list of requirements is not only of interest for the specific GiraffPlus system, but also outlines general preference of primary and secondary users with respect to technological solution. The complete list is available at the GiraffPlus website (www.giraffplus.eu) in deliverable D1.1 in the “Project” tab.

3.2 User Expectations

The project's guiding principle is that it should be driven forward by user centered design. In this case, it means active involvement during the entire development of cycle ranging from the initial prototype in the laboratory, the intermediate version installed in the homes and the final complete system.

In the first year, efforts were made to create a group of older users that serves as test subjects throughout the project. This was based on earlier experiences that such a strategy creates more engaged participants willing to serve as proactive users and being partly responsible for the result compared to being merely objects for testing. The advantage for the project manager is a secure process that creates stability and prevents drop outs and problems in case of projects delays.

Initially several workshops were organized inviting potential elderly users to proactively elaborate on the concept playfulness and its bearing on inviting telehealth systems and its artifacts into the home. The results showed the importance of considering cultural differences both in terms of design and in terms of monitoring the contacts between patients and caregivers. Minimalist design was more desirable in Sweden since it was experienced as discrete and stylish and possible to personalize. In Italy and Spain users preferred furniture like design that was experienced as offering more functionality. The knowledge of the variation in health care systems was found important to facilitate adjustment of telehealth care services to national and local conditions.

The next step was to test a preliminary GiraffPlus version in three separate evaluations. First with eleven elderly users in a test flat in Sweden, second, with six professional nurses in Italy, and third, installing a pilot in one real life setting – a home of an elderly user in Italy.

The main aim in the test flat was to identify users expectations and identify problems to facilitate the adoption process and pave the way for the next step i.e. installation of the system in homes of elderly users. Telehealthcare research most often has a technological focus without any chance to understand the user’s point of view. Such an approach neglects the experiences of users which in turn limits the ability to implement new products and services in a meaningful way in their everyday lives. Based on the user requirements derived from the initial literature study and focus groups described above elderly users were exposed to three scenarios: the situation of de-hospitalization; daily activity monitoring by a care giver; and monitoring a physiotherapy protocol. The analysis of the expectations was guided by Rogers´s diffusion of innovation theory [Roger 2003] which helped to understand the adoption or rejection of the GiraffPlus system. An innovation is likely to become successfully adopted if it is perceived as better than the idea/product it succeeds – *relative advantage*; if it is perceived as being consistent with exciting values, past experience and the person's needs – *compatibility*; if it is perceived as easy to use – *complexity*; and if the use of the innovation is visible and liked by others – *observability*. The result shows that the crucial factor for adoption of telehealthcare systems is not usability but the system's ability to really support autonomy in everyday life. Results from the user lab tests indicate that a telehealthcare system needs to: (a) be customized to the individual's needs, (b) be reliable, (c) be easy to use, (d) consist of as few devices as possible, and (e) offer training and education throughout the usage. The relative advantage of the system with respect to possible alternatives depends on how well integrated the telehealthcare system can become with the current healthcare system.

The assessment of the prototype in Italy involved nurses as secondary users. They were exposed to the same scenarios as the elderly users in the test flat in Sweden. They also assessed the first version of the interface for the Interaction and Visualization Services, see figure 4. Their main concern was how to integrate the system in their daily work. The means through which GiraffPlus is accessible is crucial in this respect: it needs to fit into their daily work routines and it challenges the way health care is organized. An additional aspect outlined is that to serve the patients the information need to be contextualized; a holistic perspective requires patient information that is beyond singular data, but considers a more general context. For instance, a blood pressure value can be influenced by the activities the person has done before.

The pilot study in an elderly home in Italy helped to make the preparations for implementation more realistic and effective. The main factors emerging in the pilot study were technology weaknesses, for example that sensors were more difficult to install than expected. In addition organizational and maintenance issues emerged, for example the previous relation between those visiting the home and the elderly impacted the willingness to regularly use the technology. There was also the need to adapt the expectations of technology reliability between a lab setting and a real life setting where reliability is much harder to achieve. The lessons learnt from this pilot illustrates very well the theoretically based fact that usefulness and meaningfulness is context based.

To conclude, the first evaluation of the Giraffe Plus emphasizing expectations shows the difficulty of assessing the value of innovation outside its context. It demonstrates the need to understand the daily life of the organization or the individual where it will be implemented. It also shows the value of involving users, both elderly and care givers, and not least, to test the system both in a laboratory and in a real life pilot and most important, to understand the difference between the settings and adapt the technology to the real home.

3.3 Deployment of the System in the Test Sites and Evaluation

The GiraffPlus system has been deployed in 6 homes. For each country (Italy, Spain and Sweden) two homes have been selected and a version of the system has been deployed. In Sweden the test sites are the homes of two single men both 83 and 82 years old. In Spain two men 84 and 76 years old. In Italy two women 93 and 94 years old. An example of the sensor deployment in a home in Sweden is shown in Fig. 5.

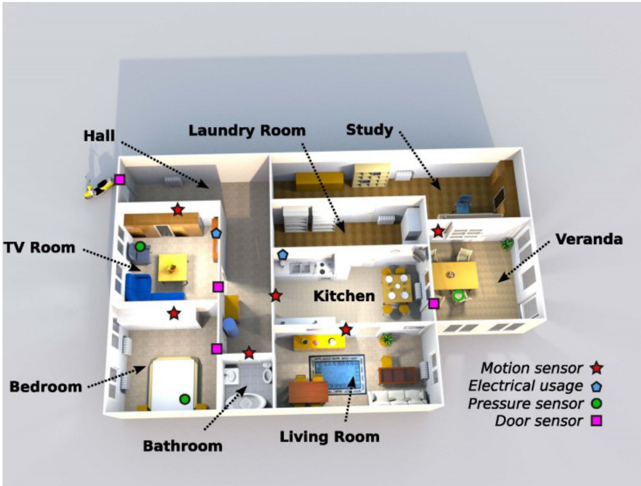


Fig. 5 Sensors deployed in a test site in Sweden

An interview has been performed for each test site to assess the expectation of the person and to decide which sensor to deploy and where depending of what was expected to be monitored. The expectations that emerge from the interviews show the presence of both drivers and barriers with respect to the use of the system. No technological barriers were considered especially significant pointing to the fact that the usability aspects of the system components are satisfactory. One barrier addressed in Spain was worries about the "big brother syndrome" i.e. the surveillance by unknown people. One driver that overshadows the barriers is the positive experience of being involved in a project and their own experiences being taken seriously.

We are in the process of performing additional interviews after an initial period of using the system.

4 Discussion and Conclusions

The GiraffPlus system aims at facilitate a prolonged independent living for elderly in their own home. The system monitors daily activities and physiological parameters and can give a more reliable view of the health status of the person both for the person himself and for health/care personal and family. The system can also give alarms and present long term data that can show health deterioration.

An essential aspect to be considered is a thorough user evaluation both in terms of user expectations and user feedback when using the system. We briefly present in the paper the methodology for acquiring the user requirements and we refer to a report available on-line for the full list of requirements due to space limitations. We also describe the acquisition of user expectations prior of the system deployment. Currently, the system has been recently deployed in 6 test sites where it will be evaluated for one year. 9 additional test sites will be started in February 2014. The system deployed is described in the paper. A first feedback from the users of the system is currently collected and can also be presented in the final version of the paper.

Acknowledgment. The project is supported by the Fp7 EU programme.

References

[Cesta et al. 2010] Cesta, A., Cortellessa, G., Rasconi, R., Pecora, F., Scopelliti, M., Tiberio, L.: Monitoring elderly people with the robocare domestic environment: Interaction synthesis and user evaluation. Computational Intelligence, Special Issue on Scheduling and Planning Applications (2010)

[Coradeschi et al. 2013] Coradeschi, S., Cesta, A., Cortellessa, G., Coraci, L., Gonzalez, J., Karlsson, L., Furfari, F., Loutfi, A., Orlandini, A., Palumbo, F., Pecora, F., von Rump, S., Štimec, A., Ullberg, J., Östlund, B.: GiraffPlus: Combining social interaction and long term monitoring for promoting independent living. In: Proc. 6th Inter. Conf. on Human System Interaction, pp. 578–585 (2013)

[Lundh et al 2008a] Lundh, R., Karlsson, L., Saffiotti, A.: Autonomous functional configuration of a network robot system. Robotics and Autonomous Systems 56(10), 819–830 (2008)

[Lundh et al 2008b] Lundh, R., Karlsson, L., Saffiotti, A.: Automatic configuration of multi-robot systems: planning for multiple steps. In: Proc. 18th European Conference on Artificial Intelligence, Greece, pp. 616–620 (2008)

[Lundh et al 2009] Lundh, R.: Robots that help each-other: self-configuration of distributed robot systems. PhD Thesis. Örebro University, Sweden (2009)

[Parker et al 2006] Parker, L.E., Tang, F.: Building multi-robot coalitions through automated task solution synthesis. Multi-Robot Systems 94(7), 1289–1305 (2006)

[Pollack et al 2003] Pollack, M., Brown, L., Colbry, D., et al.: Autominder: an intelligent cognitive orthotic system for people with memory impairment. Robotics and Autonomous Systems 44, 273–282 (2003)

[Rogers 2003] Rogers, E.M.: Diffusion of innovations, 5th edn. Simon and Schuster (2003)

[Tanget al 2005] Tang, F., Parker, L.E.: Distributed multi-robot coalitions through ASyMTRe-D. In: Proc. IEEE/RSJ Inter Conf on Intelligent Robots and Systems (IROS), Alberta, pp. 2606–2613 (2005)

Part III
Various Information Processing Systems

Textual Coherence in a Verse-Maker Robot

A. Astigarraga[1], E. Jauregi[1], E. Lazkano[1], and M. Agirrezabal[2]

[1] Department of Computer Science and Artificial Intelligence,
University of the Basque Country (UPV/EHU), Donostia, Spain
{aitzol.astigarraga,e.lazkano,ekaitz.jauregi}@ehu.es
[2] Department of Language and Software systems,
University of the Basque Country (UPV/EHU), Donostia, Spain
Manex.aguirrezabal@ehu.es

Abstract. The *Bertsobot* project aims to develop an autonomous robot capable of composing and playing traditional Basque impromptu verses –*bertsoak*. The system should be able to construct novel verses according to given constraints on rhyme and meter, and to perform it in public. The *Bertsobot* project, at the intersection of Autonomous Robotics, Natural Language Generation and Human Robot Interaction, works to model the human abilities that collaborate in the process that enables a verse-maker to produce impromptu verses. This paper provides a general overview of the system, specially focusing on the description and evaluation of different semantic similarity methods for predicting the textual coherence of the generated verses.

1 Introduction

Basque, *Euskara*, is the language of the inhabitants of the Basque Country. And *bertsolaritza*, Basque improvised contest poetry, is one of the manifestations of traditional Basque culture that is still very much alive. Events and competitions in which improvised verses, *bertso*-s, are composed are very common. In such performances, one or more verse-makers, named *bertsolari*-s, produce impromptu compositions about topics or prompts which are given to them by an emcee (theme-prompter). Then, the verse-maker takes a few seconds, usually less than a minute, to compose and sing a poem along the pattern of a prescribed verse-form that also involves a rhyme scheme. Melodies are chosen from among hundreds of tunes. Fig. 1 shows a picture of the national championship of *bertsolaritza*, which took place on 2009.

Xabier Amuriza, a famous verse-maker that modernized and contributed to the spread out of the *bertsolaritza* culture, defined *bertsolaritza* in a verse as:

Neurriz eta errimaz	*Through meter and rhyme*
kantatzea hitza	*to sing the word*
horra hor zer kirol mota	*that is what kind of sport*
den bertsolaritza.	*bertsolaritza is.*

The main objective of this research project is to develop an autonomous robot capable of generating and performing improvised verses in Basque. The interaction

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_23, © Springer International Publishing Switzerland 2014

with the robot should be speech-based; thus, the system should be able to receive the instructions to compose the verse in Basque, to generate the most appropriate verse according to the given instructions and to sing it with the proper melody. The robot should also show the degree of expressiveness that Basque troubadours, *bertsolari*-s, show in their performance. And all those tasks must be accomplished concurrently in a extemporaneous performance.

We believe that the *Bertsobot* project provides a unique opportunity to join together the capabilities of autonomous robots to sense their environment and interact with it, and the natural language processing tools devoted to automatic verse generation, in an attempt to generate improvised and context-specific poetry.

We decided to decompose the task of performing a *bertso* into several smaller tasks or abilities, so that each one might be modeled correctly before attempting to combine them in the composition of the overall performance.

Fig. 1 2009 national championship

The main concern of this paper is to measure the textual coherence of a poem automatically generated under meter and rhyme constraints.

2 Related Work

Computer-based poetry has received attention in the research community in the last years, and several interesting systems have been developed. In our opinion, automatic generation of poems resembles the creation of a *bertso* in these three aspects:

1. They have to satisfy very specific technical requirements: on the one hand metric restrictions, that is, the number of syllables per line; and, on the other hand, they have to meet certain rhyme pattern.
2. They are allowed a certain poetic license which implies, sometimes, deviations from the rules of syntax and semantics.
3. The result must be meaningful for the user, more specifically, the resulting text must arouse specific emotions amongst the audience.

Taking aforementioned similarities in mind, we reviewed the existing literature on the automatic creation of poetry. A good overview can be found in [Oliveira 2009; Gervas 2013]. Most recent -and relevant- ones include:

- **Haiku Generation Using Vector Space Model.** Wong and Chun [Wong et al. 2008] presented an approach to generate "modern haikus" using text collected from blogs. The proposed approach uses a keyword and a line repository containing sentence fragments found in blogs. The haiku generation process starts by choosing three keywords from the lexicon, and then, sentence fragments in the blogosphere containing those keywords are searched. Two keywords are extracted from each sentence, using a *tf-idf* weighting scheme to evaluate how important each word is to a sentence. Afterwards, vectors are used to compare each sentence pair. The cosine of the angle between two vectors is used to measure their semantic relation. Finally, the most semantically related pairs of vectors are chosen for the resulting haiku.
- **Full-FACE Poetry Generation.** This system [Colton et al. 2012] is a corpus-based poetry generation system, aimed to be a fully automatic computer poet. The generation is driven by a four stage process that involves: simile retrieval from a simile corpus, generation of variations for each simile, combination --they combine similes with their variations and key phrases extracted from newspaper articles-- and instantiation, which chooses one of the possible combinations randomly.
- **Computational Modelling of Poetry Generation.**In this paper [Gervás 2013] Gervás shows a review of different works done in this area and presents a redesigned version of the well known WASP poetry generator which combines different AI techniques. The poems generated by this system have been accepted for publication in a book about the possibilities of computers writing love poems.

To the best of our knowledge, there are no poetry generation systems implemented in a real robot. But *bertsolaritza* belongs to the oral genre, and the public performance is extremely important. Therefore, it is not enough the development of an automatic verse generation system, the created poem has to be part of a performance. Thus, a real body that interacts with the public and sings the improvised verse with a proper melody is needed.

The design and implementation of such an autonomous interactive robot is a challenging task. There is much to learn from the potential of the integration of these technologies, as there is from the embodied and speech-based human robot interaction [Beck et al. 2010; Scheutz et al. 2011].

3 System's Task

Bertsos can be composed in a variety of settings and manners. At a formal competition the verse-makers are called upon to compose and sing different kinds of bertsos by the theme-prompter or emcee. For our particular challenge we have chosen the popular exercise of "Rhymes Given": the robot is given the four rhyming words and it is required to compose the bertso in Zortziko Txikia meter "around" these rhyming words.

Zortziko Txikia (see Fig. 2) is a composition of eight lines in which odd lines have seven syllables and even ones have six. The union of each odd line with the next even line, form a strophe. Each strophe has 13 syllables with a caesura after the 7th syllable (7 + 6) and must rhyme with the others.

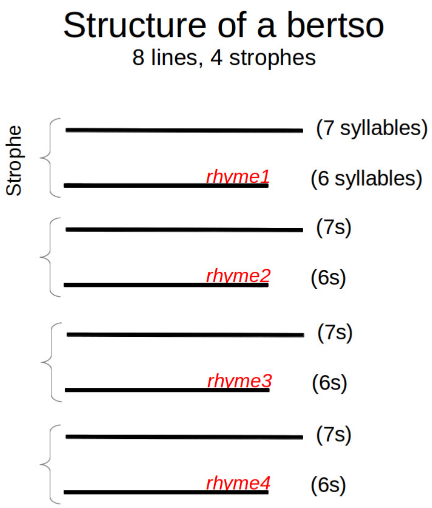


Fig. 2 Structure of the Zortziko Txikia

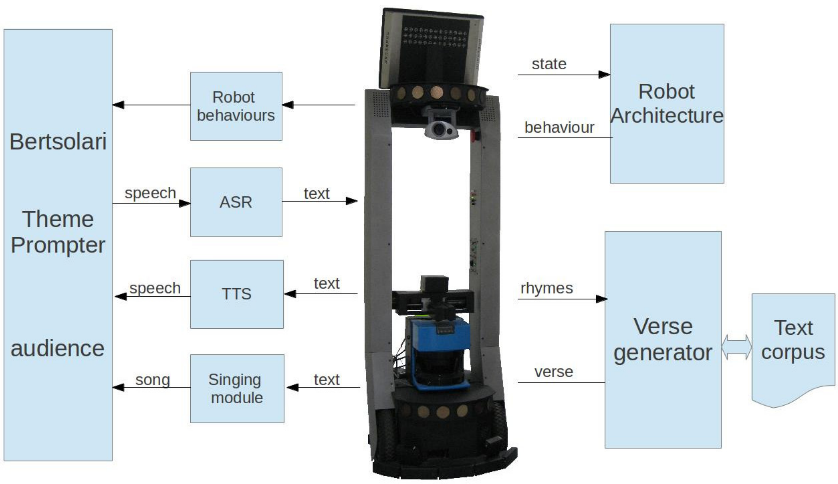


Fig. 3 Overall software architecture

It must be said that human-robot interaction is entirely speech-based: the theme-prompter (or emcee) proposes the four rhymes to compose the verse in Basque, and the system answers in Basque too. The overall performance is represented as natural as possible by the robot, interacting as much as possible with the emcee and the audience (see [Astigarraga et al. 2013] for more details).

4 Verse Generator

As it can be deduced from Bertsobot's software architecture (see Fig. 3), several research areas are involved in the project. This paper concentrates on the verse generation subpart of the project, which generates a poem text under meter and rhyme constraints and using two methods to achieve textual coherence.

In the basic scenario, the four rhymes to compose a *bertso* are received as input, and the verse generator module then should give as output a novel and technically correct verse, and (hopefully) with coherent content.

The poetry generator is based on the following principles:

- **Form**: rhyme and metric compound the technical requirements of a *bertso*. Thus, finding rhymes and counting syllables are essential abilities that the system must perform.
- **Content**: the output of the verse generator module must be meaningful. Methods to measure the semantic coherence of the generated text are needed.

It can be said that the final verse is a marriage of form and content in which each works together to convey a meaningful text with formal precision.

The verse generation process then consists of the following steps:

1. Receive as input the four rhymes to compose the verse.
2. Find sentences in the corpus that rhyme with the input words.
3. Generate the verse with the highest textual coherence.

We next describe in some more detail the background corpus construction process as well as the similarity methods used to measure textual coherence.

4.1 Text Corpus

Our approach for verse-generation is corpus based. We have chosen initially to work with documents mined from the Basque newspaper *Berria* [WWW_1] (%85) alongside verses extracted from the work of well-known *bertsolari*-s [WWW_2] (%15). The later reflects the desire to maintain the language-model of the *bertsolaritza*, and the former tends to increase quality while not appropriating text intended for poems. Next, we divided the text collection into passages with coherent meaning. For our purposes, we selected as document the minimum semantically coherent block of text, where each document deals with a single idea. Therefore, we created a corpus of individual sentences. The sentence selection and extraction process consists of the following two steps:

- **Splitting**: split the text collection into sentences.
- **Selection**: extract the sentences that fulfill the meter requirements of the *Zortziko Txikia* (13 syllables).

Once preprocessing step is finished, we have a corpus of 56274 sentences, with 20098 unique terms in it. This sentence repository will be used in the *bertso* generation process.

4.2 Semantic Relationship/Similarity Measures

The four rhymes received as input data are used to find sentences in the corpus ending with those rhymes. During the search process, a set of sentences for each initial rhyme are found. Those sentences will then be fed into the similarity measuring stage to perform the sentence matching process.

A system capable of validating its output from a metrical and rhythmical point of view shows a hint of aesthetic sensibility, but it is missing out one important aspect that humans use to value poetry: meaningfulness. The main purpose of this stage is to compare the relationship between pairs of sentences to choose the most semantically related ones for the final verse. To do that, the sentence relation system computes, for each pair of sentences, a score that evaluates how those sentence are semantically related. That is, once the technical requirements are fulfilled, sentences that maximize the internal coherence of the generated verse are selected.

The starting point for our similarity methods is the observation that related strophes in a *bertso* tend to use the same or similar words. It is true that the use of the same or similar terms does not always guarantee relatedness, but can be considered as a precondition, and we believe it should work well as a measure of relatedness. We have tested the two similarity measures described below.

Method 1: Standard Vector Space Model (VSM)

The Vector Space Model (VSM), introduced first by Salton [Salton 1971], is a model for representing text in a vector space. Each dimension of a vector corresponds to a term that is present in the text collection. This is the simplest vector-based similarity approach, and similarity calculation is based on word matching. Before any semantic processing takes place, we applied the following text preprocessing tasks:

- **Tokenization**: sentences are split into individual words. Punctuation marks are also removed from text.
- **Lemmatization**: due to its agglutinative characteristic, a given word lemma makes many different word forms in Basque language. Therefore, a lemmatization step is accomplished to reduce term quantity.

In order to represent the importance of a term within a document and within the whole document collection, the well known *tf-idf* (term frequency-inverse term frecuency) [Ramos 2003] weighting scheme has been applied, which combines local and global weighting. The *tf-idf* weighting scheme assigns to term *t* a score in sentence *s* given by

$$tf - idf_{t,s} = tf_{t,s} \cdot idf_{t} \tag{1}$$

Finally, to measure the semantic relation between a pair of sentences, we make use of the classical cosine similarity of their vector representations:

$$cos\theta = (V_1 \cdot V_2)/(||V_1|| \cdot ||V_2||) \tag{2}$$

where the numerator represents the dot product (or inner product) of the vectors, and the denominator is the product of their Euclidean lengths.

Shortcomings of this standard vector-based method are known: on the one hand, only sentences that overlap in vocabulary will be considered similar. Two documents with a similar topic but different vocabulary will not be considered similar. In other words, it cannot deal with synonymy and polysemy. And on the other hand, it assumes complete independence among the terms.

Method 2: Latent Semantic Analysis (LSA)

LSA was developed as a special vector space approach to solve the aforementioned issues [Deerwester et al. 1990]. In the standard VSM, term associations are ignored, and therefore, synonymy and polysemy are not captured. By contrast, LSA tries to find the latent semantic structure of a document collection. That is, to find hidden relations between terms, sentences or other text units [Landauer and Dumais 1997]. In other words, LSA analyzes term co-occurrence of higher orders, so that it is able to incorporate the relationship of words A, B which only co-occur in documents through word C, and never appear in the same document together directly.

LSA involves the application of the matrix algebra method Singular Value Decomposition (SVD) to the document-by-term matrix in order to reduce its rank and construct a semantic space. The matrix is then decomposed in such a way that similar vectors will be conflated, or moved closer within the space. And thus, LSA enables the calculation of semantic similarities between words, sentences or paragraphs.

To construct a semantic space to capture textual coherence, a large representative text corpus is needed. Thus, a large collection of documents collected from the newspaper *Berria* was used for the construction of the semantic space. The text collection was divided into paragraphs, that lead to a corpus of 161.113 documents and 10.121.624 words. Again, the text collection was first lemmatized in order to reduce term quantity with no loss in meaning. Thus, a corpus of 180.616 unique words was obtained. Then, the entries of the document-by-term matrix were weighted using the already presented *tf-idf* scheme. Sentence similarity in the reduced space is measured by the cosine distance between their vector representations.

We used Gensim [Řehůřek and Petr 2010] Python framework to compute the LSA and perform similarity queries. After applying the SVD, the semantic space dimension was reduced to 250, a value chosen empirically. We thus map each term/document to a 250 dimensional space. The consequence of this dimension reduction is that some dimensions are combined on more tha one term.

Alg. 1 summarizes the verse-generation process. The algorithm receives as input the four rhymes to compose the verse, extracts rhyming sentences from the corpus and generates all possible verse-combinations. For each generated verse a cosine similarity of adjacent sentences is calculated and the verse with the highest score is returned for each of the above mentioned methods.

It must be noted that for each *s1-s2-s3-s4* sentence combination, *s1-s2, s2-s3* and *s3-s4* similarity comparisons are made. That is, the overall semantic measure is calculated for each *bertso* by adding up the cosines between the vectors for all pair of adjoining sentences. This is so because we seek a progressive relationship between sentences. And for the overall coherence of the *bertso*, it is more important the relationship between *s1* and *s2* (adjacent strophes), than the relationship between *s1* and *s4* (non-contiguous strophes). Finally, the system returns the verse with the highest cosine measure.

