PROCEEDINGS OF THE MULTIDISCIPLINARY DOCTORAL SCHOOL 2006-2007 ACADEMIC YEAR FACULTY OF INFORMATION TECHNOLOGY PÁZMÁNY PÉTER CATHOLIC UNIVERSITY

BUDAPEST

July, 2007

Pázmány University ePress
Budapest, 2007

Contents

Introduction

The Interdisciplinary Doctoral School at the Faculty of Information Technology of the Péter Pázmány Catholic University in Budapest has been established as a single doctoral program in 1993 as a collaboration of four universities and two laboratories, and moved to this university as an extended doctoral school in 2000.

This year, our Proceedings contains not only the new results of the doctoral students and related research grants found in the Proceedings published a year ago, however, a few new directions have started as well. Namely, the sensory wireless networks, the extensive studies related to our new 3 Tesla fMRI equipment, some new directions in nanotechnology, as well as the synchronization related cellular nonlinear dynamics.

We do believe that these short reports will give the reader an opportunity to have an insight into the research having been made recently. In addition we do encourage to contact the authors or supervisors for any additional information.

We acknowledge the many sponsors of the research reported here. Namely, the Hungarian National Research Fund (OTKA), the Hungarian Academy of Sciences, the National Research and Development Office (NKTH), the Gedeon Richter Company, the Office of Naval Research (ONR) of the US, the National Science Foundation of the US, and the Pázmány P. Catholic University. Thanks are also due to the collaborating Universities and Research Laboratories in Hungary and worldwide.

Budapest, July 2007.

TAMÁS ROSKA

Head of the Doctoral School

Fabrication and Characterization of Uncooled Infrared Nanosensors

Kristóf Tahy
(Supervisor: Árpád I. Csurgay)
kristof@tahy.hu

Abstract—**In this semester, my goal was to explore and understand the absorption mechanisms and the interaction properties of carbon nanotubes (CNTs) and infrared (IR) photons. I have fabricated interconnects in which CNTs make contact between Cu electrodes (300 nm in width) separated by distances ranging from 300 to 2500 nm far from each other. According my investigations this structure could be functioning as infrared sensor.**

I. Introduction

Modern IR detectors capable of sensing IR radiation in the two atmospheric windows – middle wavelengths 3-5 μm (MWIR) and long wavelengths 8-14 μm (LWIR) – are largely divided into two categories based on the principles of their operation, known as thermal and photon (quantum) detectors.

Current versions of photon IR detectors provide superior performance, which is essential for high-end applications, where performance requirements cannot be compromised. However, the combination of manufacturing difficulties and cooling requirements make these detectors very costly.

II. Carbon nanotube based solution

The ballistic electronic transport in CNTs drastically reduces the scattering probability, and the small diameter and the resulting strong electron confinement lead to suppression of short-channel effect. These advantages enable the CNT to be a potential infrared material. A CNT behaves as a quasi-one-dimensional wire with an electron energy spectrum (band structure) containing a series of sharp peaks known as van Hove singularities. The energy spacing between van Hove singularities varies from less than 0.1 eV to more than 2 eV depending on the nanotube diameter. Optical excitation and recombination across pairs of van Hove peaks provides the possibility to create nanotube optical detectors or emitters that operate across a wide range of optical wavelengths. Experiments have subsequently found that photoexcitation across pairs of van Hove peaks enhances the conductivity of semiconductor nanotubes. These peaks corresponds to the maximums in the IR absorption map of CNTs.[1]

The IR absorption of carbon nanotubes has already examined by several groups.[2,3] They found that CNTs have high absorption in the LWIR range and have many absorption peaks in the MWIR range. Since the band-gap is inversely proportional to the nanotube diameter and tubes with more than 10 nm width has less than 100 meV band-gap, the measured LWIR absorbance is not unexpected. In the MWIR range different bonding structures causes absorption increases such as sp^2 and sp^3 C-C bonds or –COOH, $-CH_x$ groups.

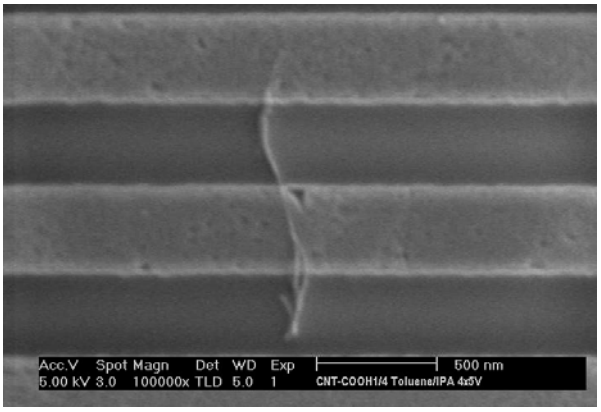


Fig. 1. SEM image of the sensor. d(CNT)~30 nm

III. Graphene based solution

The recent discovery of graphene, a single atomic sheet of graphite, has ignited intense research activities to explain the electronic properties of this novel two-dimensional (2D) electronic system. When graphene is patterned into a narrow ribbon, and the carriers are confined to a quasi one-dimensional (1D) system, the opening of an energy gap is expected. Similar to CNTs, this energy gap depends on the width and crystallographic orientation of the graphene nanoribbon. While to engineer the band-gap of CNTs is not too easy or even impossible, the same for graphene is possible. In addition depositing graphene with an exact width (and thus with exact band-gap) can be carried out lithographically.