It must be noted that for each *bertso* three comparisons are made. For example, if for each (r1-r4) rhyme group we have **s**=10 sentences in the corpus, then the cosine similarity between two vectors will be calculated *10x10x10x10x3 =30000* times. As **s** grows, computing the cosine similarities can become computationally expensive.

```
Input: r1 , r2 , r3 and r4 rhymes
Output: the verse with the highest cosine measure

group1 = sentences that have r1 as rhyme;
group2 = sentences that have r2 as rhyme;
group3 = sentences that have r3 as rhyme;
group4 = sentences that have r4 as rhyme;

for all s1 in group1 do
    for all s2 in group2 do
        for all s3 in group3 do
            for all s4 in group4 do
                verse = (s1, s2, s3, s4 );
                cos_12 = cosine_similarity(s1, s2);
                cos_23 = cosine_similarity(s2, s3);
                cos_34 = cosine_similarity(s3, s4);
                cos_all = cos_12 + cos_23 + cos_34
            ;
                store(file, verse, cos_all);
            end for
        end for
    end for
end for
return find_max_cosine(file)
```

Alg. 1 General algorithm to generate verses

It must be noted that for each *s1-s2-s3-s4* sentence combination, *s1-s2, s2-s3* and *s3-s4* similarity comparisons are made. That is, the overall semantic measure is calculated for each *bertso* by adding up the cosines between the vectors for all pair of adjoining sentences. This is so because we seek a progressive relationship between sentences. And for the overall coherence of the *bertso*, it is more important the relationship between *s1* and *s2* (adjacent strophes), than the relationship between *s1* and *s4* (non-contiguous strophes). Finally, the system returns the verse with the highest cosine measure.

It must be noted that for each *bertso* three comparisons are made. For example, if for each (r1-r4) rhyme group we have **s**=10 sentences in the corpus, then the cosine similarity between two vectors will be calculated *10x10x10x10x3 =30000* times. As **s** grows, computing the cosine similarities can become computationally expensive.

5 Evaluation

The evaluation of computer-generated poetry is a difficult task. Taking into account that our goal is to build a robotic system that performs verses in public, like real *bertsolari*-s do, we believed that empirical validation by human subjects was the best way of evaluation. Thereby, we contacted with the *Bertsozale Elkarte* [WWW_3] (Association of the Friends of *Bertsolaritza*) and 5 *bertso*-judges participated in the evaluation.

We selected the following 4 different rhyming groups (rhymes used in a verse-campionship), where each group was composed by four rhyming words to generate the bertso:

Hor-inor-zor-gogor	There-nobody-debt-hard
gera-batera-atera-bera	we're-together-go out-he
dakar-zakar-azkar-alkar	bring-rough-fast-together
dira-tira-erdira-begira	they're-come on-center-look

For each group, 6 computer-generated verses were selected:

- 2 generated with no similarity measure. Sentences were selected randomly, without applying any cohesion measure between them
- 2 generated following VSM and cosine similarity approach
- 2 generated following LSA approach.

We mixed all the verses and a a blind experiment was carried out. Verse-judges were asked to evaluate them based on the following scoring:

0→ The verse has no meaning. There is no cohesion between the strophes.
1→ Minimal cohesion.
2→ Average cohesion (some strophes maintain meaninful relation).
3→ Good cohesion level.
4→ Really well related strophes. Very well overall cohesion.

Table 1 shows how many verses have achieved 0 scoring, how many 1 and so on, for each method used.

Table 1 Score frequencies and percentages

Scoring	Random	VSM	LSA
0	14 (35%)	7 (17.5%)	4 (10%)
1	12 (30%)	8 (20%)	8 (20%)
2	12 (30%)	13 (32.5%)	14 (35%)
3	2 (5%)	8 (20%)	6 (15%)
4	0 (0%)	4 (10%)	8 (20%)

Table 2 shows the average scoring obtained by each method.

Table 2 Average scores

	Rhyming gr. 1		Rhyming gr. 2		Rhyming gr. 3		Rhyming gr.4	
	Verse1	Verse2	Verse1	Verse2	Verse1	Verse2	Verse1	Verse2
Random	1,2	1	0,4	1,6	1	1	1,2	1
VSM	2,2	2,8	2,2	2,2	0,2	1,8	1,2	2,2
LSA	2	3	1,2	1	3	3,2	2	1,8

Looking to the above tables, it is notorious that there is a clear distinction between the verses generated randomly, and the verses generated applying any of the similarity measures presented. Whereas lower scores are the most frequent scores given to the randomly generated verses, highest scores are reserved to VSM and LSA verses. It is also worth noting that although VSM shows the best mean values, LSA gets highest scores.

For illustration purposes, we show a *bertso* generated by our system using LSA method. We also give an English translation, even though part of its aesthetic value is lost in translation.

Gu euskaldunak berez	*We, Basque people*
ez gerala inor	*are nothing at all*
guk inori ez diou	*we do not owe our live*
gure bizitza zor	*to anyone*
berak gogor badaude	*if they are hard*
gu ere hain gogor	*we also*
eta euren jarrera	*and that will show*
igarriko da hor	*their stance*

The objective of the evaluation was none other than finding the most suitable method for measuring the internal consistency of the verses. Once the method is implemented in the real robot, the improvised verses will be judged by the public with applause or whistles, as in the events of improvisers.

6 Integration of the Components into a Real Robot

So far we have described the tools used for creating verses and the methods to predict their textual coherence. But *bertsolaritza* belongs to the oral genre, and thus, needs to be performed in public. Therefore, the *Bertsobot* system needs additional functionalities, such as a robotic body to interact with the environment and display emotional body-language, and a synthesized voice.

In this last step we have integrated the automatic verse-making system in a pair of real autonomous robots.

- ***Galtxagorri*** is a Pioneer 3-DX robot
- ***Tartalo*** is a PeopleBot robot built on the P3-DX base.

Fig. 4 shows a picture of the robots in a public performance.

Fig. 4 Verse-duel between robots and real *bertsolari*-s

The *Bertsobot* system can be run in a simple computer, but as mentioned before, *bertsolaritza* implies interaction with the environment, such as the other *bertsolaris* of the performance and the audience. In improvised oral festivals, the *bertsolari*-s wait sitting their turn to sing, and when their turn comes, (s)he approaches the microphone that is placed in front. Once in their location, they have to look to the theme-prompter, that will prescribe a topic which serves as a prompt for the *bertso* (in our case the four rhymes to compound the *bertso*). After that, they have a specific time to create the verse and finally sing it. Once finished, they await a few seconds to perceive the audience's reaction and they go back to their chair.

The aim of the robotic behaviors is to represent as faithfully as possible *bertsolari*'s movements on stage, interacting as much as possible with the other improvisers and the audience.

Videos of the public demonstrations carried out with real improvisers can be found on our website [WWW_4].

7 Conclusions and Future Work

We have presented an embodied system, Bertsobot, capable of improvising and performing Basque traditional verses. The main contribution of this paper consists on the verse-maker module, which composes verses automatically taking into account the semantic relation of strophes within a *bertso*.

We are already working on further enhancements to each of the processes involved, including:

1. Implementing improved methods to generate phrases for templates and working with other corpora.
2. Exploring new models to measure the semantics relationship between sentences.
3. Letting the robots display facial expressions by means of a 3D avatar.
4. Enabling the robot to adequate the message of the poem based on the reactions perceived from the public.
5. Improving the emotional body language of the robots and increasing their degree of autonomy while acting the performance. For the time being, we are trying to identify the microphone using a Kinect, so the robot will know where it is.
6. Identify the faces of the participants (*bertsolari*-s and the theme-prompter). Thus, the robot will turn exactly towards its competitor, or mentioning its name in the improvised verse, as it occurs sometimes.

Acknowledgments. This work has been partially conducted under the SAIOTEK projects XIMUR and POETAUTO, the Basque Government Research Team grant and the University of the Basque Country UPV/EHU, under grant UFI11/45 (BAILab).

References

[Alegria et al. 1996] Alegria, I., Artola, X., Sarasola, K., Urkia, M.: Automatic morphological analysis of Basque. Literary and Linguistic Computing 11(4), 193–203 (1996)

[Astigarraga et al. 2013] Astigarraga, A., Agirrezabal, M., Lazkano, E., Jauregi, E., Sierra, B.: Bertsobot: the first minstrel robot. In: Proc. 6th Inter. Conf. on Human System Interaction, pp. 129–136 (2013)

[Beck et al. 2010] Beck, A., Hiolle, A., Mazel, A., Cañanamero, L.: Interpretation of emotional body language displayed by robots. In: Proc. 3rd International Workshop on Affective Interaction in Natural Environments, pp. 37–42 (2010)

[Colton et al. 2012] Colton, S., Goodwin, J., Veale, T.: Full face poetry generation. In: Proc. 3rd Inter. Conf. on Computational Creativity, pp. 95–102 (2012)

[Deerwester et al. 1990] Deerwester, S.C., Dumais, S.T., Landauer, T.K., Furnas, G.W., Harshman, R.A.: Indexing by latent semantic analysis. J. of the American Society for Information Science 41(6), 391–407 (1990)

[Gervas 2013] Gervás, P.: Computational modelling of poetry generation. AISB Convention 2013. University of Exeter (2013)

[WWW_1] `http://www.berria.info/` (accessed January 5, 2014)

[WWW_2] `http://bdb.bertsozale.com` (accessed January 5, 2014)

[WWW_3] `http://www.bertsozale.com/en` (accessed January 5, 2014)

[WWW_4] `http://www.sc.ehu.es/ccwrobot/bideoak`

[Landauer and Dumais 1997] Landauer, T.K., Dumais, S.: A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological Review 104(2), 211 (1997)

[Landauer et al. 2013] Landauer, T.K., McNamara, D.S., Dennis, S., Kintsch, W.: Handbook of latent semantic analysis. Psychology Press (2013)

[Oliveira 2009] Oliveira, H.: Automatic generation of poetry: an overview. Universidade de Coimbra (2009)

[Ramos 2003] Ramos, J.: Using tf-idf to determine word relevance in document queries. In: Proc. 1st Instructional Conf. on Machine Learning (2003)

[Řehůřek and Petr 2010] Řehůřek, R., Petr, S.: Software framework for topic modelling with large corpora. In: Proc. of LREC 2010 Workshop New Challenges for NLP Frameworks, pp. 45–50 (2010)

[Salton 1971] Salton, G.: The SMART retrieval system: Experiments in automatic document processing. Prentice-Hall (1971)

[Scheutz et al. 2011] Scheutz, M., Cantrell, R., Schermerhorn, P.: Toward humanlike task-based dialogue processing for human robot interaction. AI Magazine 32(4), 77–84 (2011)

[Wong et al. 2008] Wong, M.T., Chun, Li, Q., Chen, S.Y., Xu, A.: Automatic haiku generation using vsm. In: WSEAS International Conference. Mathematics and Computers in Science and Engineering, no. 7. World Scientific and Engineering Academy and Society (2008)

Usability Evaluation of Biometrics in Mobile Environments

R. Blanco-Gonzalo, L. Diaz-Fernandez,
O. Miguel-Hurtado, and R. Sanchez-Reillo

University Group for Identification Technologies,
Carlos III University of Madrid, Leganés, Spain
rbgonzal@ing.uc3m.es, ldifernan@gmail.com,
{omiguel,rsreillo}@ing.uc3m.es

Abstract. Actual trends in HCI deal with moving systems to mobile environments. Moreover, biometrics is a technology that is entering maturity, getting involved in several security architectures nowadays. Thus, migrating biometrics to mobile scenarios is trending topic in the research community. Nevertheless, in this kind of systems, usability has been put aside unintentionally in the intent of produce better performance and it could involve undesirable results. To fill this gap a behavioural biometric modality (handwritten signature verification) is evaluated in this paper in mobile environments, in order to obtain a complete usability report. Users had to sign on an iPad with different styluses in different scenarios, relating performance results with several usability parameters (gathered through video, notes and forms) and obtaining interesting outcomes.

1 Introduction

Biometric recognition is the univocal differentiation of individuals according to specific physical or behavioural attributes such as iris, fingerprint or voice. Biometric systems are common in places where high security is needed and are included many times as reinforcement to other security techniques (smart cards, PIN codes, passwords, etc.). Usually, biometric systems are developed trying to reach the best throughputs but many times, users' satisfaction is forgotten. Being the final users who interact with the biometric products, dissatisfactions can involve misuses and this can lead to worse results or rejection to use the technology. There are studies in biometrics usability like [Kukula et al. 2007; Theofanos et al. 2008] but there are still additional biometric modalities and usability factors that must be covered.

As long as technology is moving to mobile scenarios, biometrics should be adapted to them assuring an acceptable behaviour without compromising performance. Therefore, several challenges have to be overcome, such as adapting devices, algorithms and systems architectures. Adapting biometrics to mobile environments is stirring researchers' interest nowadays [Mansfield-Devine 2011] and there are several researches in this line [de Santos Sierra et al. 2011; McLoughlin and Naidu 2009]. This work is focused on better adapt handwritten signature verification to

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_24, © Springer International Publishing Switzerland 2014

mobile environments contributing with a comprehensive usability analysis. The signatures were collected with a capacitive tablet (Apple iPad) using different styluses (specially made for signing and drawing over mobile devices). The iPad was chosen due to its popularity and due to be one of the most preferred devices by users [Blanco-Gonzalo et al. 2012]. In addition, different scenarios where users had to sign in different postures (the most common when using mobile devices) were analyzed.

The algorithm applied is a DTW (Dynamic Time Warping)-based [Pascual-Gaspar et al. 2009] modified for mobile devices, which is one of the most commonly used in research works. The main intention of this work is to obtain usability conclusions in order to improve future biometric developments. The measurements were extracted from ISO 9241-11 [ISO 1998], where usability is defined as "*The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use*". Performance results were extracted once the evaluation was completely finished and all the signatures had been collected in the database. All signatures captured were real signatures, not having captured any kind of forgeries in the database. Therefore only uniqueness was analyzed, leaving forgery robustness for future works. This paper is structured in 4 sections. After this introduction, section 2 describes the evaluation set-up in detail. The experiments done are explained in section 3. Results, conclusions and future work are shown in section 4.

2 Evaluation Set-up

The evaluation was split into 3 sessions with at least 1 week of separation between each one in order to avoid tiredness: users will not be eager to complete the evaluation in three days in a row and signatures would be done apathetically.

2.1 Evaluation Crew

There were 21 participants in the experiment, 13 men and 8 women. They were chosen randomly but representing all the main different age groups. Almost half of them have used electronic digitizing tablets to sign before but for all of them it was the first time they signed in a mobile device. There was not any special condition for joining the evaluation except being older than 16 years old (signature is not stable in children [Kekre and Bharadi 2009]). An explanation of the evaluation was made before starting and users filled a satisfaction form at the end of the third session. This form included questions about:

- Preferences of scenarios and devices.
- Comfort, time-spent, easiness, intrusion and global satisfaction.
- Familiarity with biometrics and previous experiences.
- Diseases that could modify the signature (i.e. Parkinson or sclerosis).
- Age, gender, laterality and level of studies.
- The received instructions during the evaluation
- Possibility of future uses.

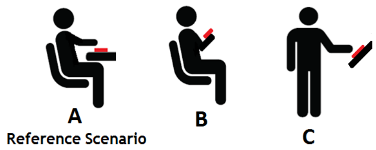


Fig. 1 Evaluation scenarios

2.2 Scenarios

Three different scenarios were completed by users in the evaluation as it is shown in Fig. 1. These scenarios were chosen in order to represent the most common situations where users may use this technology. The scenario A, which is taken as the reference scenario, consists of the user signing sat on a chair with the device resting on a table. In scenario B the user is sat on a chair handling the device with the only support of her hands. The user herself decides how to grab the mobile device. In scenario C the user signs standing up and with the device resting over a lectern. The height of the scenarios' elements was up to users' preference and they could modify it whenever they wanted in order to feel comfortable. The scenarios order was changed from one session to another, being in the 1st session A-B-C, in the 2nd session C-B-A and in the 3rd session B-C-A. This was done to minimize the effect of order in performance results as well as in users' habituation.

2.3 iPad and Styluses

The Apple iPad has a 9.7-inch (diagonal) capacitive and 1024x768 px at 132 ppi LED screen. Its processor is an A4 (at 1GHz) and it has 256 MB of RAM [WWW-1 2013]. Three styluses were used in the evaluation in all the scenarios. The differences between them are their tips' shape, length and diameter. In this work a colour code is used to identify them (black, white and pink). An illustration of the styluses can be seen in Fig. 2. The pink stylus is made of plastic, its length is 10.7 cm and its diameter is 6 mm. Its' tip is made of sponge and ends in a plane shape. The white stylus is made of plastic also. It has 11 cm of length and 8 mm of diameter. Its tip is composed by soft rubber and has rounded form. Both styluses (pink and white) have a deformable tip. Finally, the black stylus is made of steel. It has 12.6 cm of length and 8 mm of diameter. Its tip is made of hard plastic and steel, so it is not deformable but its position changes when it contacts with the device and therefore it is always plane to the touched surface. The 3 styluses are conductive and made to write over capacitive devices. The order in using the different styluses was changed in each session, being the order in the 1st session pink-white-black, in the 2nd session black-pink-white and in the 3rd session white-pink-black.

Fig. 2 Styluses utilized in the evaluation (Black, White and Pink from left to right)

2.4 Guidance and Training

There was not any training process scheduled before the evaluation and users started the experiment without previous knowledge. It was considered an important experiment to test how the users' signature changed through sessions and time (habituation) although users were previously told about the evaluation structure and proceedings. Also, users had the possibility to delete the signature made as many times as desired; so this can be considered as a self-training process. All the deleted signatures are accounted for obtaining the efficiency rate as it is explained in section 3. Furthermore, an operator responsible for the evaluation offered support to users if needed, though no instruction about how to sign was given. When a user asked for advice about how to sign, the operator responded: "the signature should be as similar as possible to the one made in a paper with a regular ink pen". As a behavioural biometric modality, in handwritten signature verification, user mistakes are not easy to control due to the fact that there is not an incorrect way of signing (excluding signing out of bounds or repeating strokes, for example). In other modalities, such as fingerprint or iris recognition, it is feasible to instruct the user on how to present the finger or the eye correctly to the system.

2.5 Algorithm

The objective of the algorithm is to decide whether the user is the one who claims to be or not. The algorithm used for obtaining the performance rates is a DTW based online signature verification. This allows an optimal alignment between two sequences of vectors of different lengths using dynamic programming. From this alignment a measure of distance between two temporal patterns is obtained. The signals used as input from signatures are the X and Y time-series coordinates. Although it is known that pressure information improves signature algorithm's performance, this parameter is not used as it is not available in the iPad. Error rates, reported using the same algorithm, in previous experiments with non portable devices are EER = 1.8% for random forgeries and EER = 7.6% for skilled forgeries [Miguel-Hurtado 2011].

3 Experiments

It is important to differentiate the two parts of the evaluation: data collection and its posterior processing. Thus, all the results gathered are calculated in the processing step. In this section all the usability and performance experiments are shown.

3.1 Usability

The tests done are according to the definition of usability from ISO 9241-11 cited before, where effectiveness, efficiency, and satisfaction parameters are defined as follows:

- *Effectiveness*: Is defined as the task completion by users. Applying this definition to the evaluation, the effectiveness error is the ratio of wrong signatures accepted as correct divided by the total amount of them. If users are not able to complete the whole process they feel frustrated and this influences directly in the satisfaction and has a bad impact over the signature quality. Accordingly, when users deliver low quality signatures (bad task completion) it has a direct effect on the algorithm performance. A signature is considered as wrong in the following cases:
    - The signature is totally or partially out of the devoted space for signing.
    - User has to repeat at least one stroke (e.g. due to latency problems).
    - Part of the signature is made without touching the device (signing in the air).
    - The accept button is touched being the signature incomplete or not desired.
    - User' wrist touches the screen.
- *Efficiency*: Is the completion of tasks on time. The efficiency error rate is calculated dividing the number of deleted signatures by the total amount of them. Deleting too many signatures makes the user feel frustrated and involves dissatisfaction and tiredness. At the same time, the average time to complete the scenarios was calculated.
- *Satisfaction*: Is the user experience in the evaluation. It was measured through the satisfaction forms, the operator notes and the video taken. Furthermore, both the efficiency and the effectiveness influence this parameter as above mentioned.

Apart from these three factors, *learnability* was measured due to users' habituation to signal acquisition through the sessions. Thus, time spent and errors made are expected to be decreasing from one session to another as the user acquires skills signing in these scenarios.

3.2 Performance

Performance error rates were calculated for each scenario and stylus, in order to compare them with users' preferences. Furthermore, using these error rates it is possible to measure habituation effects and the influence of scenarios and styluses order in users. In order to obtain performance results a previous template to compare with (enrolment) is necessary. Once the enrolment is done it is feasible to carry out the verification process.

The error rates obtained from the algorithm were the FAR (False Acceptance Rate), the FRR (False Rejection Rate) and the EER (Equal Error Rate) as three of the most common error rates used in biometrics [ISO 2006]. FAR expresses the ratio of impostors who were accepted by the system and FRR is the ratio of genuines who were rejected. EER is the rate at which FAR and FRR are equal. Finally, results are showed in terms of EER in this paper as a convenient way to determine the algorithm accuracy and compare among results. In this evaluation comparisons are offline (done after all the acquisitions have been finished) and 3 signatures gathered from each user were taken for obtaining the template for each one. Thus, FRR is calculated comparing each template with the rest of the users' genuine signatures from the same user-device-scenario combination. FAR is obtained by comparing each template with all the signatures from the same combination device-scenario containing all the other users. Finally, by shifting the decision threshold a ROC (Receiver Operating Characteristic) curve is obtained representing FAR against 1-FRR (genuine attempts accepted).

4 Results, Conclusions and Future Work

It is noticeable that a great amount of data and results were extracted from this evaluation and due to space constraints it is not possible to disclose them altogether. Anyway, in this section the most relevant results and conclusions are summarized.

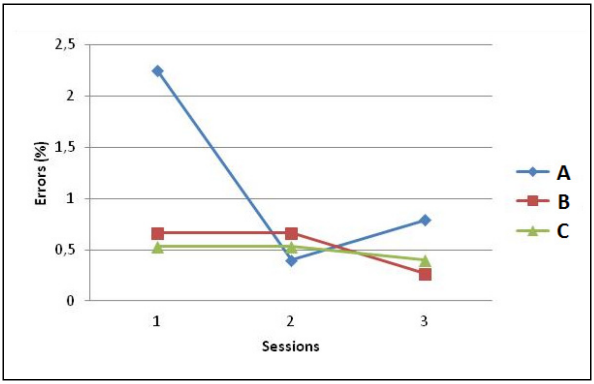


Fig. 3 Effectiveness rates

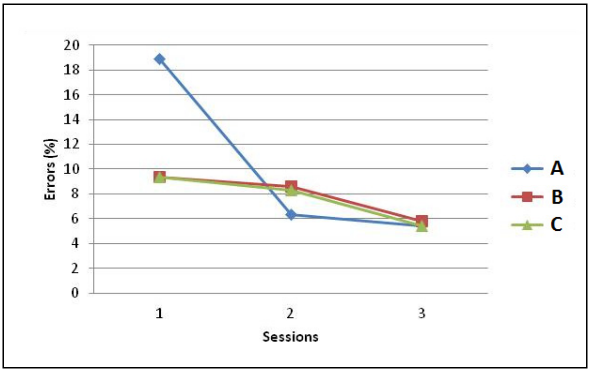


Fig. 4 Efficiency rates

4.1 Usability

In Fig. 3 effectiveness rates are shown divided in sessions and scenarios. In Fig. 4, the efficiency results can be found divided in sessions and scenarios also and in Fig. 5 time-spent averages are provided divided in styluses-sessions. The satisfaction factors obtained from forms (comfort, time spent, easiness, intrusion and global experience) were scored from 0 to 5 (as it can be seen in Fig. 5). The global experience score is 4.28 and the average of all factors is 3.87.

The effectiveness and the efficiency rates show a big decrease of the errors from 1st session to 3rd session, especially in the scenario A. It occurs due to the scenario A is the first to be completed and users had no previous experience signing in mobile devices. By the end of the evaluation both rates become stabilized. This error variation could be solved by completing a previous training process. The time average in the scenarios shows a high dependence of the order. Thus, in 1st session (A-B-C) and in 2nd session (C-B-A) the first scenarios took more time. In 3rd session the time becomes almost equal for the three scenarios, showing that users have not problems to sign in any scenario in particular.

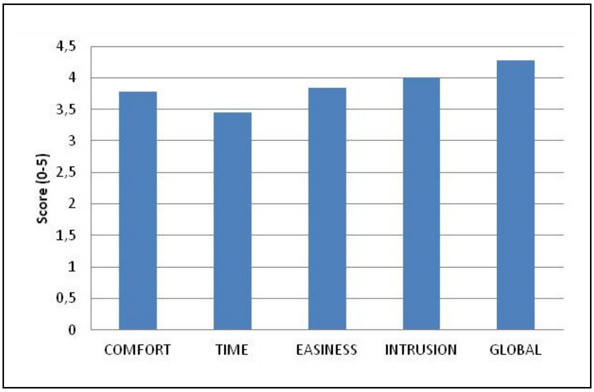


Fig. 5 Satisfaction factors

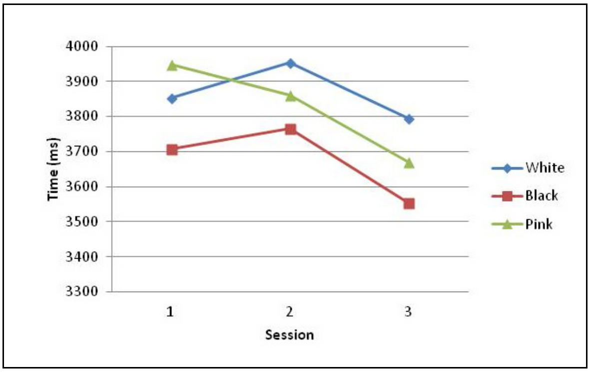


Fig. 6 Time spent with the styluses

Regarding the styluses, the black is the faster for users to complete the evaluation followed by the pink and the white. Users considered the evaluation globally good and non intrusive, in fact, 14 of them would use this kind of system habitually. The worst satisfaction parameter was the time spent. Nevertheless in a real scenario this should not be a problem because the number of signatures required would be much lower. The order of preferred scenarios is A>B>C and black>pink>white for styluses. Users felt comfortable singing in the iPad and considered that the device response was good enough. For example, one of the users said "*the visual feedback is not very different to the one perceived with pen and paper*". Regarding the learnability, users acquired part of the skills in the 2nd session and the rest in the 3rd session: according to effectiveness and efficiency, users acquired habituation in the 2nd session (in Fig. 3 and Fig. 4 these rates become stable from the 2nd session). Though, according to the time spent (Fig. 6) it is descendant until the end of the evaluation, so that, users are still gaining ability signing. Analyzing these results it is feasible to conclude that participants get used to the system in the 2nd session, once efficiency and effectiveness become steady (they suit better the space to sign and commit fewer errors). The time spent, always decreasing, shows that users do not stop improving their habituation to the signing process.

4.2 Performance Results

In Fig. 7 a comparison between the EERs obtained from the 3 styluses (within the legend box) is done along with the ROC curves, taking all the signatures gathered (from all the scenarios). In Table 1 the EERs obtained with the different styluses in the different scenarios are revealed. Previous results obtained signing with the fingertip are included also. It is noticeable the EER=3.96% of the pink stylus in the scenario A.

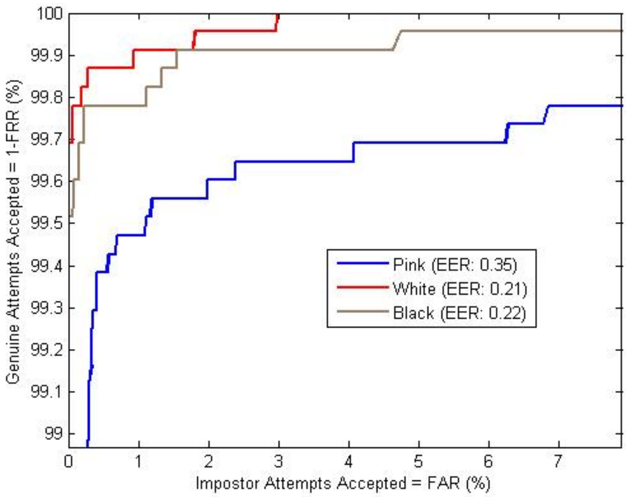


Fig. 7 ROC curve of styluses performance

This high error rate is due to be the first stylus-scenario combination in the evaluation. As in the usability results, the lack of previous training is remarkable and the order seems to be highly important also. For the 3 devices the best performances are obtained in scenario C which is the least preferred by users. This result can be due to users had not to handle the iPad with the hands (it was placed in a lectern).The fact of having the device resting in a surface avoids bad effects of pressure changes produced when is the user who handles the device. This shows also that a small tilt of the surface makes the signing process easier (results are better in scenario C than in scenario A). In order to test which stylus offers better performance, the templates for calculating the ROC curve were chosen randomly (minimizing the effect of order). In Fig. 7 it is shown that the EER of the white stylus is the best (0.21%), followed closely by the black stylus (0.22%) being the pink the worst one (0.35%). This is almost in accordance with users preferences. In addition these results improve previous results obtained (EER= 1.8%). This improvement can be due to improvements in the capture software or iPad features such as screen quality or processor (different from digitizing tablets' features used to design the algorithm).

Table 1 Styluses and fingertip EER divided by scenarios

Scenarios	Pink	White	Black	Fingertip
A	3.96%	0.26%	0.78%	7.27%
B	0.52%	0.93%	0.52%	0.49%
C	0.26%	0.13%	0.13%	-

Table 2 Usability comparison between stylus and fingertip

	Effectiveness errors	Efficiency errors	Satisfaction
Styluses	0.72%	8.62%	4.3/5
Fingertip	1.49%	5.63%	3.9/5

4.3 Fingertip versus Stylus

One interesting point in this research is the comparison between signing with the fingertip and signing with a stylus. Results regarding signing on an iPad with the fingertip are extracted from previous works [Blanco-Gonzalo et al. 2013]. The performance and usability results are given in Table 1 and Table 2 (the iPad was not used with the fingertip in the scenario C). Regarding the performance it is noticeable that signing with the fingertip is less straightforward for users than with the stylus (7.27% fingertip – 3.96% stylus) and the signatures have different appearance, in particular at the beginning (Fig. 8). This is due to the lack of habituation of users (none of them had ever signed with the fingertip). Once the user is habituated to the system both results become similar (2nd session) involving similar performance in both approaches. The number of effectiveness errors is bigger with the fingertip due, in most of the cases, to the different position in which users sign with the finger (the gesture is very different for some of them). It is remarkable that the efficiency errors are less with the fingertip due to the different styluses used: when changing from one stylus to another the provided signatures look different involving users to delete them more than usual. Finally, satisfaction rates (4.3/5 stylus – 3.9/5 fingertip) and users' opinions reveal the general preference for stylus based systems as expected. In short, both sub-modalities (fingertip and stylus) offer similar results in performance and styluses are better rated in usability. Nevertheless, as long as handling mobile devices is more common with the finger only, getting used to sign with the fingertip requires only some previous training.

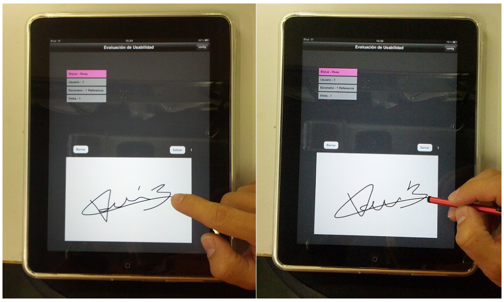


Fig. 8 Signing with the fingertip and with a stylus

4.4 Future Work

In addition to the study itself, one of the intentions of this research is to establish guidelines for future usability evaluations. Then, to apply the conclusions obtained in this work to future experiments is planned for the short term. Integrating the black stylus with the iPad as a final product seems to be interesting also as good results were achieved and users felt comfortable with this combination. Test forgery robustness of this system is an experiment that will be made also. Using other approaches to obtain usability outcomes (further research over the HCI field is required) could add new parameters to our analysis and can help to design a complete model to measure usability.

References

[Kukula et al. 2007] Kukula, E.P., Elliott, S.J., Duffy, V.G.: The effects of human interaction on biometric system performance. In: Duffy, V.G. (ed.) Digital Human Modelling, HCII 2007. LNCS, vol. 4561, pp. 904–914. Springer, Heidelberg (2007)

[Theofanos et al. 2008] Theofanos, M., Stanton, B., Wolfson, C.A.: Usability and biometrics -ensuring successful biometric systems. Handbook of National Institute of Standards and Technology (NIST), Gaithersburg, US (2008)

[Mansfield-Devine 2011] Mansfield-Devine, S.: Biometrics for mobile devices struggle to go mainstream. Biometric Technology Today 2011(9), 10–11 (2011)

[de Santos Sierra et al. 2011] de Santos Sierra, A., Sanchez Avila, C., del Pozo, G.B., Guerra Casanova, J.: Gaussian multiscale aggregation oriented to hand biometric segmentation in mobile devices. In: Proc. 3rd World Congress on Nature and Biologically Inspired Computing, pp. 237–242 (2011)

[McLoughlin and Naidu 2009] McLoughlin, I.V., Naidu, N.: Keypress biometrics for user validation in mobile consumer devices. In: Proc. IEEE 13th International Symposium on Consumer Electronics, pp. 280–284 (2009)

[Pascual-Gaspar et al. 2009] Pascual-Gaspar, J.M., Cardeñoso-Payo, V., Vivaracho-Pascual, C.E.: Practical on-line signature verification. In: Tistarelli, M., Nixon, M.S. (eds.) ICB 2009. LNCS, vol. 5558, pp. 1180–1189. Springer, Heidelberg (2009)

[Blanco-Gonzalo et al 2012] Blanco-Gonzalo, R., Miguel-Hurtado, O., Mendaza-Ormaza, A., Sanchez-Reillo, R.: Handwritten signature recognition in mobile scenarios: Performance evaluation. In: Proc. 2012 IEEE Inter. Carnahan Conf. on Security Technology, pp. 174–179 (2012)

[ISO 1998] ISO 9241:1998. Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability

[Kekre and Bharadi 2009] Kekre, H.B., Bharadi, V.A.: Ageing adaptation for multimodal biometrics using adaptive feature set update algorithm. In: Proc. of IEEE Advance Computing Conference, pp. 535–540 (2009)

[WWW-1 2013] iPad technical specifications, `http://support.apple.com/kb/SP580` (accessed August 29, 2013)

[WWW-2 2013] IPcam specifications, `http://www.axis.com/products/m10_series` (accessed August 29, 2013)

[Miguel-Hurtado 2011] Miguel-Hurtado, O.: Online Signature Verification Algorithms and Development of Signature International Standards. PhD Thesis at University Carlos III of Madrid (2011)

[ISO 2006] ISO/IEC 19795-1:2006, Information Technology - Biometric Performance Testing and Reporting - Part 1: Principles and framework

[Blanco-Gonzalo et al. 2013] Blanco-Gonzalo, R., Miguel-Hurtado, O., Sanchez-Reillo, R., Gonzalez-Ramirez, A.: Usability Analysis of a Handwritten Signature Recognition System Applied to Mobile Scenarios. In: Proc. IEEE Inter. Carnahan Confe. on Security Technology, ICCST (2013) (in printing)

Posture Independent Model for Hand Detection and Tracking

M. Alsoos and A. Joukhadar

Department of Artificial Intelligence, Faculty of Information Technology Engineering,
Damascus University, Damascus, Syria
msoossoos@gmail.com, ajoukhadar@el-ixir.com

Abstract. Human Computer Interaction is an active research field, many researches all over the world try to invent more natural and intuitive communication ways between humans and computers. Hand interaction methods occupy distinguished position among other interaction methods because of hand intensive use in human everyday life. In this paper we present a new appearance based method to recognize and track moving bare human hand in an unknown environment. Our method can distinguish between hand and other moving objects, by using proposed posture independent hand features. Results show that the presented method can efficiently and effectively recognize and track all human hand postures in real-time.

1 Introduction

Computers nowadays are the most indispensable machines in all over the world, we find them almost in all fields because of their efficiency, reliability, low price and fast processing speed. This wide spread leads to a real revolution in hardware and software manufacturing toward making computers more useful and powerful. Although we have been used computers for many decades, we still interact with these machines using ancient ways like keys, touch screens or remote controls.

Some researchers started to fill the interaction gap between human ways and computer methods by trying to make computers understand human intuitive interaction ways. This will make a revolution in human computer interaction by giving humans the ability to communicate with computers in the same way they are communicating with each others. Speech recognition and text-to-speech methods were used in automatic reservation machines to understand human oral requests and reply them orally. Speech recognition helped also in making more intelligent robots which can understand and execute their owners orders [Kohda et al. 1976]. Face and emotion recognition were used to make computers interact with some emotional situations. Many researches focused on disabled people and developed customized methods for them like eye tracking, eye blink recognition or Brain-actuated interaction which can change life of millions of people [Norris and Wilson 1997; Millán et al. 2004].

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_25,

Human hand morphology and human hand intensive use in interaction with objects, motivate researchers to use human hand in interaction with computers. Hand morphology gives hand large state space with 27 Degrees of Freedom (DOS) which gives capability to interact almost with any complicated object. This large state space also makes accurate human hand recognition and tracking more difficult than many other objects, especially in unknown environments where background is unknown or there are other moving objects.

There are mainly three approaches to recognize and track human hand: traditional approach, model based approach and appearance based approach. Traditional approach is the oldest one, it mainly depends on wearing some devices like data gloves to track hand movements. Gloves usually have sensors or retroreflective markers to collect joints angles and determine hand posture. Hardware usage makes traditional approach accurate and gives it the ability to build 3D model for human hand [Dorfmuller-Ulhaas and Schmalstieg 2001]. In practice, traditional approach is not widely used because hardware usage also makes it expensive and restricts hand movement. This leaded to innovate non-intrusive approaches like model based and appearance based approaches which are based on computer vision. Model based approach is used to recognize hand posture by using linear parameters and angular parameters to build a 3D model for the hand [Lien and Huang 1998]. Linear parameters represent fingers lengths and palm size, so it can differentiate between different persons. Angular parameters represent joints angles, so it can differentiate between different postures. To reduce state space, linear parameters are usually fixed and only angular parameters are changed to find the correct 3D hand model. To determine correct values of angular parameters, initial values are used and then reconfigured to minimize the error between projected image of 3D model and the input image [Stenger et al. 2001]. Model based approach is suitable for interaction with objects in virtual 3D environment but it is computationally complicated, also the final result totally depends on initial parameters values. Appearance based approach is used to choose the most appropriate hand posture from multiple pre-known postures set [Wang and Popović 2009]. It depends on extracting multiple features form each pre-known posture and using them to classify the input hand image based on its features. Mapping between hand postures space and 2D visualization space is not linear because multiple postures can have the same 2D image, also multiple images can be related to same posture [Haiying et al. 2006]. This means that accurate learning of 2D hand image features requires a huge number of samples.

In this paper we present a novel posture independent appearance based method to recognize and track human hand in an unknown environment. Our method uses posture independent features to distinguish between human hand and other moving objects in the scene. Results show that our method can efficiently and effectively recognize and track human hand in real-time.

2 System Overview

Our proposed system takes video stream from a digital camera as an input and gives the position of hand palm as the final output. System consists of two major phases: initial phase and tracking phase (Fig. 1). Initial phase finds hand window in the input frame while tracking phase tracks it. Initial phase is called for each input frame until we find a hand window, then tracking phase is called to track it. When hand goes out of camera view, initial phase is recalled to find hand window again.

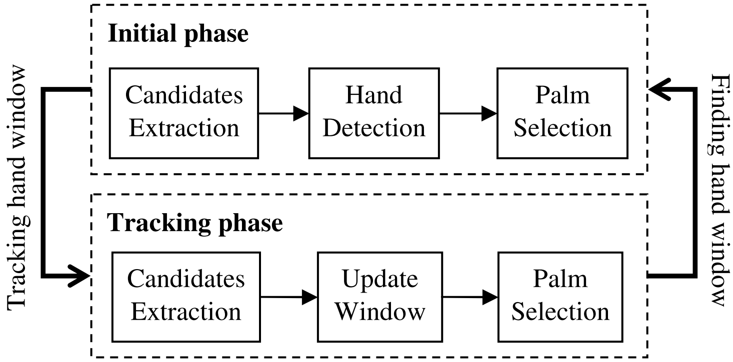


Fig. 1 System overview

Tracking phase reduces execution time as it replaces hand detection step with none computationally complicated update window algorithm.

2.1 Candidates Extraction

Candidates extraction step is the first step in both initial phase and tracking phase, it depends on motion and color to find candidate hand windows in the input frame. Fig. 2 shows candidates extraction step in details. First foreground areas and skin areas are detected in parallel. Then these areas are merged using "and" operation to find foreground skin areas. Finally foreground skin areas are segmented based on their motion to find candidate windows.

Foreground Detection

To detect foreground areas we try 3 different methods; frames subtraction, Mixture of Gaussians (MOG) [Stauffer and Grimson 2000] and Foreground Detection (FGD) [Li et al. 2003]. Frames subtraction method subtracts each pixel in frame *N-1* from the same pixel in frame *N* to find foreground pixels in frame *N*. MOG and FGD methods give many possible values for each pixel which allow them to learn background objects and distinguish them from foreground ones.

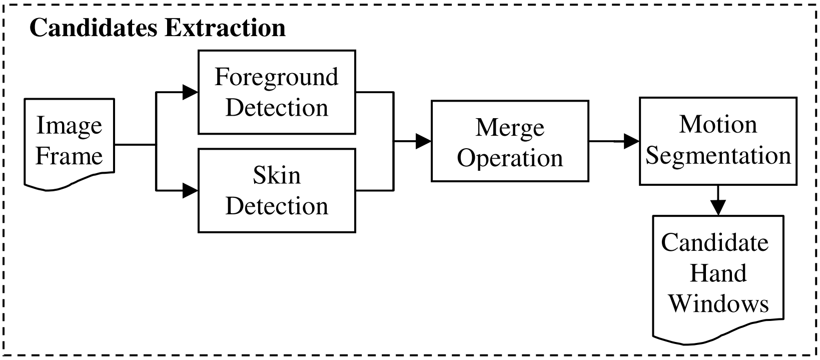
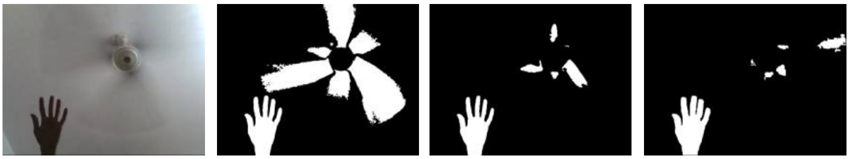


Fig. 2 Candidates extraction step

Practically, frames subtraction method fails to distinguish between background moving pixels and foreground pixels when there is repeated motion in background. MOG and FGD methods learn repeated background motions and show better results in detect foreground pixels (Fig. 3). We depend on MOG method in our work because it is faster than FGD method and shows almost the same results.

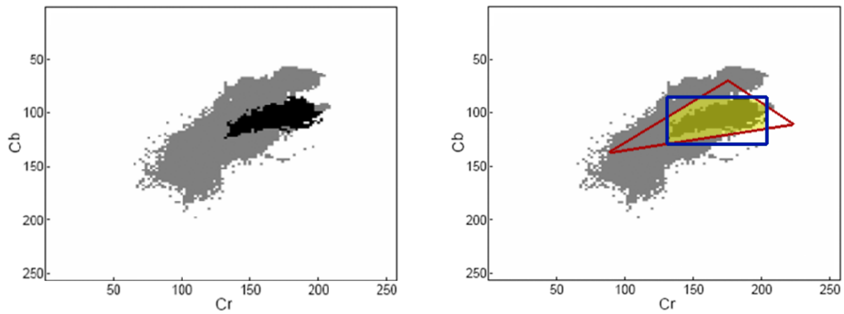


(a) Input frame (b) Frame subtraction (c) MOG method (d) FGD method

Fig. 3 Foreground detection methods. White color represents foreground areas

Skin Detection

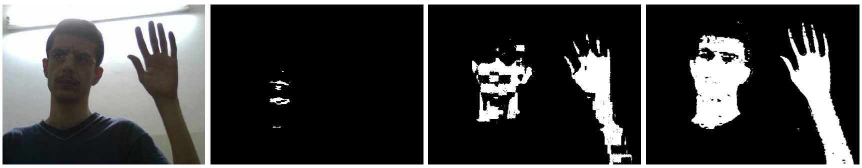
Many researches use color space ranges to detect skin areas. Color spaces that can isolate luminance component give better result than other color spaces. In this work we depend on a skin statistical study with more than 0.5 million positive samples and 1.2 million negative samples [de Campos 2006]. This study shows skin distribution along Cb and Cr axes in YCbCr color space as in Fig. 4 (a) where black area represented skin and gray area represented other objects. We can see that there is no CbCr range that can detect skin area and only skin area. To detect skin area with none computationally complicated method, we approximate skin area to the intersection area between a rectangle and a triangle as in Fig. 4 (b). We find that the smallest rectangle that contains skin area is the range: $130 \le Cr \le 204$ and $86 \le Cb \le 131$. The smallest triangle that contains skin area is determined by points: (177,70), (224,112), (88,138).



(a) Skin distribution (b) Skin area detector

Fig. 4 Skin statistical study result

Fig. 5 compares between 3 methods to detect skin areas: RGB range [Sung et al. 2008], YCbCr range [Chai and Ngan 1999] and our proposed method. RGB range method almost detects nothing, while YCbCr range method shows good detection result. Our proposed method shows better result than YCbCr range especially in detecting very dark and very light skin areas.



(a) Input image (b) RGB range (c) YCbCr range (d) Our method

Fig. 5 Skin detection methods comparison. White color represents skin areas

Merge Operation

After detecting foreground areas and skin areas, foreground skin areas (shared areas) are found by using "and" operation. Any pixel is considered foreground skin pixel if it is belong to a foreground area and belong to a skin area. We fill any holes inside foreground skin areas because hand area could not have holes inside it. Foreground detection is independent from skin detection, so we execute them in parallel. This allows us to dramatically reduce the computational complexity of candidates extraction step.

Motion Segmentation

After finding foreground skin areas, these areas are segmented based on motion speed and direction to find candidate hand windows (motion areas). We use Motion Template algorithm [Bradski and Davis 2000] to find candidate windows. This algorithm depends on timed Motion History Image (tMHI) which stores many sequential frames into one motion history gradient image.

2.2 Hand Detection

The second step in initial phase is hand detection. Posture independent features are used to select hand window from candidate windows. Hand detector depends on machine learning algorithm to learn features so it has two phases: offline training phase and online detection phase (see Fig. 6).

In offline training phase features are extracted from multiple positive/negative hand windows and passed to machine learning algorithm as a training set. Learning output is a trained model which is used in detection phase. In detection phase the same features are extracted from each candidate window and passed to the trained algorithm to determine which one is a hand window. We use 4 posture independent features to detect hand window: Eccentricity, Convexity defects, CbCrCount and FGNearCount.

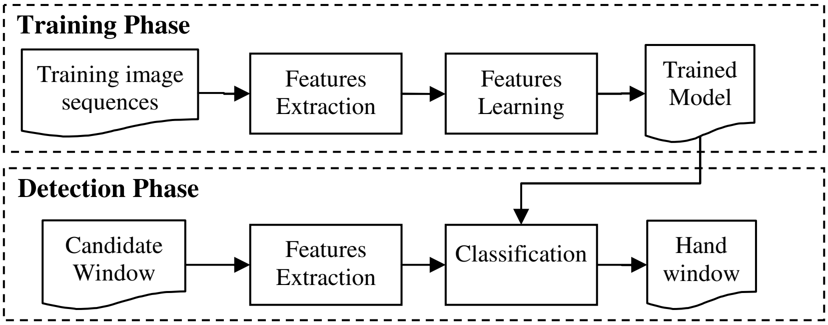


Fig. 6 Hand detection

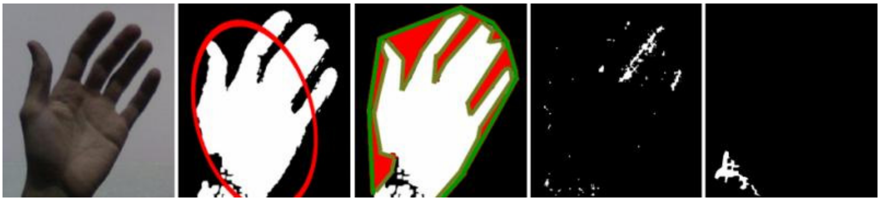
Eccentricity feature is the eccentricity value of the best ellipse that surrounding foreground skin area in the window. Ellipse eccentricity is calculated as the ratio between focal points distance and ellipse biggest diameter. This feature value is always in range [0,1]. The larger eccentricity feature value is, the more elliptical is the foreground skin area in candidate window.

Convexity defects value is the number of foreground skin area convexity defects which are placed between foreground skin body and foreground skin convex hull. This value counts areas between fingers in case of hand window.

CbCrCount feature represents number of non skin colors (in CbCr space) inside foreground skin area. Foreground skin area maybe contain some pixels that are not skin pixels. These pixels are added at closing holes step. To calculate this feature value, foreground skin area colors are converted to YCbCr space. After omit Y component, CbCr colors are passed to skin detector filter as mentioned in candidates extraction step to count number of non skin colors.

FGNearCount feature calculates the area of foreground objects around foreground skin area. This foreground objects are moving in the same direction as foreground skin area because they are in the same candidate window.

Fig. 7 shows main components to calculate these 4 features on hand window.



(a) Original image (b) Eccentricity (c) convexity defects areas (d) CbCrCount (e) FGNearCount

Fig. 7 Hand detection features

2.3 Update Window

After detect initial hand window in input frame, hand tracking phase is used to track hand window in following frames. In tracking phase, candidates extraction step is still executed for each frame. After candidates extraction, update window step is used to update hand window position. Update window step finds hand window in current frame based on hand window position in previous frames without calculating hand features. Update window depends on motion estimation where first and second derivative are used to compute the position of prediction hand window. Candidate hand windows are compared to prediction hand window position to choose the new hand window.

2.4 Palm Selection

After detecting hand window, we find palm region inside hand window. To find palm region, we find the center C of all edge points in hand window. Then we take the biggest window centralized on point C that is fully contained in hand window as palm region.

3 Results

We use F-Measure and ROC Curve to evaluate our system accuracy. Because our hand detection uses machine learning algorithm, system accuracy depends on learning algorithm accuracy. We try four common machine learning methods: Neural Network, Decision Tree, Naïve Bayes and Logistic Regression. Each algorithm is trained and tested on N images that are classified manually to positive and negative images. Positive images contain human hand while negative ones don't. 70% of these N images are used as training set and 30% are used as testing set.

Accuracy is measured for each algorithm based on F-Measure which is calculated from precision and recall equations as follows:

$$precision = \frac{TP}{TP + FP} \quad (1)$$

$$recall = \frac{TP}{TP + FN} \quad (2)$$

$$F - Measure = \frac{2\, precision.recall}{precision + recall} \quad (3)$$

Where TP is true positive samples, FP is false positive samples, FN is false negative samples and TN is true negative samples. Table 1 and Fig. 8 show F-Measure values for used machine learning algorithms where *N*=100, 250, 500, 750, 1000 and 1100 respect**ively.**

Table 1 F-Measure for used learning algorithms

	Neural Network	Decision Tree	Naïve Bayes	Logistic Regression
N=100	87.09%	80.41%	87.69%	86.66%
N=250	98.84%	89.76%	87.69%	95.08%
N=500	97.47%	99.14%	95.00%	95.08%
N=750	98.30%	97.47%	97.43%	94.30%
N=1000	97.47%	98.30%	97.43%	93.54%
N=1100	98.84%	98.30%	98.30%	95.86%

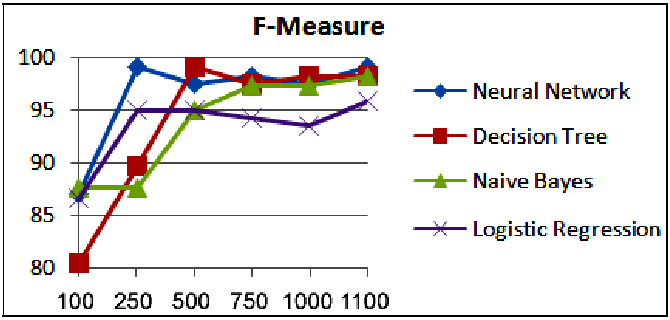


Fig. 8 F-Measure for used learning algorithms

We can see that all learning methods can't learn features when they are trained on a small number of images (less than 250 images). Neural network method learns hand features from 250 images while other methods need more images. All methods show almost the same results when they are trained on enough images (more than 500 images). This shows the superior of neural network method to all other common methods, because it needs less training set size than others. To implement neural network method we used multi-layer neural network with 4 input neurons, 11 hidden layer neurons and 2 output neurons.

F-Measure measures only sensitivity of methods not specificity. To measure both sensitivity and specificity we use ROC curve which measures the relation between true positive rate and false positive rate. Fig. 9 shows ROC curve for our neural network when *N*=1100 on testing set.

We can see that the best point for both true positive rate and false positive rate (upper left corner) gives us about 96% for true positive rate and about 6% for false positive rate, which confirms that our method has a high sensitivity and specificity for hand detection.

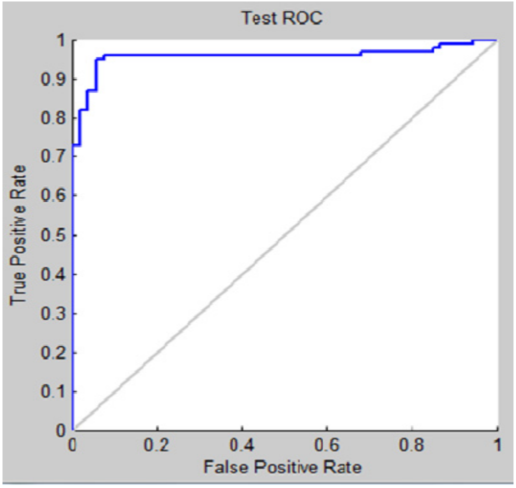


Fig. 9 ROC Curve for neural network method when *N*=1100

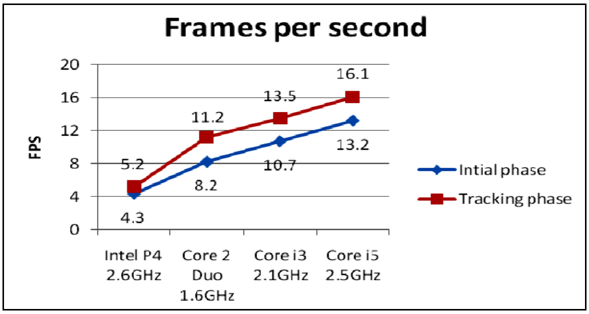


Fig. 10 System test results on different CPUs

To measure performance of our method we implemented it using C++ with OpenCV library and tested it using Genius eFace 2025 webcam with 640×480 size live videos. We measure detection performance improvement when CPU speed is increased. Fig. 10 displays both initial phase and tracking phase test results for different CPUs. We can see that tracking phase is 20% faster than initial phase.

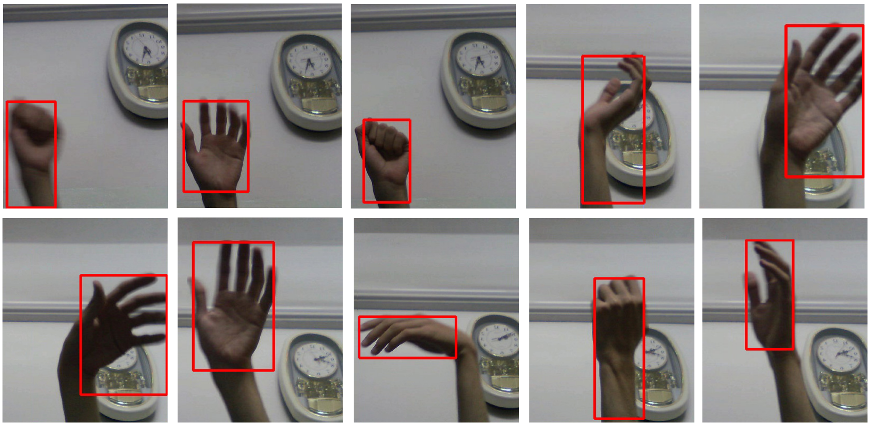


Fig. 11 Samples of detected hand regions

We can roughly say that system overall performance depends only on tracking phase performance, because initial phase is used only when there is no hand in the scene. Results show that our method can process more than 16 fps on 2.5 GHz CPU and number of processed frames per second is increased when CPU speed or number of cores are increased. Fig. 11 shows samples of detected hand regions from 3 different test videos in random postures.

4 Application

To test our method practically and get users feedback, we developed an interactive TV application that simulates normal TV and uses hand motion as interaction method instead of normal TV remote control. Our TV can recognize direction of hand movements (up, down, right, left) regardless of hand posture to simulate the 4 arrows on normal TV remote control. Also it can detect closed hand posture which simulates “OK” button. By using hand to send up, down, right, left and ok signals to TV you can show a favorite channels menu on TV screen, move a marker on its items, select a channel and also control volume up and down.

To evaluate our interactive TV, we asked 31 users (21 males and 10 females) between 21 and 55 years old to interact with it for about 5 minutes. After that each user evaluates her\his experience about comfort, accuracy and excitement on a scale of -3 to 3 comparing to normal TV remote control experience. Fig. 12 shows users evaluation results.

We can see that hand interaction is more excited than normal remote control while remote control is still a little more accurate. Hand interaction with TV is almost as comfort as using a remote control.

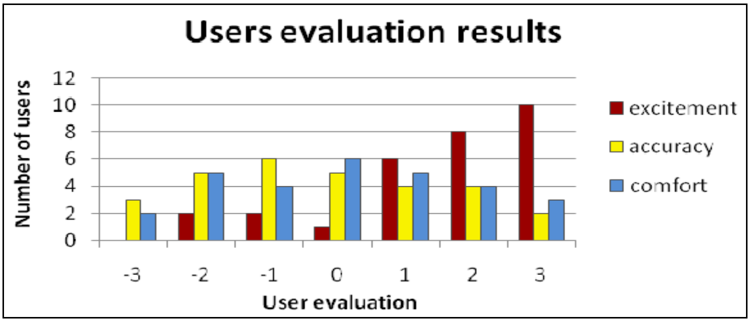


Fig. 12 Interactive TV interaction result

5 Conclusion

In this paper we proposed a new appearance based method to recognize and track bare human hand in an unknown environment without any constraints on hand posture. The proposed method depends on posture independent features to recognize and track hands. Results show that neural network learning algorithm is superior to other common used learning algorithms with higher accuracy rate. Results also show that our method is suitable for real-time applications.

References

[Bradski and Davis 2000] Bradski, G.R., Davis, J.: Motion segmentation and pose recognition with motion history gradients. In: Proc. 5th IEEE Workshop on Applications of Computer Vision, Palm Springs, CA, USA, pp. 238–244 (2000)

[Chai and Ngan 1999] Chai, D., Ngan, K.N.: Face segmentation using skin-color map in videophone applications. IEEE T. Circ. Syst. Vid. 9(4), 551–564 (1999)

[de Campos 2006] de Campos, T.E.: 3D Visual Tracking of Articulated Objects and Hands. Department of Engineering Science, University of Oxford, USA (2006)

[Dorfmuller-Ulhaas and Schmalstieg 2001] Dorfmuller-Ulhaas, K., Schmalstieg, D.: Finger tracking for interaction in augmented environments. In: Proc. IEEE and ACM International Symposium on Augmented Reality, New York, NY, USA, pp. 55–64 (2001)

[Haiying et al. 2006] Haiying, G., Feris, R.S., Turk, M.: The isometric self-organizing map for 3D hand pose estimation. In: 7th International Conference on Automatic Face and Gesture Recognition, Southampton, UK, pp. 263–268 (2006)

[Kohda et al. 1976] Kohda, M., Nakatsu, R., Shikano, K.: Speech recognition in the question-answering system operated by conversational speech. In: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, PA, USA, pp. 442–445 (1976)

[Li et al. 2003] Li, L., Huang, W., Gu, I.Y.H., Tian, Q.: Foreground object detection from videos containing complex background. In: Proc. ACM International Conference on Multimedia, Berkeley, CA, USA, pp. 2–10 (2003)

[Lien and Huang 1998] Lien, C.-C., Huang, C.-L.: Model-based articulated hand motion tracking for gesture recognition. Image Vision Comput. 16(2), 121–134 (1998)

[Millán et al. 2004] del R. Millán, J., Renkens, F., Mouriño, J., Gerstner, W.: Brain-actuated interaction. Artificial intelligence 159(1-2), 241–259 (2004)

[Norris and Wilson 1997] Norris, G., Wilson, E.: The eye mouse, an eye communication device. In: Proc. IEEE 23rd Northeast Bioengineering Conference, Durham, NH, USA, pp. 66–67 (1997)

[Stauffer and Grimson 2000] Stauffer, C., Grimson, W.E.L.: Learning patterns of activity using real-time tracking. IEEE T. Pattern Anal. 22(8), 747–757 (2000)

[Stenger et al. 2001] Stenger, B., Mendonca, P.R.S., Cipolla, R.: Model-based 3D tracking of an articulated hand. In: Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, HI, USA, vol. 2, pp. 310–315 (2001)

[Sung et al. 2008] Sung Kwan, K., Mi Young, N., Phill Kyu, R.: Color based hand and finger detection technology for user interaction. In: Proc. International Conference on Convergence and Hybrid Information Technology, Daejeon, Korea, pp. 229–236 (2008)

[Wang and Popović 2009] Wang, R.Y., Popović, J.: Real-time hand-tracking with a color glove. ACM Trans. Graph. 28(3), 1–8 (2009)

Vanishing Point Estimation in Urban Roads for Omnidirectional Image

D.C. Hernández, V.-D. Hoang, and K.-H. Jo

Intelligent Systems Laboratory, Graduate School of Electrical Engineering,
University of Ulsan, Ulsan 680-749, South Korea
{danilo,hvzung}@islab.ulsan.ac.kr, acejo@ulsan.ac.kr

Abstract. Regarding the autonomous of robot navigation, vanishing point (VP) plays an important role in visual robot applications such as iterative estimation of rotation angle for automatic control as well as scene understanding. Autonomous navigation systems must be able to recognize feature descriptors. Consequently, this navigating ability can help the system to identify roads, corridors, and stairs; ensuring autonomous navigation along the environments mentioned before the vanishing point detection is proposed. In this paper, the authors propose solutions for finding the vanishing point in real time based density-based spatial clustering of applications with noise (DBSCAN). First, we proposed to extract the longest segments of lines from the edge frame. Second, the set of intersection points for each pair of line segments are extracted by computing Lagrange coefficients. Finally, by using DBSCAN the VP is estimated. Preliminary results are performed and tested on a group of consecutive frames undertaken at Nam-gu, Ulsan, South Korea to prove its effectiveness.

1 Introduction

Autonomous ground navigation is still facing important challenges in the field of robotics and automation due to the uncertain nature of the environments, moving obstacles, and sensor fusion accuracy. Therefore, for the purpose of ensuring autonomous navigation and positioning along the environments aforementioned, a visual based navigation process is implemented by using an omnidirectional camera. The camera must operate alongside a variety of portable sensing devices such as Global Positioning System (GPS), Inertial Measurement Unit (IMU) and Laser Range Finder (LRF) and online interactive maps. Based on the perspective drawing theory, one VP is a point in which a set of parallel lines converge and disappear into the horizon. Ultimately, by using VP for indoor and outdoor in 2D images, roads as well as corridors can be described as a set of orthogonal lines due to the set of parallel lines that typically exhibit these structures in 3D scenes, see Fig. 1a. In the field of autonomous navigation systems, efficient navigation, guidance and control design are critical in averting current challenges. In this sense, referring to the need for estimation of rotation angle for automatic control,

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_26,

one VP plays an important role at this stage. By detecting VP in the 2D image, autonomous unmanned systems are able to navigate towards the detected VP. Several approaches for VP detection have been used, for example, Hough transforms (HT), RANSAC algorithm, dominant orientation, and more recently equivalent sphere projection. In the case of HT, for estimating lines into the 2D images [Samadzadegan 2006] propose the randomized HT to estimate the lane model based VP voting, [Ebrahimpour et al 2012] propose a VP detection method based HT and K-mean algorithm mainly based on the straight lines orientation given by the edges of corridors. In [López et al 2005] authors use the RANSAC approach to describe a parametric model of the lane marking into the image. The clear examples of dominant orientation approaches are [Miksik et al 2011; Kong H et al 2009] in which authors detected the VP by using a bank of 2D Gabor wavelet filters with a voting process. In case of spherical representation, [Bazin and Pollefeys 2012; Mirzaei and Roumeliotis 2011] authors used the 3-D line RANSAC approach in real time. In the case of omnidirectional scenes, parallel lines are projected as curves that converge and disappear into the horizon, see Fig. 1b. Therefore, in order to address the challenges of VP detection in omnidirectional scenes, the authors decided to present a real time VP detection based on the polynomials approach and clustering. To this end, the main contributions of the presented method are:

- Implementation of real time curve fitting algorithm for VP detection,
- Implementation of DBSCAN, due that the number of cluster in the image are not specified.

The rest of this paper is structured as follows: (2) Proposed Method, (3) Experimental Result, (4) Conclusions and Future Works.

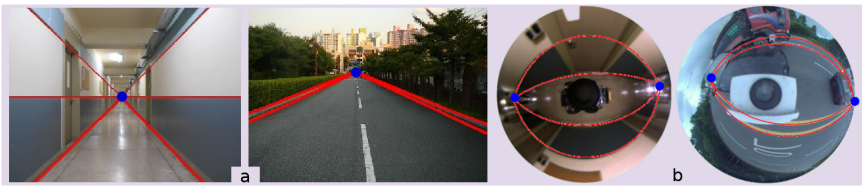


Fig. 1 Perspective projections of 3D parallel lines are projected as straight but diagonal lines or as curves in the 2D images space. Lines in represent orthogonal lines set towards the vanishing point. Points in blue represent the vanishing point on the horizon where all the orthogonal lines converge. (a) VP perspective in single CCD images. (b) VP perspective in omnidirectional images

2 Proposed Method

Essentially, the proposed method consists mainly in extracting the information surrounding the ground plane. Frames are extracted in short time intervals that started just before the earliest detection from the video capture sequence. In this section, the proposed algorithm for Iterative Vanishing Point Estimation Based on DBSCAN for Omnidirectional Image has three steps: (1) extracting line segments, (2) curve fitting, (3) cluster extraction by DBSCAN.

2.1 Extracting Line Segments

Road scenes can be described as structures which contain lane marking, soft shoulders, gutters, or a barrier curb. Therefore, in 2D images these features are represented as a set of connected points. The main idea of this section consisted of extracting the longest line segments around the ground by applying the canny edge detector from a group of consecutive frames captured in Nam-gu, Ulsan, South Korea. After the edge detection step, the subsequent task was to remove the smallest line segments by extracting the longest line segments after applying canny edge detection. This was achieved by using a 3 chain rule for each possible curve candidate, see Fig 2. From the tracking algorithm were able to extract the basic information such as: length (l), number of point per line segments as well as the pixel position location of the points of each line segments into the image plane. As a result, extraction from an image sequence the set of longest line segments was completed. Fig. 3 shows us the result of this step.

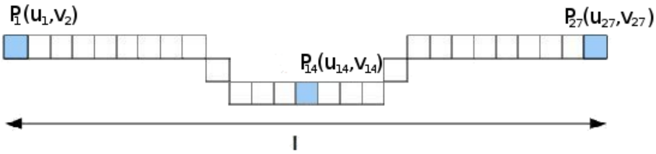


Fig. 2 Curve model

2.2 Curve Fitting

At this point the set of line segments in the image plane are known. Hence, the new task consists of defining the function for each line segment. The process starts at every iteration by selecting a set of three random points for each line segment. This will help to describe the function that determines the curvature of the lines. As far as we know, in numerical analysis there are various approaches for solving a polynomial system. For example, in [Bazin and Pollefeys 2012] authors use polygonal approximation to extract that lines segments. On the other hand, in [Mirzaei and Roumeliotis 2011] authors use an Auzinger and Stetter method. In that sense, the main contribution of the authors work is to estimate VP in real time processing, required by autonomous navigation. To this end, Lagrange polynomials are used to develop the fitting polynomial. For example: given the data set of points (x1,y1), (x2,y2), ..., (xn,yn) the interpolation polynomial is defined as follows:

$$fn(x) = \sum_{j=1}^{n} L_j(x) \cdot y_j \tag{1}$$

$$L_j(x) = y_j \prod_{\substack{k=1 \\ k \neq j}}^{n} \frac{x - x_k}{x_j - x_k} \tag{2}$$

Thus, the given 3 points of the interpolation polynomial equate to:

$$fn(x) = y_1 \frac{(x - x_2)(x - x_3)}{(x_1 - x_2)(x_1 - x_3)} + y_2 \frac{(x - x_1)(x - x_3)}{(x_2 - x_1)(x_2 - x_3)} + y_3 \frac{(x - x_1)(x - x_2)}{(x_3 - x_1)(x_3 - x_2)} \tag{3}$$

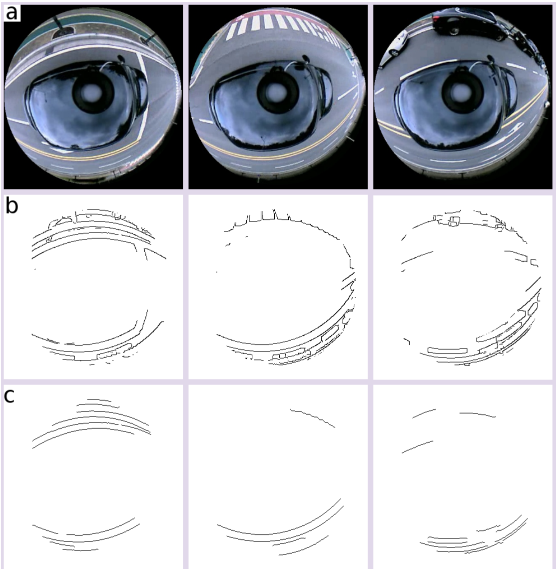


Fig. 3 Longest line segments. (a) Selected frames sequences taken at different time steps. (b) Edge detection results. (c) Extracting line segments after applying edge detection step

Once the polynomials have been disclosed (see Fig 4b), the next step consists of selecting the appropriate curve fit model by finding the function that contains the largest numbers of inliers by using RANSAC. In other words, the curve model that has the largest number of points based on the given data set becomes the points that are extracted, see Fig. 4c. As a result, the set of curves are clearly defined. The next step entails the extraction of the data set of intersection point by finding the same for each pair of defined curves at the point where they intersect, see Fig. 4c.

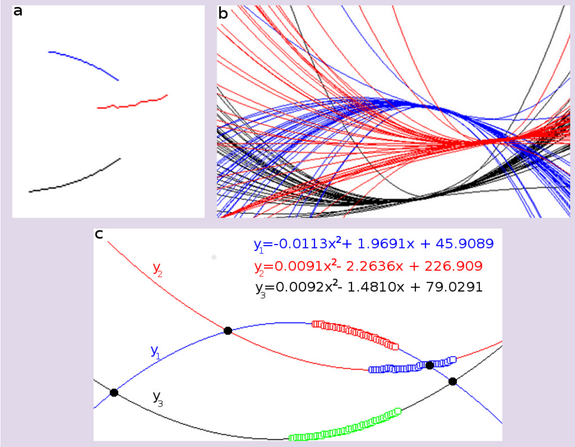


Fig. 4 Polynomial fitting curve results. (a) Candidate line segments. (b) Iterative curve fitting.(c) RANSAC curve fitting results. Curves in color show the function that contains the largest numbers of inliers. Points in black show the intersection point between the defined curves.

2.3 Cluster Extraction by DBSCAN

Given the data set of intersection point from the previous step, the algorithm should be able to extract clusters. From Fig. 5, it is clear that the projected data into the image plane give us vital information about the data set points, as follows:

- The data does not depict a well-defined shape,
- Presence of noise in the data due to the lack of a pre-processing model for road analysis. To compensate, all line segments are considered as a part of the road,
- The number of cluster could not be described in advance.

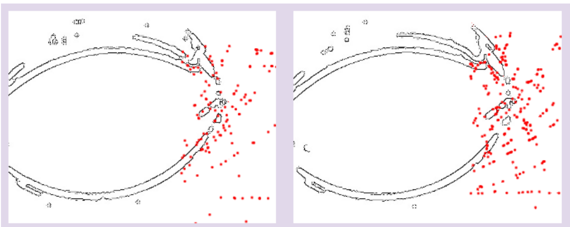


Fig. 5 Result of the intersection point step from two different frame sequences

Considering the above mentioned occurrences, among the various clustering algorithms it was proposed to use the DBSCAN algorithm [Ester et al 1996]; an unsupervised method, due to the algorithm having achieved a good performance with respect to some other algorithms. In essence, the main idea of DBSCAN is that for each point of a cluster, the neighborhood of a given radius has to contain a minimum number of points.

The DBSCAN algorithm depends mostly on two parameters:

- Eps: number of points within a specified radius
- MinPts: minimum number of points belonging to the same cluster.

After the candidate clusters are formed, the idea is to define the cluster which contains the largest amount of points. Then, the centroids of the cluster are calculated, see Fig.6. It's important to remark that one of the main advantages of using omnidirectional cameras lies in the fact of the possibility to extract clusters which are located either in the front most or rear most part of the vehicle. Finally, once the VP has been extracted, the angle (θ) between the line formed by the VP and the center point of the omnidirectional camera, in respect to the x-axis of the coordinate system (which passes through the center point of the omnidirectional camera) is now able to be accurately computed, see Fig. 7

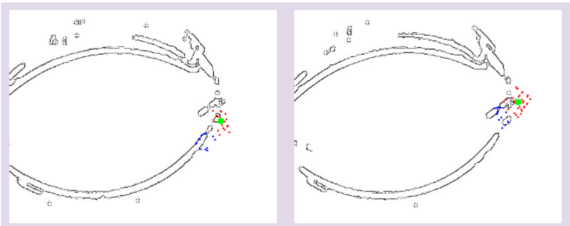


Fig. 6 DBSCAN results from different frame sequences. The point in green shows the estimated VP result. Points in red show the cluster with the largest amount of points. Points in blue show the cluster with the lowest amount of points

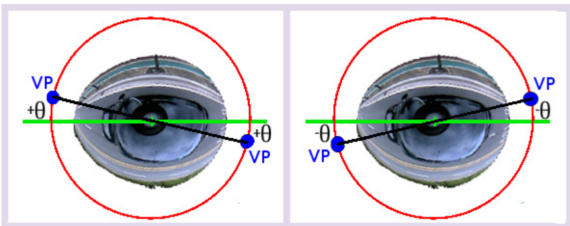


Fig. 7 Coordinate system. The green line shows to the x-axis of the coordinate system (which passes through the center point of the omnidirectional camera). The line in black is the line formed by the VP and the center point of the omnidirectional camera. Theta (θ) is the angle between the line formed by the VP and the center point of the omnidirectional camera, in respect to the x-axis

3 Experimental Results

In this section, the ending results of the experiment will be introduced. All the experiments were done on Pentium Intel Core 2 Duo Processor E4600, 2.40 GHz, 2 GB RAM and the implementation was done in C++ under Ubuntu 12.04. The algorithm used a group of 4,355 consecutive frames taken at Nam-gu, Ulsan, South Korea with a frame size of 160x146 pixels. On the one hand, the chosen path contains a set of a good example of different road geometric design, such as crossroad, turning lanes, sag and crest curves, fork, straight roads, underpass and overpass sections. On the other hand, the scene also contains different urban objects like cars, trees, buildings, as well as frames with problems given by the illumination distribution, strong sunlight and dark shadows. The map in Fig. 8 shows the chosen trajectory, it consists of approximately 3 km drive from A to B. The consecutive frame consists on urban road scene image. The proposed algorithm had a frame rate of approximately 20 frames per second. The depicted results are to the implementation of the Lagrange polynomial, RANSAC curve fitting and DBSCAN. The experimental result of the proposed method in different frame sequences are shown in the closing set of images. Fig. 9 shows the extracting line segments and the curve fitting steps and Fig. 10 shows the VP estimation result. Fig. 11 and Table 1 show the computing time for main phase of processing.

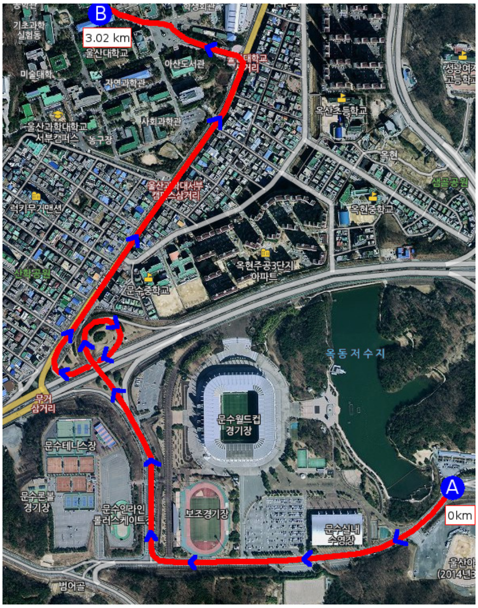


Fig. 8 The map shows the location of Nam-gu area in Ulsan, where the consecutive frames were taken. The average distance between A and B is approximately 1.87 miles (3.02 km.). The line in red shows the route that has been used for testing the proposed algorithm; the blue arrow shows the trajectory

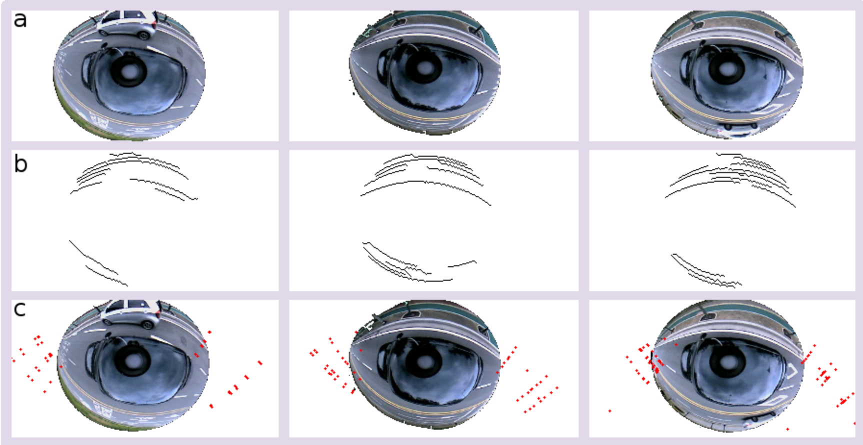


Fig. 9 Experimental results from different frame sequences. (a) Selected frames sequences taken at different time steps. (b) Extracting longest line segments after applying edge detection step. (c) Intersection point results by using Lagrangian polynomial interpolation

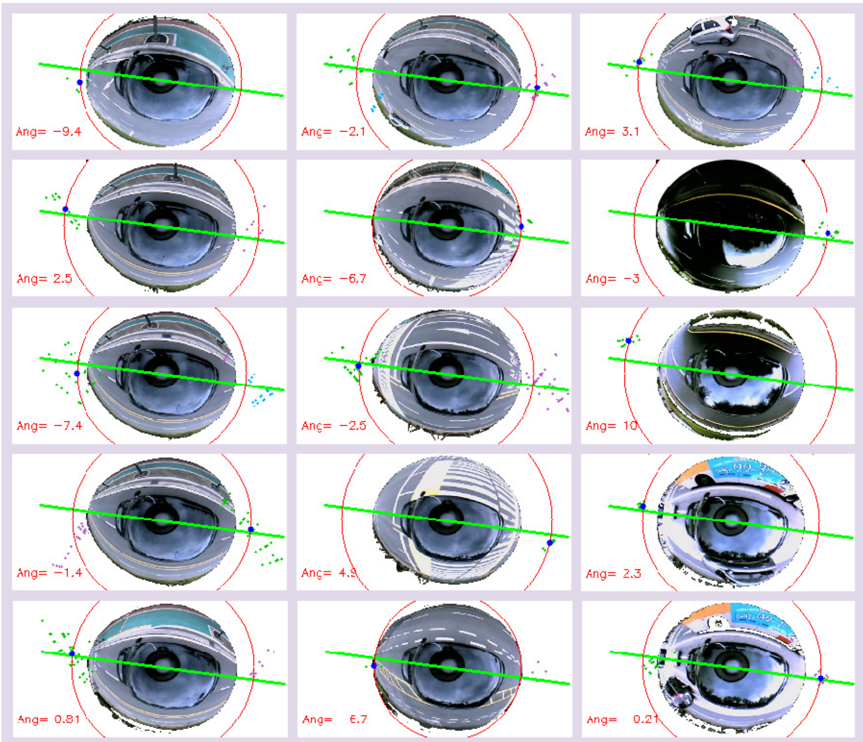


Fig. 10 VP estimation results from a different frame sequences. The point in blue shows the estimated VP. Points in others colors show the detected set of clusters in the current frame

Table 1 Computing time for main phase of processing

Process	Average time (msec.)
Capture	3.10
Preprocessing	3.47
Line extracting	24.80
RANSAC	10.11
DBSCAN	5.63
Total running time	47.11

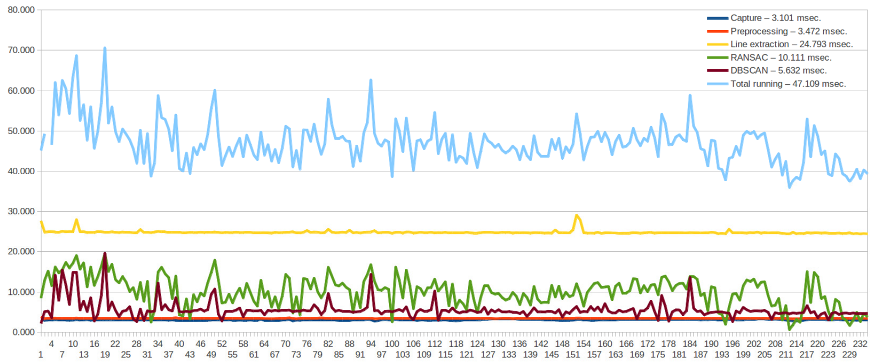


Fig. 11 Frame-rate image sequences at a resolution of 160x146 pixels. Data in blue sky represent the total running time. Data in yellow represent the line extraction processing time. Data in green represent the RANSAC curve fitting processing time. Data in brown represent the DBSCAN processing time. Data in orange represent the preprocessing time

4 Discussion and Conclusions

The preliminary results of the proposed method presented relevant information for finding the VP estimation in road scenes. A group of imaging results shows the performance of our proposed work over a set of possible scenarios. For example, in Fig. 12 the proposed algorithm works well despite the problem given by the illumination distribution, strong sunlight, and dark shadows. Initially, it could be concluded that these problems did not appear to affect the image, thus the algorithm was able to estimate de VP. On the other hand, Fig. 13 shows the result of false VP estimation. The first two images show the VP extracting from the largest amount of polynomials segments located on the left side of the camera system in contrast to the right side. The last two images show the VP when the driver of a vehicle is turning to the left (the same phenomena also happen when turning to the right), due to the projection properties of the lines segments into the mirror at that moment. This results indicate that the vanishing point has an error in rotating sensitive directions as well as places where single or double lines do not appear at all due to weather conditions, wildlife, pavement deterioration or where turn lanes are not marked in places like junctions, intersection, etc. On the one hand, this is due to the fact that pictures lose a lot of detail during resizing. On the other hand, the continues white lines along the road segments were not extracted from the binary image obtained after applying the Canny edge detector with a length l > 25, where l is the Euclidean distance between the endpoints of the line segments. As a result, VP were extracted using the spatial density information from the side strip areas (bicycle and/or planter) or sidewalks. As a result, the proposed algorithm is able to determine VP in the 3D space. However, in order to improve the VP robustness, there will be improvements to the performance and the experiment by using road model features, surface as well path planning approaches. Regarding autonomous navigation, this result reinforces the point of usage a set of sensor (GPS, IMU, LRF and online interactive maps) for dealing with the autonomous navigation problem in real time.

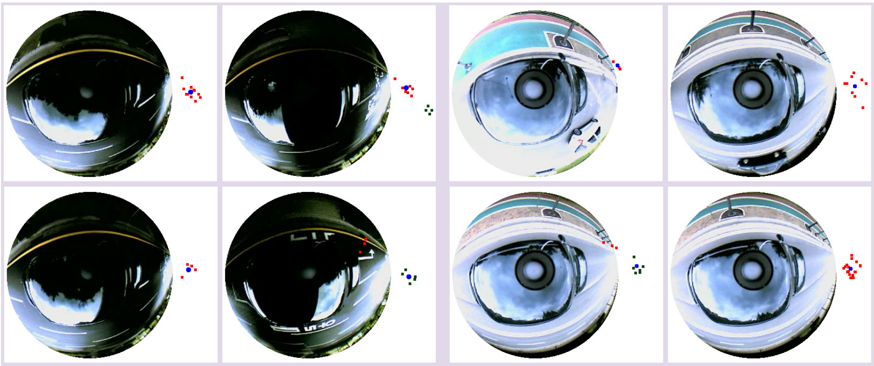


Fig. 12 Experimental results for a frame sequences with strong sunlight and dark shadows areas. Point in red and green represent the extracted cluster. Points in blue represent the VP

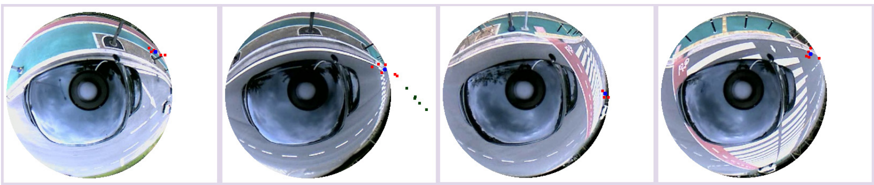


Fig. 13 Experimental results for a frame sequences with false VP estimation. Point in red and green represent the extracted cluster. Points in blue represent the VP

Acknowledgment. This research was supported by the MOTIE (The Ministry of Trade, Industry and Energy), Korea, under the Human Resources Development Program for Converge Robot Specialists support program supervised by the NIPA (National IT Industry Promotion Agency)(H1502-13-1001).

References

[Bazin and Pollefeys 2012] Bazin, J.-C., Pollefeys, M.: 3 – line RANSAC for Orthogonal vanishing point detection. In: IEEE/RSJ Inter. Conf. on Intelligent Robots and Systems, Algarve, Portugal, pp. 4282–4287 (2012)

[Bosse M et al. 2002] Bosse, M., Rikoski, R., Leonard, J., Teller, S.: Vanishing point and 3D lines from omnidirectional video. In: Proc. Inter. Conf. on Image Processing, New York, USA, vol. 3, pp. 513–516 (2002)

[Ebrahimpour et al. 2012] Ebrahimpour, R., Rassolinezhad, R., Hajiabolhasani, Z., Erahimi, M.: Vanishing point detection in corridors: using Hough transform and k-means clustering. IET Computer Vision J. 6, 40–51 (2012)

[Ester M et al. 1996] Ester, M., Kriegel, H.-P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proc. of 2nd International Conference of Knowledge Discovery and Data Mining, Oregon, USA, pp. 226–231 (1996)

[Kong H et al. 2009] Kong, H., Jean-Yves, A., Ponce, J.: Vanishing point detection for road detection. In: IEEE Computer Society Conference on Computer Vision and Patter Recognition, Miami Beach, USA, pp. 96–103 (2009)

[López et al. 2005] López, A., Cañero, C., Serrat, J., Saludes, J., Lumbreras, F., Graf, T.: Detection of lane marking based on ridgeness and RANSAC. In: Proc. of the 8th International IEEE Conference on Intelligent Transportation Systems, Vienna, Austria, pp. 733–738 (2005)

[Miksik et al. 2011] Miksik, O., Petyovsky, P., Zalud, L., Jura, P.: Detection of shady and highlighted roads for monocular camera based navigation of UGV. In: IEEE Inter. Conf. of Robotics and Automation, Shanghai, China, pp. 4844–4849 (2011)

[Mirzaei and Roumeliotis 2011] Mirzaei, F.M., Roumeliotis, S.I.: Optimal estimation of vanishing points in a Manhattan World. In: IEEE International Conference on Computer Vision, Barcelona, Spain, pp. 2454–2461 (2011)

[Samadzadegan et al. 2006] Samadzadegan, F., Sarafraz, A., Tabibi, M.: Automatic lane detection in image sequences for vision-based navigation purposes. In: Proc. of the ISPRS Commission V Symposium Image Engineering and Vision Metrology, Dresden, Germany, vol. XXXVI(5) (2006)

Improvement of Assistive Robot Behavior by Experience-Based Learning

P.M. Nauth

Department of Computer Science and Engineering, University of Applied Sciences
Frankfurt a. M., Frankfurt a. M., Germany
pnauth@fb2.fh-frankfurt.de

Abstract. Robots designed for assisting humans in their homes need to adapt to the changing requirements of daily life. This requires multimodal sensor systems as well as learning strategies for understanding new goals and for recognizing new objects. However, coping with changes is not limited to environmental sensing. In order to achieve full autonomy, the robots must adapt their behaviour due to good and bad experiences made. Concepts and first results of modelling intelligent sensing and adaptive behaviour in an artificial mind as well as of merging mind and machine are presented in this paper.

1 Introduction

Nowadays, assistive robots can derive goals from instructions such as spoken commands and achieve them by executing pre-programmed trajectories [Mertens et al. 2011]. A certain degree of adaptability is provided by fine-tuning these trajectories by means of proximity sensors and smart cameras which enable the robot to avoid obstacles and to guide a robot gripper to an object to be grabbed. Robots such as the Care-O-Bot (Fraunhofer Institute, Germany) or the Asimo (Honda, Japan) come with these abilities.

However, in order to perform tasks more efficiently, assistive robots must be able to cope with changes in the home environment and with new demands of the user autonomously, i.e. they must learn from experiences. A robot equipped with an experienced-based learning strategy would avoid a dangerous or non-effective behavior after having experienced the consequences of this behavior as negative, once. A positive experience would enforce a given behavior.

Our approach for adaptive behavior is based on artificial feelings, which are modeled by Quality of State (QoS) functions. First studies with a 30 cm tall humanoid robot showed that using artificial emotions can change the robot's behavior based on previous experiences.

Currently, we embody these algorithms for intelligent behavior in a larger robot, i.e. we merge mind and machine in order to test our approach in real applications.

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_27, © Springer International Publishing Switzerland 2014

2 Embodiment and Functional Model

As a first step to merge mind and machine we are developing a semi-humanoid robot (Fig. 1) which is 1.40 meters tall. The name of the robot is “Roswitha" (Robot System with Autonomy). Most of the mechanical components are based on a Volksbot [Surmann et al. 2008]. Both front wheels are driven by with one Maxon motor RE40each, while the two rear wheels run freely. This mechanical design has been expanded by several intelligent sensors and two robot arms: the joints and links of the right one are composed of modules of the company “Schunk”, whereas the left one consists of servo motor joints and links from the company “Item”. The programs for sensor data processing, decision making, and motor control are executed on a laptop and an embedded system “cRIO” from the company “National Instruments”, which are integrated in the robot’s mechanics.

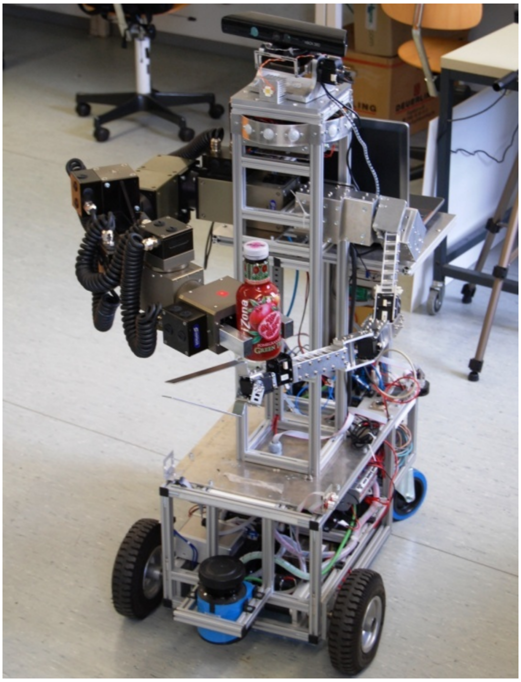


Fig. 1 Mechanical design of the robot “Roswitha”

Programming is carried out in LabVIEW which provides us with a unified development environment and an extensive driver library. As autonomous robots are complex systems and consist of many different components, this is important to management and execution of development processes. It also fosters the research environment of universities of applied sciences, because their research activities are based in large part on sequences of short-term theses. The periods of vocational adjustment for students at the beginning of their thesis are minimized.

3 Intelligent Multimodal Sensor System

Our multimodal sensor system consists of sensors as follows:

- Speech recognition sensor or touch for human machine interaction.
- 2-D Laser - proximity sensor in the front lower part of the robot to detect obstacles and landmarks that allow to calculate the robot's position. The sensor delivers a distance array (Fig. 2) and a remission array which contain the distance to the nearest object and its reflection intensity for each angle scanned. The angle resolution is set to 0.5°.
- Several ultrasonic sensors for obstacle detection in the areas of the laser-scanner does not cover (e.g. in the rear part of the robot).
- Contact and temperature sensors for feedback of dangerous situations.
- Vision sensor consisting of a 3-D camera and an image processing algorithm for the detection and classification of objects such as beverage bottles and the determination of their position in order to guide the gripper to the object.

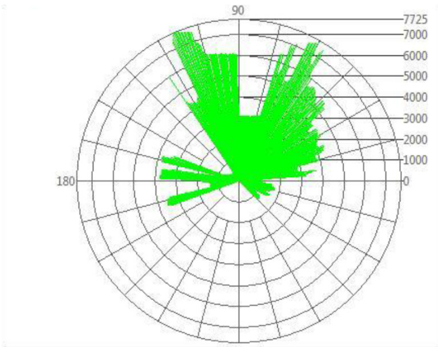


Fig. 2 Display of a distance array for a range of 0 to 7725 mm and -45° to 225°

The vision and the speech recognition sensor access a knowledge base for classification. It is generated by supervised learning.

A typical scenario is the following: The user says or enters the commands related to the item that he wants to have brought such as a water bottle and, if available, the destination area in which the item's location is to be expected such as the kitchen table. By means of the sensors and a map which represents the geometry of the rooms and known obstacles [Jin et al. 2003], the robot moves to the destination area. Obstacles that are not included in the map are detected by the sensors and cause evasive maneuvers. In the destination area the vision sensor identifies the desired object, calculates the object's position, grabs it and brings it to the user.

Fig. 3 shows the functional model addressing such or similar scenarios. First, the robot must understand the goal or task to be solved. Sensing the environment and fusing the sensor data, "Roswitha" knows her own position as well as the identity and position of objects and obstacles around her. Using these and the map data, the robot can now develop a strategy to solve the task by calculating the optimal trajectory and controlling the motors accordingly. As for behavior improvement [Nauth 2010], a planning algorithm must be implemented that use a experience base which store the positive and negative experiences represented by weight vectors. Finally, an action planning module generates the control sequences for the actuators and executes them.

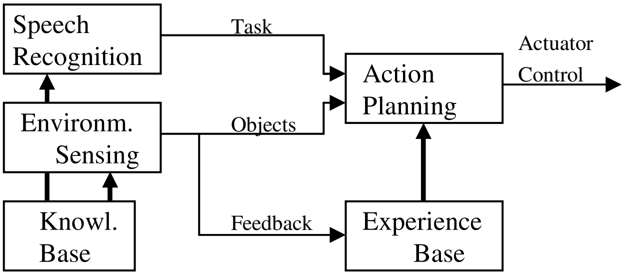


Fig. 3 Functional model of the robot

Due to the complexity of the overall system, in the current stage of development we realize the robot as two platforms: One stationary where the algorithms and hardware for the camera-guided grabbing are being explored [Gonzalez and Woods 2008], and one mobile where the sensor- and map-based driving and first experiments for adaptive behavior are carried out.

Our approaches for navigation, identifying objects with the vision system and adaptive behavior are discussed below.

4 Navigation

The navigation system guides "Roswitha" to the location where the object to be grabbed is placed. There she searches for it by means of the vision system (see chapter 5) and grabs it if it is in reach of the actuator. Otherwise, "Roswitha" moves closer to the object before grabbing it.

"Roswitha" navigates from her current position to the destination position by means of a map in which areas of known obstacles and areas the robot is not allowed to enter are marked.

Two highly reflecting foils of the size of 15 cm x 20 cm are used as landmarks for calculating the current position. Those landmarks are placed at two adjacent walls perpendicular to each other and are detected by the Sick laser scanner. The procedure to distinguish the landmarks from other objects is as follows:

- Calculate the mean remission out of all reflections from the whole scanning area.
- Remissions larger than 1.3 of the mean are considered to be caused by reference marks.
- If more than two or less than two landmarks have been detected the robot moves a little bit and repeats the search for two landmarks.
- If two landmarks have been detected the distances to the robot, r_1 and r_2, and the angles to the robot's main axis, θ_1 and θ_2, are stored.
- "Roswitha's" current position is calculated by means trigonometry using r_1, r_2, θ_1, and θ_2 as well as the known distances a and b of the landmarks from the corner of the adjacent walls they are placed.

The map consists of a grid with cells of 5 cm x 5 cm resolution. Known obstacles and forbidden areas are plotted in black and red respectively by common drawing programs. Minimizing the costs regarding all possible paths, the optimal path consisting of a set of sub-paths is calculated by means of the Occupancy Grid method. The result is a list of control sequences consisting of the distance and direction to move in each sub-path. In order to measure the distance travelled each wheel on the rear axis is equipped with an encoder resolving 0.18° per pulse.

Moving from the current to the goal position, the calculated path is left temporarily in case the laser scanner detects an object other than this to be fetched. “Roswitha” makes a detour around it and returns to the optimal path to reach the goal. In case of using Experience-based Learning she would avoid or approach objects depending on the experience made previously.

5 Actuator Guidance by Robot Vision

The smart camera consists of a Kinect 3D camera and an image processing program that has been developed with LabVIEW and runs on a laptop. The Kinect camera generates a color image c(x, y) by means of a CCD sensor and a depth image z (x, y) by means of structured illumination with IR light. As for the algorithm, we use the pattern matching function of LabVIEW. It searches in the acquired image for patterns which are similar to one of previously learned reference patterns (Fig. 4). Color and shape are the comparison criteria. The pattern matching functions can be parameterized in sensitivity, location- and size-invariance. In a first approach, we applied pattern matching to the grabbed color image c(x, y) in order to differentiate bottles regarding color and shape. The approach reaches its limits if the background contains colors and shapes similar to the object to be recognized, e.g. in pictures on the wall. Applying this approach we achieved a sufficient specificity only at the cost of a low sensitivity [Toth 2009].

Fig. 4 Pattern Matching with the color image

Therefore, we have developed a sequential process: The first step is to segment the depth image z(x, y) (Fig. 5, top) by masking the range between z = 80 cm to z = 160 cm as foreground, corresponding to the grabbing area of the robot arm and applying a pattern matching with respect to shape. If an object is found, the bounding box coordinates are being calculated and transferred to the color image c(x, y) (Fig. 5, bottom). In this area the pattern matching is repeated, but this time with respect to the color pattern. Hence, all patterns outside of the gripping area are not taken into account and a high specificity and high sensitivity can be achieved.

For the robot coordinate system is displaced from the camera coordinate system, the real position of the object to be grabbed results from the bounding box coordinates corrected by a coordinate system transformation. The control sequences for the robot arm joints are calculated by inverse kinematic algorithms converting the object position into angles the joints must be set to.

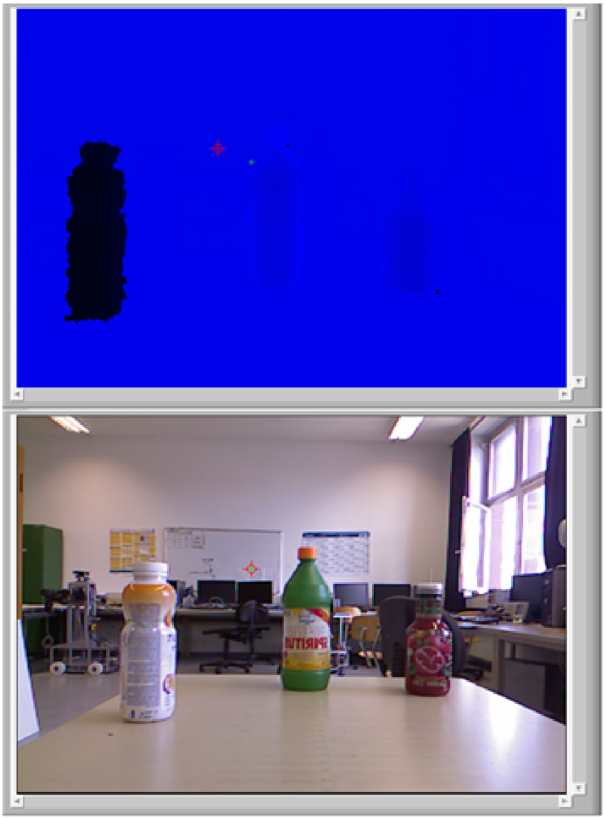


Fig. 5 Segmentation in the depth image

6 Behavior Improvement by Experiences

The intelligence of the robot described above is limited to learning new object classes and words. It understands goals and attains them straightforward by planning and executing a sequence of actions. However, good and bad experiences made do not result in an improved behaviour such as avoiding dangerous situations which might harm the robot. In order to enable the robot to adapt its behaviour by means of experiences, we propose to equip the robot with artificial feelings (Fig. 6). For the related algorithms must run on the limited resources of the microcontroller embedded in the robot, our approach focuses on lean algorithms rather than modelling functions of the brain [Dietrich et al. 2009; Doeben-Henisch 2009].

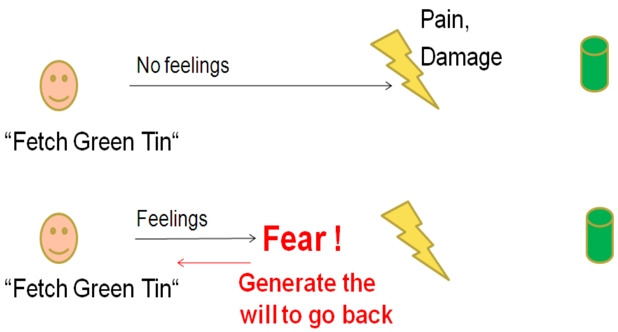


Fig. 6 An initial fragment of a cognitive map with false transitivity

The lean algorithm mirrors the way humans acquire feelings: Human feelings ("fear" and "well-being") about a situation, caused by unconscious judging information gathered from the human's task-related sensors such as eyes and ears, are enforced if their body-related sensors such as pain receptors sense bad ("pain") or good ("comfortable") feedbacks in presence of the situation. The feelings that develop when the human confronts the same situation again will result in a change of his behaviour and his will, subsequently.

Implementing this model requires to gather environmental data in the sensory vector $\underline{y}$

$$\underline{y} = \left(\underline{x}^T, \underline{d}^T, 1\right)^T \tag{1}$$

in order to calculate the feelings. It consists of the task-related sensory vector $\underline{x}$, which contains features derived from the visual and proximity data, and, if appropriate, some elements of the body-related sensory vector $\underline{d}$ for acquiring data such as temperature as well as for drives such as hunger (low battery status) and the desire for praise by the user for having achieved a goal and "social contact".

Artificial feelings with respect to a given state (or situation) are generated by assessing the sensory input. Therefore, we introduce a weighting vector $\underline{w}$ which represents the experiences made previously. Each object class and obstacle class, characterized by the object - ID - number (OID), has its own weighting vector $\underline{w}$. All weighting vectors together build the experience base. Hence, each state

$$s_j\left(\underline{y}, \underline{w}\right) \tag{2}$$

depends on the sensory input $\underline{y}$ and experiences $\underline{w}$.

In order to assess the artificial feelings about the current state, a quality of state function [Sun 2007; Stubbs and Wettergreen 2007]

$$Q\left(s_j\left(\underline{y},\underline{w}\right)\right)=\underline{y}^T\,\underline{w} \tag{3}$$

is introduced, which classifies the state either to the feeling “well-being” or to the feeling “fear”. A quality of state larger than or equal 0 corresponds to “well-being”. This will cause the robot to continue the actions necessary to achieve its goal. A quality of state value less than 0 means the robot has “fear”, e.g. of another “aggressive” robot, of standing on an inclined ramp or of low battery charge. Rather than executing the actions to meet the goal the robot develops alternative behavior strategies such as returning to its (safe) start position and stay there or retry to fulfill its mission after having waited for a while.

The experience gathering process, i.e. the training of $\underline{w}$ by forced feedback, is triggered if the body-related sensors (or a part of them) exceed a threshold which indicates either “pain” or “comfortable” and if this feedback is not in line with the feelings calculated before. They can be compared with receptors of the human body, e.g. the with pain receptors of the skin, which give a feedback about the body’s situation or with receptors which generate the feedback “hungry”. Hence, experiences are made if the feedback “pain” is sensed under the presence of the feeling “fear” or the feedback “comfortable” is sensed but the robot develops the feeling “well- being”. In either case the weighting vector $\underline{w}$ is adapted accordingly . A possible adaption approach is minimizing the amount of false classifications by the 1st gradient method and the Robbins-Monro method. This implies correcting $\underline{w}$ by means of $\underline{y}$ scaled by a factor k ($0 < k \leq 1$) in case of a false classification, i.e. a false assessment of the feeling:

$\underline{w}_{i+1}=\underline{w}_i+k\,\underline{y}$ if the feeling "fear” has been classified but the physical condition feedback is “comfortable”
$\underline{w}_{i+1}=\underline{w}_i-k\,\underline{y}$ if the feeling “well-being” has been classified but the body-related feedback is “pain”.

$\underline{w}$ is not modified in case of a right classification. It is important for the adaptation to use the sensory data $\underline{y}$ which have been acquired a short time before the adaptation has been triggered. On that condition feelings calculated allow the robot to act appropriately before a situation has deteriorated dangerously. The linear adaption of $\underline{w}$ converges if the features used in the sensory vector $\underline{y}$ are linearly independent.

In our first investigations with the assistive robot we limited the size of the sensory vector to the measurement of the distance to an object. Thus, the sensor vector contains the two elements:

y_1 = distance (in meters) and
y_2 = 1 as the offset.

In order to acquire the body-related data for experience gathering, we use a contact sensor which generates the negative feedback "pain" and a touch sensor which generates data related to “comfortable”.

For the calculation of the artificial emotions through the Quality of State function we choose initial values as follows:

w_1 = 0,8 (weighting of the distance) and
w_0 = 0,2 (weighting of the offset).

The calculation of the feelings and the comparison with the physical condition is carried out in an interval of 2 seconds in which the robot covers a distance of 20 cm roughly or at an emergency stop, e.g. if the robot has collided with an obstacle or object.

If the comparison between feelings and the physical condition requires an adaptation for the behavior shown so far has proven to be not appropriate, the adaptation is carried out with a step size of k = 0.4. As for the sensory vector used in the adaptation algorithm for calculating the new weighting vector, the second to last acquired data, i.e. prior to the initiation of the adaptation, have been applied.

The first experiments were done with a cabinet that is not registered as an obstacle in the map. At the distances 0.48 m, 0.29 m and 0.11 m the robot generated during the final approach to the cabinet the Quality of State - values 0.584, 0.432, and 0.288 (Tab. 1). For these are greater than 0 the robot develops positive feelings (Quality of State data) which are in line with the physical conditions because the contact sensors sense no “pain”. Hence, the robot moved further toward the cabinet.

Table 1 Quality of State data with initial weighting vector

	Position 1	Position 2	Position 3
Distance [m]	0.48	0.29	0.11
QoS	0.584	0.432	0.288

The inevitably crash taking place at 0 cm distance closes a contact sensors and triggers the adaptation of the weighting vector to w_1 = 0.756 and w_0 = -0.2. According to the theory described above, the second to last distance has been used in the calculation in order to generate an experience which develops “fear” prior to the collision. The experience gathering causes the following change in behavior:

When the robot performed its job the next time and recognized the closet again at the distances 0.5 m, 0.31 m and 0.09 m the Quality of State values of 0.178, 0.034, and -0.132 have been calculated (Tab. 2). The negative value at 9 cm distance to the cabinet generates a feeling of "fear", which causes the robot to stop, instead of colliding with the cabinet. Next, it computed a new trajectory to the destination area which it avoided the collision and tried to get accomplish its goal moving on the new path.

Table 2 Quality of State data with initial weighting vector

	Position 1	Position 2	Position 3
Distance [m]	0.5	0.31	0.09
QoS	0.178	0.034	-0.132

Without positive feedback so far the approach has its limits for objects that are useful for the robot and will also be avoided. E.g., it developed after one or more (accidental) collisions at a charging station "fear" and then stopped without charging the battery to execute his task. Pressing the touch sensor generated the feedback "comfortable" and triggered an adaptation to positive quality of state - of values. The robot now has no "fear" of the charging station and recharges its battery. Alternatively, a magnetic field sensor could be used as a body-related sensor to trigger kind of "motivation" to charge the battery.

7 Conclusion

In this paper the development of a mobile assistive robot with visual and auditory skills is presented. It navigates map- and sensor- based within and between rooms and recognizes objects by means of a 3-D camera that have been requested a user. The recognized object is grabbed and brought to the user. The robot is able to change its behavior on the basis of past experiences. In a first scenario, the robot avoids a cabinet after having experienced one or more collisions with it. By "motivation" it can correct disadvantageous risk avoiding behaviour such as staying away from a charging station it has collided previously.

References

[Dietrich et al. 2009] Dietrich, D., Bruckner, D., Zucker, G., Müller, B., Tmej, A.: Psychoanalytical model for automation and robotics. In: IEEE Africon 2009, Nairobi, Kenya (2009)

[Doeben-Henisch 2009] Doeben-Henisch, G.: Humanlike computational learning theory: a computational semiotics perspective. In: Proc. IEEE Africon 2009, Nairobi, Kenya (2009)

[Gonzalez and Woods 2008] Gonzalez, R., Woods, R.: Digital image processing. Prentice Hall, Englewood Cliffs (2008)

[Jin et al. 2003] Jin, T., Lee, B., Lee, J.-M.: AGV Navigation using a space and time sensor fusion of an active camera. Inter. J. of Navigation and Port Research 27(3), 273–282 (2003)

[Mertens et al. 2011] Mertens, A., Reiser, U., Brenken, B., Lüdtke, M., Hägele, M., Verl, A., Brandl, C., Schlick, C.: Assistive robots in eldercare and daily living: Automation of individual services for senior citizens. In: Jeschke, S., Liu, H., Schilberg, D. (eds.) ICIRA 2011, Part I. LNCS, vol. 7101, pp. 542–552. Springer, Heidelberg (2011)

[Nauth 2010] Nauth, P.: A method for goal understanding and self-generating will for humanoid robots. In: Proc. 3rd International Conference on Human System Interaction, Poland, pp. 344–349 (2010)

[Stubbs and Wettergreen 2007] Stubbs, K., Wettergreen, D.: Anatomy and common ground in human-robot interaction: a field study. IEEE Intelligent Systems, 42–50 (2007)

[Sun 2007] Sun, R.: Cognitive Social Simulation Incorporating Cognitive Architectures. IEEE Intelligent Systems, 33–39 (2007)

[Surmann et al. 2008] Surmann, H., Bredenfeld, A., Christaller, T., Frings, R., Petersen, U., Wisspeintner, T.: The volksbot. In: Proc. Inter. Conf. on Simulation, Modelling and Programming for Autonomous Robots, Italy, pp. 551–561 (2008)

[Toth 2009] Toth, D.: Object recognition of humanoid robots using visual sensors. Diploma Thesis, University of Applied Sciences Frankfurt a.M., Germany (2009) (in German)

Evaluation of Team-Sport Training Effort Control Systems

F.J. Parrado-García[1], P. López-Matencio[1], D. Chaves-Diéguez[2], J. Vales-Alonso[1], J.J. Alcaraz[1] and F.J. González-Castaño[2]

[1] Department of Information Technologies and Communications, Technical University of Cartagena, Spain
{fco.parrado,pablo.lopez,javier.vales,juan.alcaraz}@upct.es
[2] AtlantTIC, University of Vigo, Spain
dchaves@alumnos.uvigo.es, javier@det.uvigo.es

Abstract. This work studies different analytical systems to evaluate and control effort in team-sport training. They analyze real-time training data obtained by means of biometric belts and provide instructions to direct athletes' training. The decision techniques estimate the ratios of each effort regime, based on three different methodologies: (i) best-fit polynomial approximations, (ii) Kalman filters and (iii) sliding-window distribution estimation. The goal is to predict the future effort regimes and to provide suitable training orders to control that effort. The complete system results in a virtual coach, operating in real time and automatically. This methodology has been piloted in an experiment with the UCAM Volleyball Murcia team, top of the Spanish national women's volleyball league. Data obtained during training sessions have provided a knowledge base for the algorithms developed and allowed us to validate results.

1 Introduction

The performance and processing capabilities of Wireless Sensor Networks (WSN) are growing rapidly. In the mid-term, WSNs will be seamlessly embedded in the physical world, leading to sensitive environments enabling context-aware personalized services, which may automatically act on the behalf of the individual. This paradigm is called Ambient Intelligence (AmI) and is currently the focus of significant research. AmI relies on decision-making processes, which are usually hidden to users and give rise to smart environments. Some envisioned environments include smart homes, health monitoring, education, workspaces, sports, assisted living, and so forth. Moreover, these environments are becoming increasingly complex, making it difficult to make suitable decisions that support human activity.

This work introduces a novel integrated control system for sports training, based on supervised learning techniques, following the AmI paradigm. Commonly, high level athletes make use of wearable computing in their training. These devices monitor athletes' performance in sports like football, basketball and volleyball, and provide useful information for their coaches. Among other

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_28, © Springer International Publishing Switzerland 2014

parameters, they are able to measure heart rate, series of exercises, speeds and distances, and so forth. Often the complexity of the data collected requires the aid of specific software for analysis by a human coach at the end of the training session. Furthermore, in elite team sports training the aim is to augment athletes' aerobic performance and also to improve the quality of exercises (i.e. technical training). This requires studying athletes' movements and working with specific groups of muscles that might differ if we consider the particular role performed by each athlete within the team. In these scenarios, personalized real-time training feedback may suppose a remarkable difference with competitors.

Additionally, the schedule of matches throughout the competition season leads to variation in the team's training policies. Controlling each athlete's performed effort, alternating periods of low intensity exercise, or even rest, with periods of normal or high intensity exercise, is required to avoid unnecessary over-training. In short, training elite athletes implies monitoring and evaluating their performance in real time, and making personalized training decisions, which cannot be tackled by ordinary training systems.

Our system comprises a global decision engine whose mission is to provide personal training commands to athletes, as well as an infrastructure that enables monitoring the athletes (position, biometric data, and movement) and sensing the environment (humidity, temperature, etc.). The result is an ambient intelligence system focused on assisting human players and coaches. The advantages of such a system include the following:

- Full time availability.
- Data analysis takes place in real time. Coaches currently analyze biometric records at the end of the training.
- Management an unlimited number of athletes, compared with the limited resources of a human coach.
- Study of "hidden" variables that may influence athletes' performance. In particular, environmental data.
- Autonomy of operation after the learning phase.

The sensing infrastructure has already been introduced in previous works [Vales-Alonso et al. 2010; López-Matencio et al. 2010], together with decision engines oriented to supporting demonstrative applications in the sport domain. These works show the suitability of AmI in that domain, as well as the usefulness of the architecture proposed.

In this new work we extend our former research by developing suitable decision-making methodologies to solve the important practical problem of training for high-level team sports. The full system addresses both the effort and the technical and tactical training in the sport of volleyball [Moras 1994]. Selection of volleyball resides mainly in the need for coaches to personalize the training of individual players. In fact, each player has a different role in the team that also requires different preparation. For example, the *libero* or defensive player trains muscle groups for good control of the ball pass so that the team can run the offense. Hitters on the other hand, train specific muscle groups to improve jumping and ball strike. Therefore, the preparation of a competitive volleyball team requires a high level of control that considerably hinders the work of the coaches.

The heterogeneity of measures and the need for individual plans have been decisive factors in choosing and validating the decision-making methodologies. In particular we have implemented three control modules:

- *Effort control in customized matches*: control players' *heart rate* (HR) by numerically estimating their evolution.
- *Technical and tactical*: controls the levels of training effort and fatigue by computing the durations of activity and rest periods in repeated exercises.
- *Exercise quality*: controls the quality of movements by evaluating their intensities.

This paper is focused on the first module: aerobic control during customized matches. Our approach can be extrapolated to other elite sports (e.g. basketball or football), where technical, tactical and aerobic preparation play an important role in practice.

2 Related Works

Nowadays, the use of wearable computing is common in athletes' training. As described in [Pérez and Llana 2008], the use of chronometers, photocells, contact platforms, microphones, photo or video cameras, magnetic resonance and X-rays machines, movement sensors and, in summary, every sensing device for physical or chemical parameters, is now a standard feature of athletes' training. Many of these devices are not exclusive to elite athletes, but available to the general public (e.g. heart rate monitors). In fact, the widespread use of commodity hardware (e.g. Apple's iPhone) has also lowered the barrier to create mobile applications for sports training [Saponas et al. 2008]. To some extent, the availability of these devices is just the first step towards the appearance of contextual services and AmI environments. Sport monitoring systems now introduce real-time data collection, as well as user localization. Future developments will aim to expand the range of monitored data (environmental conditions, detailed user data) and to provide useful actions and information based on them. WSNs [Amstrong et al. 2007] represent one of the enabling technologies for such a development.

Several context-aware applications for athletes' training have already been introduced. The focus of these works is on collecting data related to specific applications. However, these data are not directly used to obtain real-time feedback but are externally analyzed by human experts. Below we outline works belonging to this group.

In MarathonNet [Pfisterer et al. 2006] a WSN monitors runners in marathon events. Sensors on runners collect data on heart rate, time, and location. These data are sent via base stations along the track to a central database, where they are analyzed off-line. Base stations can communicate with the central database by means of GPRS, WLAN, or a wired network link. In the system developed in this work, the sensor nodes have a similar functionality, but they also act as information routers. The system in [Michahelles and Schiele 2005] assists professional skiers. Using accelerometers and force-sensing resistors, skiers can obtain data about their movements and visualize them, along with video footage, once exercise has finished. SESAME [Hailes 2006] proposes the combination of two data sources, video and sensed parameters, to improve the performance of novice and elite athletes. Major competitions like the Olympic games typically pioneer the adoption of new technologies. By way of example, in [Chi 2005], the author introduces the SensorHogu systems, which attach piezoelectric force sensors to body protectors, in order to recognize valid scoring kicks in Taekwondo. The commercial product Team^{2}Pro enables the recording and study of fitness data in real time for up to 28 players.

More intelligent systems have been developed for sports scoring purposes. For example, in [Spelmezan and Bolchers 2008], the authors present a sensor system intended to provide immediate feedback to alert users of incorrect movements and body positions. The prototype uses sensors attached to the human body and inserted into boots for the purpose of detecting mistakes during snowboarding. In [Arvind and Bates 2008], the authors develop a score system for the golf swing motion. Golfer body motion is captured by a wireless network of inertial sensors that extract snapshots of the orientation data of the user's body during the golf swing. The orientation information is compared with correct motion rules, and a score of the exercise is computed. Paper [Ghasemzadeh et al. 2009] also develops a feedback system, based on sensors which capture the movements of a golf swing. The swing motion is pre-processed locally and sent to a base station, e.g. a PDA, for further analysis. Swing motion quality is expressed as the amount of deviations from the target line and computed by linear discriminant analysis.

None of the works outlined above consider ambient conditions and thus does not incorporate devices embedded in the environment. In previous works [Vales-Alonso et al. 2010; López-Matencio et al. 2010], we showed the usefulness of the ambient intelligence paradigm in sport. The goal in both papers consists in selecting, for a given user, suitable tracks where the athletes HR will be in the desired HR range most of the time. In [Vales-Alonso et al. 2010], decision-making is based on (m,s)-splines interpolation of the HR signal, whereas in [López-Matencio et al. 2010] a *k*-NN classification engine was developed. Moreover, [Vales-Alonso et al. 2010] demonstrates the feasibility of the communication and location system architecture by means of tests performed in real outdoor training scenarios.

Finally, in [Vales-Alonso et al. 2013] we introduced a preliminary system to control training system for an elite volleyball team. The approach taken is based mainly on specialized training studies [Moras 1994; Zintl 1991]. In [Moras 1994], the author established six levels of volleyball training, ranging from global body development and enforcement, not specific to volleyball practice (levels 0-2), to actions focused on technical and tactical training (levels 3-5) in volleyball. In short, lower levels address global physical preparation, whereas higher levels are more centered on improving specific player movements. On the other hand, [Zintl 1991] presents a classification that enables the categorization of all exercise training. The author distinguishes between *continuous*, performing a constant effort over a period, *intervallic*, alternating periods of activity and rest, *repetitional*, making a series of similar actions coupled with complete rest, and *customized match*, playing a match whose intensity is controlled by the coach. In the present study we complete this previous work by introducing and evaluating new analytical approaches to predict and control de effort.

3 System Architecture

Figure 1 shows the architecture of our training system approach. The sensing and communications infrastructure of the system is based on a WSN developed in a former study [Vales-Alonso et al. 2010]. We use this infrastructure to collect two types of data:

- A set of static data, including the position of the sensors, the height of the net in volleyball, the programmed training routine, as well as players' static data, such as performance profile, age, ability, or training objectives.
- A set of real-time data of ambient parameters (temperature and humidity) and players' data (HR, minutes of training).

In sports such as volleyball, the three most important parameters to track are heart rate, levels of lactic acid accumulation, and maximum oxygen consumption (VO_2 max) [Sharma 2012]. Although the measurement of lactate is the most accurate method, it has the drawback of being invasive and can therefore only be performed from time to time. Thus, in order to measure physical effort during practice, we have chosen HR biometric (directly correlated with VO_2 max) and environmental data (temperature and humidity).

To perform such measurements, a WSN is deployed around the perimeter of the training field (see Fig. 1). Its function is to collect ambient data and to connect the user and the decision-making node (control node or CN). Similarly to the prototype developed in [Vales-Alonso et al. 2010], this *ambient* network has been implemented using standard WSN hardware (MICAz and Imote2 motes). Nevertheless, this prototype has been adapted to the particular sensing and communications needs of our project. Specifically, the player's equipment consists of two devices:

- *Zephyr BioHarness is a belt-shaped* biometric sensing device. The device senses temperature, HR, body acceleration and oxygen consumption and communicates these data through an incorporated Bluetooth interface. The player carries this device fastened to his/her body.

- *Imote2* provides the player with communication and sensing capabilities. It includes a wireless transceiver to communicate with the network, and sensing elements to monitor players' biometry. The Imote2 also includes an acoustic interface to deliver training commands and a Bluetooth interface to collect BioHarness data. The device can be placed on the training field, close to the player or as with the BioHarness, can be carried by the player. Each player has his/her own Imote2, which also responds to the localization messages from the control node.

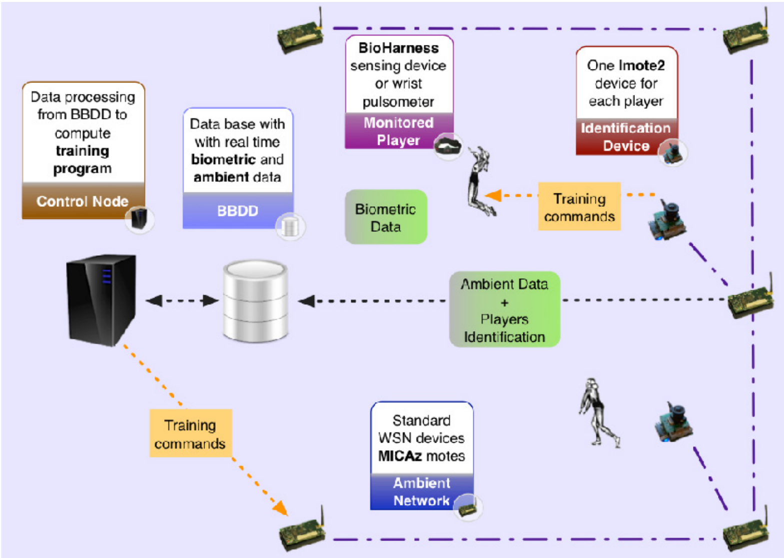


Fig. 1 Global system scheme

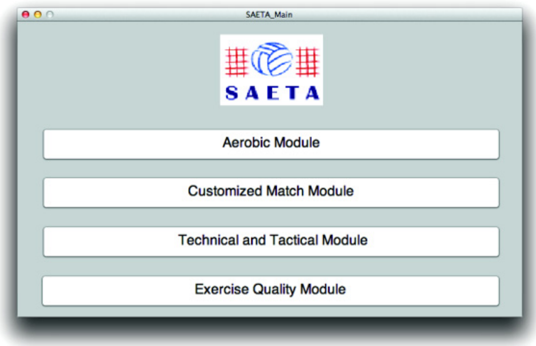


Fig. 2 SAETA main window

Moreover, the CN includes the intelligence of the whole system. It coordinates sensing and communication devices, collects data from all sensing devices, generates real-time training decisions and transmits them to the player. The interface and modules of our control system have been developed with Matlab. This interface provides easy access to configure all training functionalities (Fig. 2). Each of the training options allows the coach to set each player's initial parameters through individual menus.

4 Control in Customized Match

This section describes the customized match module of our training control system. Our system reacts to changing training conditions (of the user and the environment) in real time and provides training decisions for coaches and players. These decisions are computed by the CN (see Fig. 1). The CN is able to infer intelligent decisions by using a knowledge base of data from previous training sessions.

Customized matches are standard competition matches, where the coach may vary game conditions to address a specific training objective. For example, a common goal in customized matches is to increase aerobic capacity by extending playing time and by reducing activity. The customized match module aims to complete a training program by controlling players' HR but by using an additional methodology to the one described in our previous works [Vales-Alonso et al. 2010] and [Vales-Alonso et al. 2012]. The new module utilizes the numerical approximation of HR regimes to infer future values. Then, if the estimated HR is low, the control indicates to increase activity, and vice versa.

Since our aim is to control the player's HR, we first need to characterize his/her performance. According to the type of training activity, HR performance can be classified into different intensity levels or classes (from Low to VMAX). The HR in each class is in a percentage range of the recommended HR_{max}.

1. Low activity: $<50\%$.
2. Moderate activity (Maintenance/Warm up): $[50\%,60\%]$.
3. Weight Control (Fitness/Fat burn): $[60\%,70\%]$.
4. Aerobic (Cardio training/Endurance): $[70\%,80\%]$.
5. Anaerobic (Hardcore training): $[80\%,90\%]$.
6. VMAX (Maximum effort): $>90\%$.

An accurate formula to compute the maximal HR for a healthy person follows the relationship $\mathrm{HR}_{max} = 208 - 0.7 \times age$.

4.1 Decision-Making Process

For each player under control the coach selects total activity time and a training profile (TP). The TP is represented by a vector $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_6)$, where each position corresponds to an HR class, from low activity (ρ_1) to VMAX (ρ_6). Each position of ρ represents the ratio of training time that the player should accomplish in the corresponding HR class during the total activity time. Thus, the coach inputs TP values such that $\sum_{i=1}^{6} \rho_i = 1$. An example of this module input window is presented in Fig. 3.

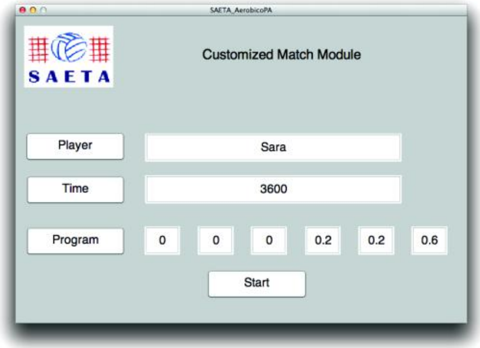


Fig. 3 Input window for customized match module

Once the total activity time and TP are selected, the control module computes HR ranges (using the athlete's profile) and starts sampling the player's HR, classifying each sample in its range. Thus, after the j-th sample the module now has the ratios in every HR range observed so far. These ratio values are utilized to compute predictions on the future evolution of the players' effort. Next, we discuss three different methods used for this task:

1. Best-fit polynomial approximation
2. Kalman filters
3. Sliding-Window distribution estimation

4.2 *Best-Fit Polynomial Approximation*

The first approach was already discussed in [Vales-Alonso et al. 2013]. It consists of using individual polynomials for each ratio ($p_i, i = 1,\ldots,6$), providing best-fit curves for the observed curves. To compute these polynomials, we have selected the least square method, which requires at least a number of points greater than the degree of the polynomial m. In our implementation, after the j-th $(j > m)$ sample the algorithm utilizes the last $n > m$ samples, n being a parameter of our module program. With this estimation, the CN computes the difference between the selected TP and the estimated TP at each instant j, following the expression:

$$\Delta[j] = \sum_{i=1}^{6} \theta_i \left(\rho_i - p_i[j + \mathrm{STEP}]\right) \tag{1}$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_6)$ are weighting factors of the difference between the approximated values and those required by the coach. Besides, STEP is an algorithm parameter that represents the estimation horizon in seconds (120 seconds in our tests). Thus, $p_i[j + \mathrm{STEP}]$ is the value of polynomial i at instant j +STEP. For the conformity tests we have considered $\boldsymbol{\theta} = (-1, -1, -1, 2, 2, 4)$. These weights balance the ratio of time that a player must train in each HR regime and can be modified, according to the athlete or coach's needs to fit training needs. Should the player's training decrease, these values will lead to negative Δ 's, or positive in the contrary. Therefore, the output of this decision-making process indicates to players and coach whether the intensity of training must increase or decrease (respectively, positive or negative $\Delta[j]$ values). In addition, we have considered the importance of underlining the difference between estimated and required TP for the VMAX regime ($\theta_6 = 4$), due to the difficulty of achieving that exercise level. This weight stresses the trend of increasing training intensity if required to fulfill the selected TP.

4.3 Kalman Filters

The second alternative that we have considered is the use of a predictor based on Kalman filters, which allows us to solve, in a computationally efficient and recursively manner, least-squares estimation problems. These problems consist on estimating parameters by minimizing the squared discrepancies between observed data (the effort ratios in our work), on the one hand, and their expected values on the other.

The algorithm finds the least squares best estimate by going through data one by one. The filter gets the new best estimate for a given set of data once a new measurement is added by using, the new measurement, the old estimate, and some measure of confidence in the old estimate. As a result it provides an estimate of the state each time a new measurement is included. In order to define our Kalman filter, we must set the *state vector*. In our case it contains the ρ vector values and its first derivative (velocity):

$$x = \left(\rho_1, \rho_2, \rho_3, \rho_4, \rho_5, \rho_6, \dot{\rho}_1, \dot{\rho}_2, \dot{\rho}_3, \dot{\rho}_4, \dot{\rho}_5, \dot{\rho}_6\right) \tag{2}$$

Then, the Kalman estimation proceeds iteratively with the next two steps (see next Figure), executed at each time instant *j*. Namely, a prediction and correction process called Kalman filter cycle.

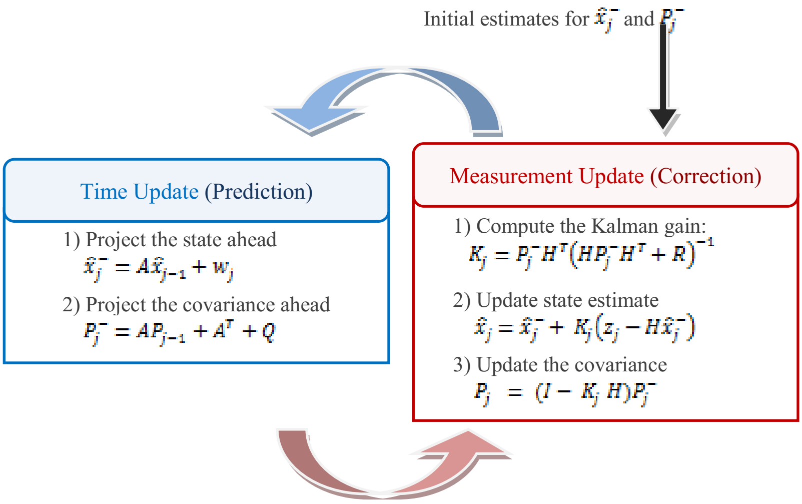


Fig. 4 Kalman filter cycle

1. *Prediction step*. The algorithm determines a state estimate $\hat{x}_j^-$ and an estimate for the error covariance P_j^- which provides an indication of the uncertainty associated with the current state estimate. The state-evolution equation is:

$$x_j^- = A\hat{x}_{j-1} + w_j \tag{3}$$

where

$$A = \begin{pmatrix} I_6 & \Delta t I_6 \\ 0 & I_6 \end{pmatrix}$$

)

is the state transition matrix, and w_j represents a normal distributed process noise sample with covariance matrix *Q*. Besides, in our case we have set the initial value $P = I_{12}$. In our model, empirical value Q = 0 gave us the best results.

1. *Correction step*. It provides a feedback by incorporating a new measurement value into the *a priori* estimate to get an improved *a posteriori* estimate. In this step the expected state $\hat{x}_j^-$ is corrected by following the equations provided in

Fig. 4. The main idea is relating the current state to a measurement z_j through a lineal model with a matrix H, and a normal distributed process noise sample v_j with covariance *R*. This is:

$$z_j = \boldsymbol{H}\hat{x}_j^- + v_j \tag{5}$$

where

$$H = \left(I_6 \,|\, 0_{6x6} \right) \tag{6}$$

Afterwards the model corrects the estimation $\hat{x}_j^-$ by means of this measurement z_j. The empirical value $R = I_6\,0.1$ gave us the best results. Besides, the Kalman gain *K* denotes that while the measurement error covariance *R* gets closer to zero, the measurement z_j is more trusted, while the predicted measurement is less trusted.

Therefore, we can estimate future ρ values if a set of previous measures is known (such us our HR ratio measurements). This estimation is updated every time a new measurement is obtained. To infer ratios with a prediction horizon of STEP stages, the Kalman prediction must be applied STEP times. Note that each estimated ρ value obtained by Kalman filter must be normalized. Moreover, as in the previous case of the polynomial approximations, once the expected ratio is computed, the difference between the selected TP and the estimated TP is obtained using expression (1). The same weighting coefficients have been used in our evaluation tests to correct the training offset.

4.4 Liding-Window Distribution Estimation

Finally, in order to estimate the ratios of each effort regime for an athlete during a training session, a sliding-window distribution estimation method was also tested. The foundation of this last method is to assume that, in absence of external stimulus, the athlete will continue to perform with similar effort in the near future as he or she has done in the recent past.

To do this, let us assume that $r_i[j]$ is a discrete sequence of the values zero and one, where one means that the heart rate of the athlete sampled at time *j* lies within the *i*-th HR defined range, and zero means that the heart rate is outside this range. At any given time *j*, the fraction of time an athlete is within each HR range is simply calculated as the fraction of accumulated HR samples in each range over the total of HR samples:

$$\rho_i[j] = \frac{\sum_{\tau=1}^{j} r_i[\tau]}{\sum_{i'=1}^{6}\sum_{\tau=1}^{j} r_{i'}[\tau]} \quad (5)$$

If we define the accumulated value of HR samples within each range as $HR_i[j] = \sum_{\tau=1}^{j} r_i[\tau]$, then eq. (5) can be simplified as shown in the next equation.

$$\rho_i[j] = \frac{HR_i[j]}{\sum_{i'=1}^{6} HR_{i'}[j]} \quad (6)$$

After a number of new samples (defined by STEP in our case), the updated ρ_i values must be between the ranges specified in the next equation. This is, between zero and STEP new values have been added to the *i*-th HR range, and exactly STEP new samples have been added to the total.

$$\frac{HR_i[j]}{\mathrm{STEP} + \sum_{i'=1}^{6} HR_{i'}[j]} \leq \rho_i[j + \mathrm{STEP}] \leq \frac{\mathrm{STEP} + HR_i[j]}{\mathrm{STEP} + \sum_{i'=1}^{6} HR_{i'}[j]} \quad (7)$$

None of the external factors that can affect athletes' heart rate will be taken into account for the forecast method described in this section so, as stated before, we will assume that the athlete will keep a similar intensity in the near future exercise performance as he or she has kept in the recent past. This is, if we look at the most recent *n* samples of HR, the prediction made by this method will be to forecast that the next STEP samples will be classified following the same distribution or, to put it in another words, if $\Delta HR_i^n[j] = HR_i[j] - HR_i[j-n]$ new samples have been added to the *i*-th HR range within the last *n* samples, each one of the next STEP samples has an independent probability of $\frac{\Delta HR_i^n[j]}{n}$ of being classified within the *i*-th HR range.

Taking this approximation, and working only with the expected value for the random variables involved, the ratios for each HR range can be expressed as shown in the following equation:

$$\tilde{\rho}_i[j] = \frac{HR_i[j] + \frac{\mathrm{STEP}}{n}\Delta HR_i^n[j]}{\mathrm{STEP} + \sum_{i'=1}^{6} HR_{i'}[j]} \quad (8)$$

This formulation of the predicted ρ_i ratios can be updated with each new sampled HR value, and also ensures that the sum of all predicted $\rho_i[j]$ for any given j is always one, as in the Kalman filter case. Finally, the difference between the expected and the required TP are obtained also with expression (1).

4.5 Validation

We have tested the methods discussed for this module in a real customized match training session of the UCAM volleyball team. The training intends to increase the effort in relation to a standard match. To that end, the coach eliminated serves and idle periods and immediately passed the ball when the training was interrupted. During training the players' biometrics were monitored in real time using the BioHarness system. An example of the samples collected from two different players is represented in Fig. 5. We can observe how the different roles of the players in a match can produce different HR profiles. Specifically, in our tests the physical effort of a hitter produces higher HR frequency samples than that of a setter. Hitters therefore require different TPs and indications. By using the customized match module, the coach is able to adapt training to each player.

In order to validate our methodology, we used an external feedback signal to the player's HR, which is controlled by the program module output. A rise in training intensity points to an increase in the player's HR modulated by the output of the program. This process is similar to the result obtained when a real player receives a command to increase training. The same procedure is followed when the decision is to decrease training intensity.

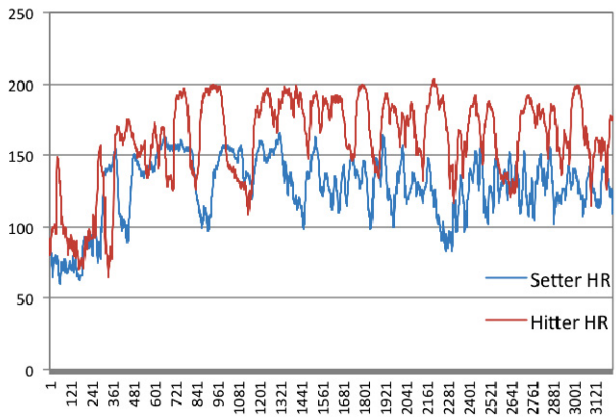


Fig. 5 Real-time HR samples during a customized match training

Thus, we modulate the HR of a player by adding a feedback signal γ. The computation is done as follows:

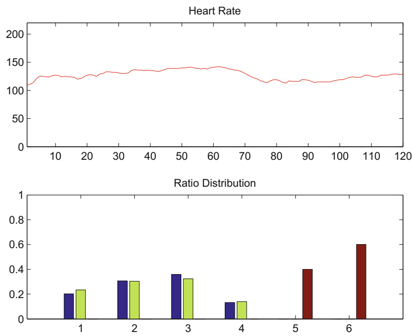
$$hr'(j) = hr(j) + \gamma(j) \tag{9}$$

where γ is the output of the single pole IIR system,

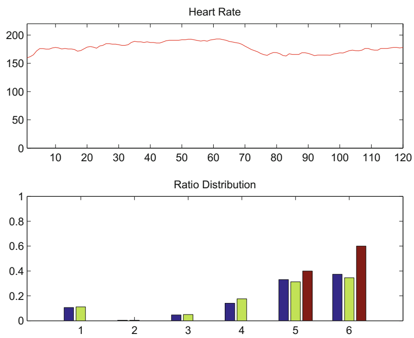
$$\gamma(j) = a\gamma(j-1) + \Delta[j] \tag{10}$$

with the weight a adjust the strength of the feedback signal. In our validation tests we have selected $a = 0.99$.

Finally, we computed the difference between the TP required by the coach and the real training performed. This difference is computed in two forms: with and without HR correction (i.e. without using the control module). Our system will be validated if the use of the feedback signal brings the real TP closer to the TP configured by the coach. The difference between both TPs shows the goodness of fit of our methodology.



(a) No feedback



(b) Feedback

Fig. 6 Real samples of hitter heart rate and their distribution in intensity classes under training time objective ρ = (0,0,0,0,0.4,0.6)

We have tested our customized match module with two players: a setter and a hitter and two different effort distributions (ρ). Fig. 6 shows an example of the output (see [Vales-Alonso et al. 2013] for the full set of results). For each HR class, the following three data ratios are shown:

- The blue bar represents the estimated HR values computed with p_i polynomials and STEP = 120.
- The green bar represents the current HR of the player
- The brown bar represents the required ratio.

In all cases the results show how the difference between objective TP (brown bar) and real training obtained (green bar) is greatly reduced by using the feedback signal. Clearly, introducing a feedback during the training session leads to a better match with the selected TP.

If the other methods (the Kalman filters or the Sliding-Window distribution estimation) are used, similar trends have been observed (but with a different final distribution). Next we compare the performance of the three methods.

Fig. 7 shows comparatively the error (measured as norm 2, i.e., the root square of the squared difference between the expected and the real ratio) in the prediction achieved by the three methods for a complete training session of 1 hour (from 1 to 3600 s.). The curves depicted show the predicted ratio with a horizon of STEP=180 seconds. Note that predictions starts in 300 seconds since we require collecting enough data to start predictions, and we have set this time as 300 seconds in our experiments. Moreover, the Sliding-Window method also uses a window of n = 300 samples. Clearly the polynomial approximation has as side effect large oscillations, as can be expected. Comparatively, both the Kalman and the Sliding-Window approximations have a better performance without oscillations. Although in some cases the Kalman filter prediction surpasses in quality the Sliding-Window method (e.g. for ρ_4 around the 600 seconds), the common trend is that the Sliding-Window method achieves less prediction error for that horizon STEP.

Another important issue is how STEP should be selected. On the one hand, if STEP is too small the predictions will be better but rather useless, since only real data is used to decide the training orders (the feedback). For example, if STEP is null this is similar to a coach that has been collecting statistics of a player during training and decides the next training based on that statistics. What our system allows is to infer what will happen in the future and add this information to the decision making process. On the other hand, when STEP is too large the prediction process will become unreliable, due to the dynamic evolution of the HR regime ratios. Therefore, the feedback obtained is likely to be incorrect.

Summarizing, there must be a balance in the STEP value. An experiment has been performed to show this. Figs. 8 and 9 depict the final Δ value (i.e. Δ at 3600 seconds instant) computed trough expression (1) for both players and both example TPs. As being stated previously, this value indicates the quality of the training, being better for values closer to 0. As shown in Figs. 6 even using feedback is difficult to reach the most intensive HR regimes since the TP is too homogeneous, and therefore the final Δ tends to be positive. Besides, for the second TP, and since it concentrates in the hardest regimes, the feedback is more effective and the final results are closer to zero and even negative, due to some over-effort.

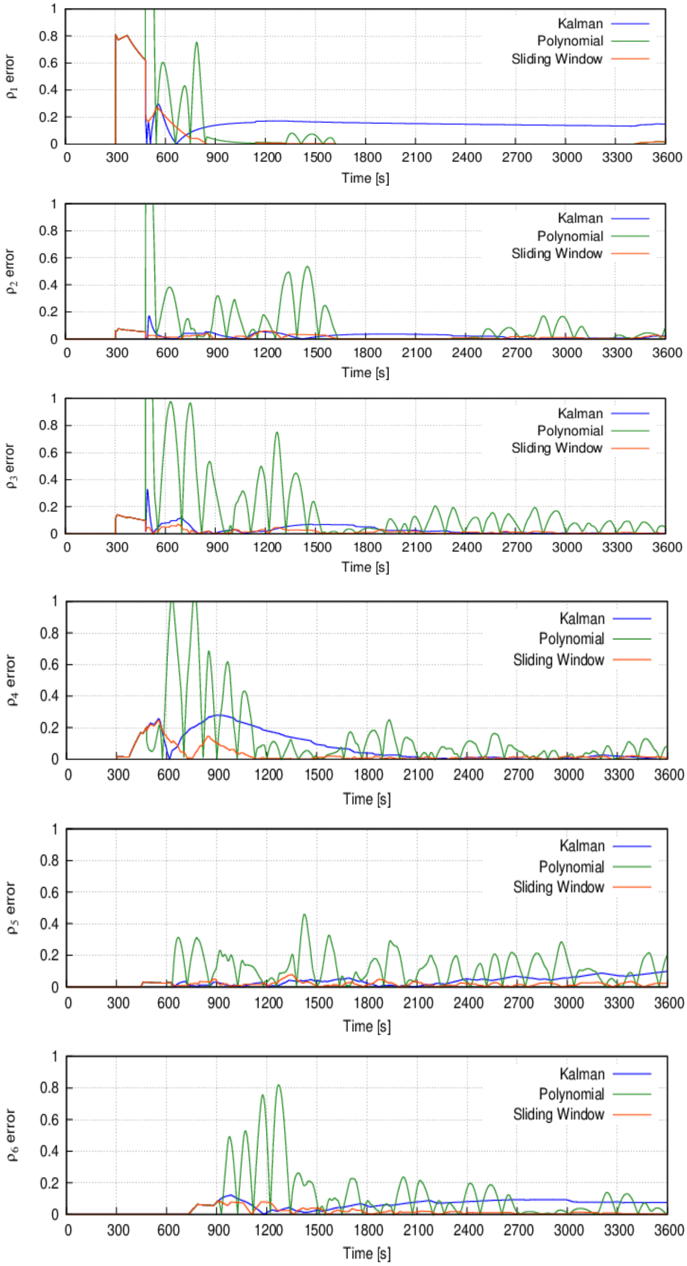


Fig. 7 Estimation error

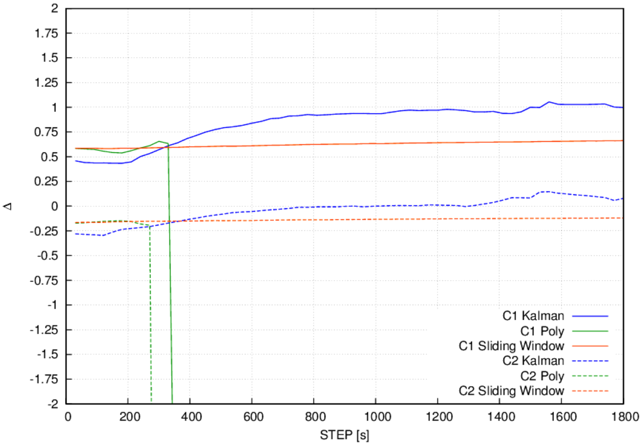


Fig. 8 Final Δ versus STEP value, player 1

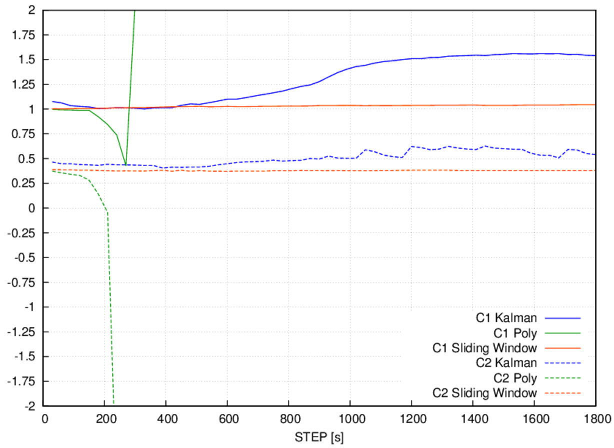


Fig. 9 Final Δ *versus* STEP value, player 2

As shown in the Figs. 8 and 9, even though the polynomial approximation achieves a reasonable performance for STEP values below 300 seconds, beyond this point this method becomes unusable, with a stepper performance drop. For the player 1, the setter, the best training for the first TP is obtained for STEP values in the order of 150 seconds, using the Kalman filter. In the case of considering the second TP, the best training is obtained again with the Kalman filter for broad values of STEP around 1000 seconds. In the case of the second player, for both TPs the best method is the polynomial approximation, for 300 and 200 seconds respectively. However, we must note that *a priori* is not possible to known the best step, and that in this case the Sliding-Window methods is safer, since, as show in these Figs., the response is almost independent from the STEP parameter.

As conclusions, if a careful preliminary study can be performed for a given player to obtain a suitable prediction methodology and parameterization, benefits can be obtained in the fine-tuning of the athletes' training. Otherwise, both the Kalman and Sliding-Window methods are worth being considered. A trend has been observed in the results where Kalman filter obtains similar or better results than the Sliding-Window technique for STEP values around 300 seconds.

5 Conclusions and Future Works

We have developed a novel training system for an elite volleyball team based on the ambient intelligence paradigm. This system monitors and controls each team player and provides useful feedback for coaches in real time. Training decisions are based on the real-time and historical data of each player which are then used to predict the effort ratios based on three different methodologies. Validation tests performed with an elite volleyball team have demonstrated the utility of our approach.

Acknowledgements. The authors gratefully acknowledge the contribution of Pascual Saurín Cutillas and Alberto Aguilar Zarco, coaches to the UCAM volleyball Murcia team for their comments and decisive collaboration.

References

[Armstrong et al. 2007] Armstrong, S., Jovanov, E., Kerwin, D.: Wireless connectivity for health and sports monitoring: a review. British Medical Journal 41(5), 285 (2007)

[Arvind and Bates 2008] Arvind, D., Bates, A.: The speckled golfer. In: Proc. 3rd International Conference on Body Area Networks. Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, no. 20 (2008)

[Chi 2005] Chi, E.: Introducing wearable force sensors in martial arts. IEEE Pervasive Computing 4(3), 47–53 (2005)

[Ghasemzadeh et al. 2009] Ghasemzadeh, H., Loseu, V., Guenterberg, E., Jafari, R.: Sport training using body sensor networks: A statistical approach to measure wrist rotation for golf swing. In: Proc. 4th Inter. Conf. on Body Area Networks. Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, no. 2 (2009)

[Hailes 2006] Hailes, S.: The sensing for sport and managed exercise project (SESAME) (2006), http://sesame-wiki.cl.cam.ac.uk/twiki/bin/view/Sesame/WebHome/ (accesed November 7, 2013)

[Lopez-Matencio et al. 2010] López-Matencio, P., Vales-Alonso, J., González-Castaño, F.J., Sieiro, J.L., Alcaraz, J.J.: Ambient intelligence assistant for running sports based on k-NN classifiers. In: Proc. 3rd Conference on Human System Interactions, pp. 605–611 (2010)

[Michahelles and Schiele 2005] Michahelles, F., Schiele, B.: Sensing and monitoring professional skiers. Pervasive Computing 4(3), 40–45 (2005)

[Moras (1994)] Moras, G.: Complete volleyball training 1000 exercises and games. Paidotribo (1994) (in Spanish)

[Pérez and Llana 2008] Pérez, P., Llana, S.: Instrumentation in sports biomechanics. Journal of Human Sport and Exercise-University of Alicante 2(2), 26–41 (2008)

[Pfisterer et al. 2006] Pfisterer, D., Lipphardt, M., Buschmann, C., Hellbrueck, H., Fischer, S., Sauselin, J.H.: MarathonNet: Adding value to large scale sport events-a connectivity analysis. In: Proc. 1st Inter. Conf. on Integrated Internet Ad Hoc and Sensor Networks, no. 12 (2006)

[Saponas et al. 2008] Saponas, T., Lester, J., Froehlich, J., Fogarty, J., Landay, J.: iLearn on the iPhone: Real-time human activity classification on commodity mobile phones. University of Washington CSE Tech. Report UW-CSE-08-04-02 (2008)

[Sharma A 2012] Sharma, A.: Physical and physiological profiles of different levels of net ball players. PhD dissertation. Bundelkhand University, India (2012)

[Spelmezan and Borchers 2008] Spelmezan, D., Borchers, J.: Real-time snowboard training system. In: CHI 2008 Extended Abstracts on Human Factors in Computing Systems, pp. 3327–3332 (2008)

[Vales-Alonso et al. 2010] Vales-Alonso, J., López-Matencio, P., Gonzalez-Castaño, N.-H.H., Baños-Guirao, P.J., Pérez-Martínez, F.J., Martínez-Álvarez, R.P., González-Jiménez, D., Gil-Castiñeira, F., Duro-Fernández, R.: Ambient intelligence systems for personalized sport training. Sensors 10(3), 2359–2385 (2010)

[Vales-Alonso et al. 2013] Vales-Alonso, J., López-Matencio, P., Veiga-Gontan, J., Banos Guirao, P., Alcaraz, J.: An effort control system for training elite team-sport athletes. In: Proc. 6th Inter. Conf. on Human System Interaction, pp. 279–286 (2013)

[Zintl 1991] Zintl, F.: Endurance Training: Foundation, Methods and Training Management. Ediciones Martínez Roca, Spain (1991) (in Spanish)

Brain Monitoring to Detect Nationality Difference Induced by Robot Gesture

S. Suzuki[1], Y. Fujimoto[2], and T. Yamaguchi[2]

[1] Department of Robotics and Mechatronics, Tokyo Denki University, Tokyo, Japan
ssuzuki@fr.dendai.ac.jp
[2] Faculty of System Design, Tokyo Metropolitan University, Tokyo, Japan
{yamachan,yfujimoto}@tmu.ac.jp

Abstract. This paper reports an experimental study to confirm whether robot's gesture can induce country-by-country different response in human observers according to their nationalities. Robot greeting actions in Japanese, Chinese, and French styles were mimicked using the communication robot ApriPoco. Brain activity in the cortex area associated with a mirror neuron was investigated by the near-infrared spectroscopy (NIRS) to evaluate the country-by-country difference objectively. As a result, it was confirmed that nationality difference between Japanese and Chinese participants appeared in means of the sum of cerebral blood flow change (SBC) in a primary motor cortex (M1) and Brodmann area 44 (BA44). Moreover, it was shown that robot gesture enhanced the observer's SBC in case of Japanese participants who watches the mother language-style greeting and Chinese participants' SBC were emphasized by greeting gesture in above-mentioned three types of language styles.

1 Introduction

The service industry holds an important position in the economy, and a ratio of the service in GDP in advanced countries reaches 70 percent. With this international background, *Service Engineering*, that is a new discipline to develop service scientifically by utilizing engineering, psychology, information-technology, and economics, was presented in 2002 from Japan [Tomiyama 2002]. Also from business community, SSME (Service Science, Management, Engineering) was proposed by IBM as the next frontier in innovation. One of significant issues about the service industry is how to consider the country-by-country (or culture) difference among customers, because same service might give different satisfaction to different customers depending on their nationality [Lovelock 1995]. Service itself is, however, difficult to be dealt with due to its ambiguous characteristics such as intangibility, simultaneity, heterogeneity, and perishability [Rugimbana 2003]. Therefore an analysis of the country-by-country impression by utilizing an artificial system is one of effective methods to evaluate some service action, since such system can execute same task with high repeatability. In order to establish a

Z.S. Hippe et al. (eds.), *Human-Computer Systems Interaction: Backgrounds and Applications 3*, Advances in Intelligent Systems and Computing 300,
DOI: 10.1007/978-3-319-08491-6_29,

measurement method of country-by-country difference of persons, however, the following questions should be made clear:

- Q1. Can country-by-country differences of persons be detected by a technological measurement method?
- Q2. Can difference be detected by the technological measurement when a person feels a nationality difference against other person's behavior?
- Q3. Can artificial system induce different response according to person's nationality?

Questions Q1 and Q2 concern the measurement of the service, and question Q3 corresponds to the control of the service. The present authors performed a preliminary experiment to obtain an answer to Q1, as reported in previous study [Suzuki 2011]. In that study, greeting actions of a human-like robot were utilized since the greeting tends to reflect the nationality of persons. Possibility of detection of the country-by-country difference was confirmed in that study by the brain monitoring of the participant who observed the robot greeting action. The reliability of the obtained result was, however, not sufficient due to a small number of participants. Therefore, in the present study, the number of participants was increased, the experimental protocol was improved, and statistical analyses were performed in order to obtain answers to the above-mentioned questions.

The latter sections are organized as follows: Section 2 introduces the related research topics which are required to determine the experimental protocol for investigation of the country-by-country difference. Section 3 explains experimental settings and tasks. Section 4 shows the measured data and discusses an evaluation method for the following analyses. Statistical analyses are mentioned in Section 5, and the obtained knowledge is summarized in Section 6.

2 Related Research Topics

Greeting and Country-by-Country Difference: Reception service is one of preliminary service activities to provide main service and it is an intervening variable of service quality [Kondo 1999]. The greeting actions play a significant role for starting of the reception service. Also from view point of the country-by-country difference, greeting is a significant factor because it differs depending on kinds of language and gesture.

Since greeting often involves gesture, many studies utilize the gesture communication, especially in the robot research field. A study considering the national difference in the gesture is, however, rarely found. As a similar study focusing on difference of gesture, the investigative studies on evaluation of human sense to the robot behavior were reported [Broadbent 2007]; however, they assumed only one culture of the home country. Further these studies used subjective investigation such as questionnaire or the SD method, that is, they were not performed based on objective measurement data.

Brain Monitoring: Then how should we measure the country-by-country difference objectively? The difference is recognized in principle through a cognitive processing in a human brain; hence, the present authors thought it was natural to use a brain monitoring method. Although non-invasive brain monitoring methods such as Electroencephalogram (EEG), functional Magnetic Resonance Imaging (fMRI), Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT), and Near Infra-Red Spectroscopy (NIRS) have been used, these methods except EEG and NIRS cannot allow the participant to move the body freely since his/her head has to be steadied inside the huge measurement device. Moreover, it is difficult to use EEG when other electrical devices like a computer and a mechatronics gadget exist around the participant, because EEG is weak biological electric signal and easily influenced from other electrical signal sources. On the other hand, the NIRS system measures changes in the concentration of oxy- and deoxy-hemoglobin using different spectra lasers in the near-infrared range. Thanks to the measurement principle, NIRS has been often used to investigate the brain activation of natural behavior in a nonrestrictive environment [Miyai 2002]. These are reasons why we adopted NIRS for the present study.

Mirror Neuron and a Motion of Greeting: The mirror neuron fires both when an animal acts and when the animal observes the same body motion performed by another. The mirror neuron was found in the premotor cortex of the macaque monkey at first, and several studies showed that similar type of neuron existed in a human brain. Recent various researches using the functional brain imaging methods demonstrated that superior temporal sulcus (STS) was activated when a human watched motion, face, and emotion of others [Rizzolatti 2004]. Especially, it had been reported that area of Broca (i.e., Brodmann area 44: BA44), an inferior parietal lobule (BA39-40) [Iacobonil 1999], and a primary motor cortex (M1) [Fadiga 1995] tend to activate as a mirror neuron during observation of other's behavior. Some researchers claimed that mirror neuron relates to non-verbal communication, sympathy, and language system besides recognition of behavior of others [Rizzolatti 1998]. Based on this scientific knowledge, it can be expected to detect the country-by-country difference of greeting as a difference of the mirror neuron's activation since greeting is verbal communication as well as non-verbal one. It is, however, difficult to measure the STS area due to the measurement limit of the NIRS system. Hence, in this study, a primary motor cortex (M1) and a Brodmann area 44 (BA44) were measured as a region of interest.

3 Method

In this experiment, ApriPoco (Toshiba Cooperation, Japan) was used to perform greetings. ApriPoco has seven movable body parts (the head, the arm, and the waist) and an utterance function. The shape resembles a penguin, and the greeting gesture was generated by the hand and head motion.

Brain blood flow was measured by the NIRS system, ETG-4000 (Hitachi Medical Corporation, Japan). In order to cover the wide area including M1 and BA44, the region covering the frontal cortex around Fz, which was determined by an international 10-20 system, was monitored. Data measured from four channels corresponding to M1 and from two channels near the right and left BA44 were investigated in the following analysis. Positional relationship of M1 and BA44 is shown in Figs. 1(a) and 1(b).

The experimental scene is shown in Fig. 1(c). The participant sat on a chair in front of the robot and the monitor. To avoid an induction of recrementitious cognitive activity caused by an environmental noise, the participant wore an active noise cancelling headphone. The participant watched the greeting motion by listing the voice through the headphones. Further, to eliminate visual disturbance, the experimental space was covered with curtain and separated from the outside.

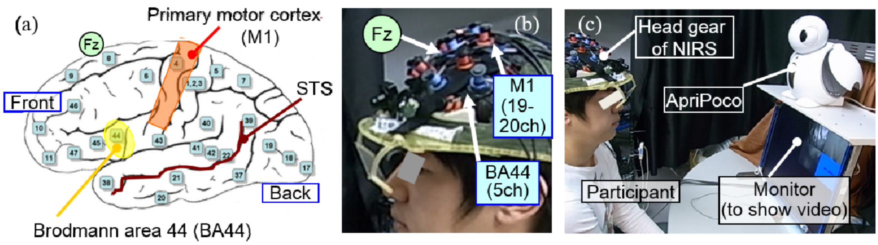


Fig. 1 (a) Brain local area associated with a mirror neuron, (b) NIRS measurement channels, and (c) Experimental set-up

Japanese and Chinese peoples were selected as participants for the experiment. French was chosen as a non-native language against both peoples. Another reason of this selection is that French-style greeting might induce difference since French is classified into 'contact culture' unlike Japanese and Chinese that are 'noncontact culture' [Michael 1970]. Therefore, three types of greeting behavior, that were mimicked to Japanese, Chinese, and French styles, were programmed into ApriPoco using the function of utterance and motion control. Illustration of each greeting motion and the serial photographs of these greeting styles are shown in Fig. 2. Japanese-style is 'bowing'; the robot bows the head and both hands move to the front of body. Chinese-style greeting was expressed as a motion bringing the arm up slowly to the head height. In case of French-style greeting, the similar hand motion was repeated twice like a hand-shake to give a friendly impression to express 'contact culture'.

Moreover, videos of greeting by humans were prepared to investigate a difference of impressions to human greeting and robot greeting. In the video, Japanese, Chinese, and French young male say a word of greeting in each country style without gesture: "Kon-nitiwa", "Ni Hao", and "Bonjour". These videos were projected to the monitor in front of the participant at the experiment.

Greeting behaviors in the experiment were prepared by combination as {*robot with gesture*, *robot with no gesture* (only voice), *human* (movie) } × {Japanese, Chinese, French}-styles. For simplicity, the former conditions are abbreviated as *robot+gesture*, *robot+nogesture*, and *human*, respectively, and these are called as *greeting-type*. On the other hand, the latter conditions that mean kinds of language in the greeting are called *greeting-style*. These nine conditions of greeting behavior were shown in random sequence to the participant.

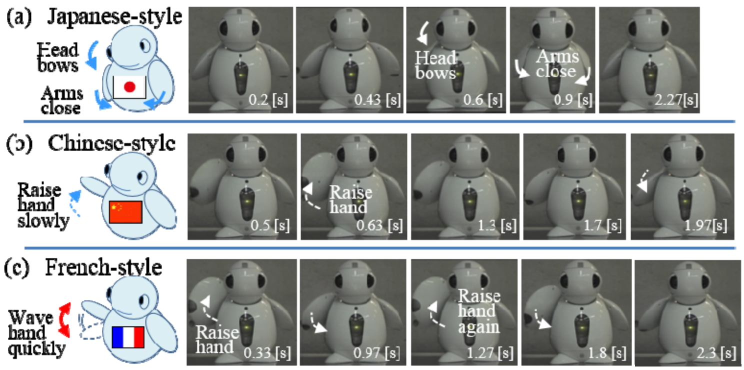


Fig. 2 Three types of robot greeting motion

For the NIRS measurement, a few minutes rest is required to calm down the brain activation which is induced by stimulus of previous task. And the total time of the experiment including the set-up preparation is desired to be less than 30 minutes to avoid fatigue. Therefore the rest period and the response period were specified to be 60 and 30 sec, respectively. The 90 sec period is called one task. Hence, the total time was 13.5 (=(60+30 sec) ×9 tasks) minutes for one participant. During the rest periods, the participant was asked to close his/her eyes in order to suppress the needless cognitive processing from visual information. The participant opens the eyes at a sign from the experimenter at five seconds prior to the start timing of the response period. At the beginning of the response period the participant watches the greeting behavior. When a thirty-second response period passed, the participant was asked to close the eyes again, and next task was started by changing the greeting conditions.

4 Result and Preparation for Analyses

The above-mentioned experimental procedure was approved by the university's ethics committee. Participants cooperated after he/she gave informed consent. Participants were eight Japanese (male, 20yrs to 22yrs) and four Chinese (three male and one female, 20yrs and 21yrs). In case of young adults like these participants, total-Hb reflects both oxy- and deoxy- change since the increase in oxy-Hb tends to be larger than the decrease in deoxy-Hb; hence, total-Hb was investigated in the present study.

Figure 3 shows examples of the time responses of the total Hb at the right M1 area for Japanese participant. Graphs (a),(b), and (c) are of the case of Japanese-style, Chinese-style, and French-style by the robot greeting with gesture, respectively. Graph (d) is other case of *no-gesture* by Japanese-style. Orange zone in each graph expresses the response period (60-90 [s]) and the change of the blood flow is evaluated for that period. On graph (e), the waveforms drawn in graphs (a)-(c) are plotted together to compare their differences in the response period. This graph shows that the absolute change in case of Japanese-style greeting (graph (a)) is larger than the other country styles (graphs (b) and (c)). This fact indicates the possibility that the country-by-country difference may be distinguished by investigating the blood flow. On the other hand, comparing the *gesture* case with the *no-gesture* one, which are plotted in graph (f), it can be expected that an effect of the gesture might be also detectable from the brain monitoring.

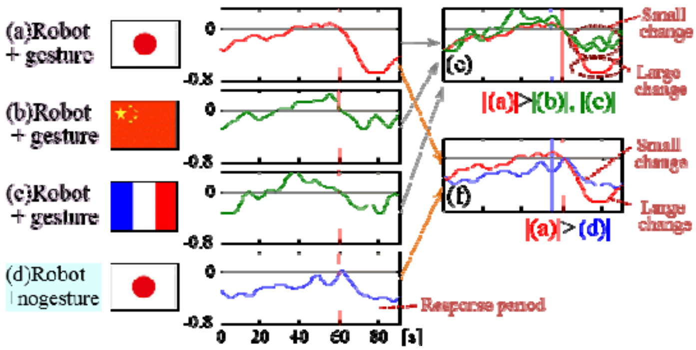


Fig. 3 Examples of change of cerebral blood oxygenation: comparison according to greeting styles and gesture types

As shown above, some kind of the blood flow response against stimulus of observation was confirmed. However, other various wave patterns like a decrease or flutter were also found, and it appeared inadequate to evaluate only the increasing pattern of the Hb change. Generally, usual changes in cerebral blood oxygenation are an increase of oxy-Hb, a decrease of deoxy-Hb, and an increase of total-Hb; this tendency is known as *pattern A*. Other *pattern B* (both oxy-Hb and total-Hb decrease) and *pattern C* (oxy-Hb, deoxy-Hb, and total-Hb increase) are also known. In case of high-level cognitive function such as linguistic activity, several patterns besides *patterns A-C* are reported [Tamura 1997]. Moreover, a seemingly paradoxical phenomenon such as deactivation, that means the decrease in both oxy-Hb and total-Hb occurs even on high nerve activity, is known. It is said that the deactivation occurs in BA44 of normal adult on linguistic activity like a reading out [Fallgatter 1998].

As introduced above, in case of linguistic activity that relates to the greeting situation that we adopted in this experiment, the increasing pattern is not always observed. Hence in the following analyses, the amount of blood oxygenation change is investigated instead of the maximum of the increase of the change, unlike other typical studies that utilized NIRS. Specifically, after computing a bias $\bar{y}$ by averaging the measured total-Hb signals $y(t)$ during two seconds just before the greeting stimulus starts, the sum of cerebral blood flow change (SBC) is computed as a sum of an absolute value of subtraction of the signal change from the obtained bias as follows:

$$S = \sum_{t=0}^{10} \left| y(t) - \bar{y} \right| \cdot \Delta \quad , \tag{1}$$

where t=0 [s] is the time count that the greeting action starts, Δ (= 0.1[s]) is a sampling interval, and the bias is computed as $\bar{y} = \frac{1}{2} \sum_{t=-2}^{0} y(t)$.

5 Analysis

After checking all data measured from all participants, it was found that the 22th signal of one Chinese participant and the 9th signal of another Chinese could not be measured adequately because of a defect of the NIRS probe setting; hence, these two channels data were eliminated for the following statistical analyses.

Analysis 1: Detection Possibility of the Country-by-Country Differences in Participants: The first analysis is to investigate whether the country-by-country differences in participants can be detected from the scientific measurement method that is a brain monitoring method using NIRS. For each case of the nine combinations of three greeting types and three styles, a two-sample test for a difference in mean was used: one group consists of SBC of all six channels concerning M1 and BA44 for eight Japanese participants (N=6×8=48), and the other group consists of same channels' SBC of four Chinese participants (N=6×4-2=22).

First, Kolmogorov-Smirnov test (p<0.05) was applied to check normality of the distribution of each group. When normality in both groups was confirmed, homoscedasticity was checked by variance ratio test (F-test, p<0.05). Two-sample t-test was applied when both normality and homoscedasticity could be confirmed for the two group selected (parametric method). When either normality or homoscedasticity was not recognized in any of two groups, Wilcoxon rank sum test was used to check a difference in mean of samples of the two groups (non-parametric method). The results are illustrated in Fig.4. The graphs (a), (b), and (c) show the case of Japanese-, Chinese-, and French-style, respectively. In each graph, averages of SBC of Japanese and Chinese groups are shown by blue and red bars, respectively. Range of standard deviation (SD) against each average is indicated by an error-bar using a 0.2-fold SD value. As shown in the figure, there were significant differences (p<0.05 or p<0.01) in averages of Japanese and Chinese participants' SBC in the six cases of total nine cases. Especially in the case of the

Japanese-style greeting (Fig.4(a)), averages of SBC in Japanese group were significantly higher than that of Chinese group in all three greeting types (p<0.05). Checking the French-style greeting (Fig.4(c)), average of Japanese participants' SBC was larger than that of Chinese, and there were significant differences in two cases of *robot+nogesture* and *human* greetings.

These results show that nationalities of Japanese and Chinese people can be distinguished by comparing the SBC value computed from the NIRS measurement data using the robot gesture observation protocol.

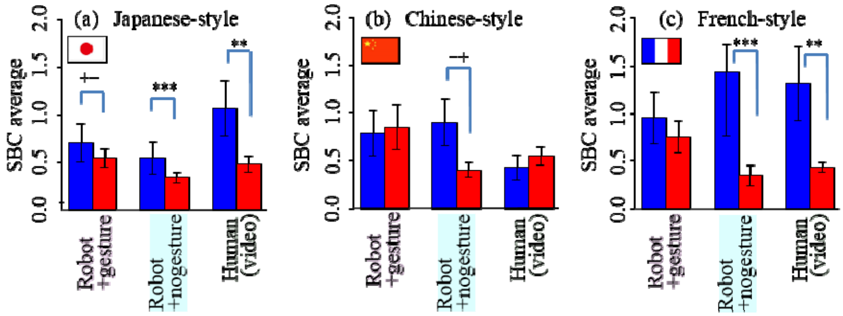


Fig. 4 Comparison of the SBC average of Japanese (blue bars) and Chinese (red bars) group (***p<0.01, **p<0.05; Wilcoxon rank sum test, ++p<0.05; t-test)

Analysis 2: Detection Possibility of Impression Difference: We investigated whether the difference of impression of a person who watched country-by-country different greeting styles could be detected by the brain monitoring method. Specifically, the difference in brain response on watching each greeting style were analyzed by ANOVA (Analysis of Variance).

Same datasets computed at Analysis 1 were used in Analysis 2. Each dataset for one participant consists of six channels' signal including the M1 and BA44. As a result of ANOVA, a significant difference was not confirmed for any combination (p<0.05). Hence, new datasets which included only M1 and BA44 separately were investigated by ANOVA again. For Japanese group, the numbers of samples for the M1 dataset and the BA44 one were N_{M1}=32 (=4×8) and N_{BA44}=16 (=2×8), respectively. Similarly for Chinese group, the numbers of samples are N_{M1}=15 (=4 × 4-1) and N_{BA44}=7 (=2 × 4-1), respectively. Computing P-values by the ANOVA, any P-value was larger than significant level 0.1; hence, there was no significant difference in brain response on watching country-by-country greeting styles. Difference in impression on watching different country's greeting could not be detected as a conclusion here.

Analysis 3: Induction of Country-by-Country Difference: Difference of brain response on watching the robot performing different country-by-country greeting was investigated in third analysis. Concretely, difference of the SBC values between two groups of *robot+gesture* and *robot+nogesture* was investigated.

The numbers of sample are N_J=48 and N_C=22 for Japanese and Chinese group, respectively, since all six channels' data per one participant were used. After checking the normality of each group, a similar test of difference in mean to Analysis 1 was executed by selecting adequate test methods according to the presence or absence of the normality in the same manner of Analysis 1. Figures 5(a) and (b) illustrate the mean values of SBC for case of *robot+gesture* (pink bars) and *robot+nogesture* (sky-blue bars), respectively. The graph format is the same as in Fig.4. As a result, Japanese participants showed significant difference (p<0.05) in case of the Japanese greeting style, as shown in Fig.5(a). In other words, the robot gesture could induce stronger brain response in Japanese participants than the robot voice without gesture, especially in case of their mother country style greeting. On the other hand in case of Chinese group, it was confirmed that Chinese tended to show difference (p<0.1) between *gesture greeting* and *non-gesture greeting* in any country greeting style, as shown in Fig.5(b). This result indicates that strong change in SBC of Chinese participants could be induced by adding gesture on robot behavior to three types of country greeting styles.

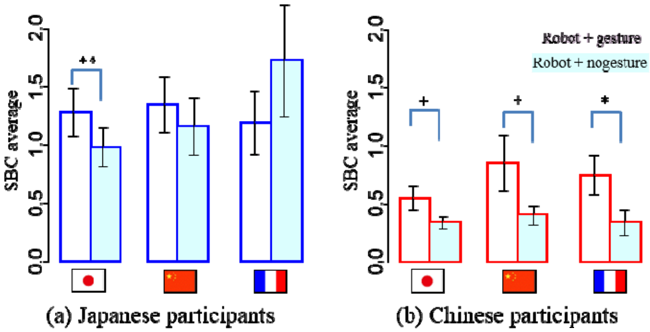


Fig. 5 Comparison of the SBC average by the presence or absence of robot's gesture (** p<0.05, * p<0.1; Wilcoxon rank sum test, $^+$ p<0.1; t-test)

Above-mentioned analysis proved that the robot gesture can induce country-by-country difference in brain responses of the Japanese and Chinese participants significantly at least. However, before that, it is necessary to confirm whether the robot can give greeting stimulus as well as a human can do, because such characteristics might be special in case of a use of robot. Therefore, additional analysis was executed to ensure that no difference in the brain response between the robot greeting observation and the human greeting one was present.

Figure 6 shows the results of a two-sample t-test to check differences of the brain response on watching the robot greeting and the human greeting. The test conditions are the same as in the above analysis. Means of SBC on observing the human greeting are shown by bars colored by leaf green. Figure 6(a) shows no significant difference in both cases of Japanese-style and French-style. For Chinese-style greeting, the robot greeting tends to induce difference (t-test, p<0.1). On the other hand, Fig. 6(b) for the Chinese participants indicates that there was a difference between the human greeting case and the robot case in only case of French-style (Wilcoxon, $p<0.05$). Conversely there was no significant difference between robot's and human greetings concerning each mother country language in both cases of Japanese and Chinese participants. Therefore, it can be said that enhancement of brain activation does not cause due to just a use of a robot, and the country-by-country differences were induced by an effect of robot gesture.

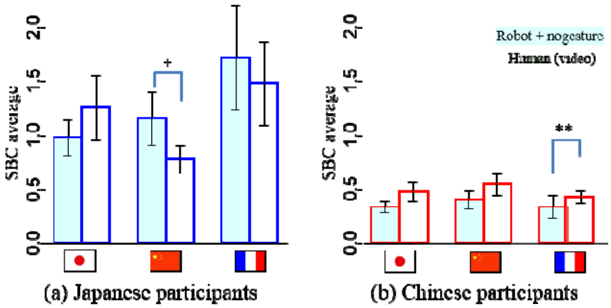


Fig. 6 Comparison of means in SBC in cases of robot and human greetings

6 Summary

Results of above-mentioned three analyses are summarized as follows:

- Investigation using nine conditions combined with three countries greeting styles and three greeting types found that SBCs of Japanese participants were significantly larger than that of Chinese under the six of nine conditions ($p<0.05$ or $p<0.01$). In other words, it was confirmed that scientific measurement technology, which is a brain blood flow measurement system, could distinguish the nationalities of Japanese and Chinese person objectively.
- Country-by-country difference in a brain response on watching different country's greeting were not able to be distinguished by the measurement protocol designed in this study.
- Addition of the gesture to the robot greeting action could enhance significantly the Japanese participants' brain response in the case of the mother country language greeting style ($p<0.05$). For Chinese participants, their brain response were increased to all three types of country greeting by addition of the gesture ($p<0.1$).

Acknowledgment. This work was supported by the Japan Science and Technology Agency for Feasibility study: the problem-solving research program for service science. Set-up of the experiment and measurement system was supported by K. Kuronuma, S. Matsunaga, and T. Kidokoro. Data analysis was supported by T. Shimazu. Motion program of ApriPoco was coded by H. Aoki. The authors appreciate their cooperation and feel grateful to many participants who cooperated with our experiment.

References

[Broadbent 2007] Broadbent, E., MacDonald, B., Jago, L., Juergens, M., Mazharullah, O.: Human reactions to good and bad robots. In: Proc. IEE/RSJ International Conference on Intelligent Robots and Systems, San Diego, USA, pp. 3703–3708 (2007)

[Fadiga 1995] Fadiga, L., Fogassi, L., Pavesi, G., Rizzolatti, G.: Motor facilitation during action observation: a magnetic stimulation study. J. of Neurophysiology 73(6), 2608–2611 (1995)

[Fallgatter 1998] Fallgatter, A.J., Müller, T.J., Strik, W.K.: Prefrontal hypooxygenation during language processing assessed with near-infrared spectroscopy. Neuropsychobiology 37, 215–218 (1998)

[Iacobonil 1999] Iacobonil, M., Woods, R.P., Brass, M., Bekkering, H., Mazziotta, J.C., Rizzolatti, G.: Cortical mechanisms of human imitation. Science 286(5449), 2526–2528 (1999)

[Kondo 1999] Kondo, T.: Service marketing. Japan Productivity Center, Japan (1999) (in Japanese)

[Lovelock 1995] Lovelock, C.H., Wright, L.K.: Principles of service marketing and management. Prentice Hall, Englewood Cliffs (1995)

[Michael 1970] Michael, W.O.: Proxemic behavior: cross-cultural study. Mouton De Gruyter, Germany (1970)

[Miyai 2002] Miyai, I., Yagura, H., Oda, I., Konishi, I., Eda, H., Suzuki, T., Kubota, K.: Premotor cortex is involved in restoration of gait in stroke. Annals of Neurology 52(2), 188–194 (2002)

[Rizzolatti 1998] Rizzolatti, G., Arbib, M.A.: Language within our grasp. Trends Neuroscience 21(5), 188–194 (1998)

[Rizzolatti 2004] Rizzolatti, G., Craighero, L.: The Mirror-neuron system. Annual Reviews of Neuroscience 27, 169–192 (2004)

[Rugimbana 2003] Rugimbana, R., Nwankwao, S.: Cross-cultural marketing. Cengage Learning, EMEA (2003)

[Suzuki 2011] Suzuki, S., Fujimoto, Y., Yamaguchi, T.: Can differences of nationalities be induced and measured by robot gesture communication? In: 4th International Conference on Human System Interaction, HSI 2011, Yokohama, Japan, pp. 357–362 (2011)

[Tamura 1997] Tamura, M., Hoshi, Y., Okada, F.: Localized near-infrared spectroscopy and functional optical imaging of brain activity. Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences 352(1354), 737–742 (1997)

[Tomiyama 2002] Tomiyama, T.: Expansion of service engineering. In: 6th Symposium on Research into Artifacts, Tokyo, Japan, pp. 45–58 (2002) (in Japanese)

Author Index

Subject Index

H

I

K

L

M

N

O

P

Q

R