IV. Conclusion

At the present state the devices are ready for the infrared measurements (Fig. 1.). Similar method will be developed for the fabrication of graphene devices what used for carbon nanotube sensors.

It is clear that the present solutions far not perfect in many sense like spectral selectivity, sensitivity or integration time. High quality devices are working at cryogenic temperatures. Carbon nanotube and graphene based sensors can be the solution for room temperature sensing with high detectivity as well as for sensing distinct frequencies at high speed.

References

[1] A. Mohite, S. Chakraborty, P. Gopinath, G. U. Sumanasekera, and B. W. Alphenaar, "Displacement current detection of photoconduction in carbon nanotubes," Appl. Phys. Lett., vol. 86, pp. 061114, Febr. 2005

[2] N. Kouklin, M. Tzolov, D. Straus, A. Yin, and J. M. Xu, "Infrared absorption properties of carbon nanotubes synthesized by chemical vapor deposition," Appl. Phys. Lett., vol. 85,, pp. 4463, Nov. 2004

[3] A. Ugawa, J. Hwang, H.H. Gommans, H. Tashiro, A.G. Rinzler and D.B. Tanner, "Far-infrared to visible optical conductivity of single-wall carbon nanotubes," Current Appl. Phys., vol. 1, pp. 45-49, Jan. 2001

Model Simplification of Reaction Kinetic Networks

Dávid Csercsik
(Supervisors: Katalin M. Hangos and Gábor Szederkényi)
csercsik@itk.ppke.hu

***Abstract*—This summary describes possible model simplification methods and applications of the reaction kinetic network system class. The availability of variable lumping model simplification methods for reaction kinetic networks is investigated. The simplified model remains in the class of reaction kinetic systems. The method is performed on one and two step cascade activation reactions.**

***Index Terms*—Reaction kinetic networks, model simplification, variable lumping transformation, systems biology**

I. Introduction

In biology there are many phenomena that cannot be explained with relevant properties of it's physical elements (e.g. molecules taking part in the phenomena), but with the dynamics of the interactions between its elements. Systems biology is the science that aims to study these interaction networks in biological systems. The spectrum of systems biology can be very wide, ranging from population dynamics to biochemical networks.

The theory of reaction kinetic systems [1], [2], [3] provides tools for analysis and model simplification of biochemical reaction systems (e.g. signaling pathways).

II. materials and Methods

In this section a short description of the engineering type model reduction called variable lumping is provided. The variable lumping is performed in a way which ensures that the resulting reduced system remains in the class of reaction kinetic networks. The intermediate product of the cascade, which acts as an enzyme in the second reaction is lumped together with the final product of the cascade. In this article two different methods of variable lumping are detailed:

- Cascade lumping with the variation of stochiometric coefficients: In this case the stochiometric coefficients are changed due to the approximation of the original input-output behavior.

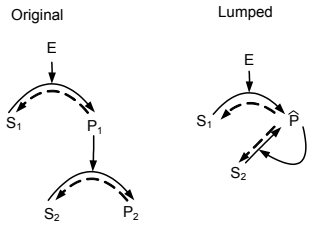


Fig. 1. The activation scheme of the original and the lumped reaction. The dashed lines denote the reversible case

- Cascade lumping with output transformation:

  Another way for performing the variable lumping transformation in irreversible case is to transform the output of the system. In this case the basic idea is that the transformed becomes product.

Furthermore the method can be extended to multi-level and feedback regulated cascades.

III. Conclusions

Model simplification provides a way to reduce reaction kinetic models of cell signaling pathways in order to obtain a minimal model, optimized for parameter estimation. Variable lumping may be a tool for model reduction in the case of enzyme-driven cascade activation reactions. The variable lumping can be performed various ways in order to find optimal approximation for the actual problem.

References

[1] G. Craciun and M. Feinberg, "Multiple equilibria in complex chemical reaction networks: II The species-reaction graph," *Siam Journal of Applied Mathematics*, vol. 66, pp. 1321–1338, 2006.

[2] ——, "Understanding bistability in complex enzyme-driven reaction networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, pp. 8697–8702, 2006.

[3] M. Feinberg, "On chemical kinetics of a certain class," *Archive for Rational Mechanics and Analysis*, vol. 46, pp. 1–41, 2004.

Grasping and Catching Tasks of Robot Manipulators Using Multimodal Sensory Information

Ferenc Lombai
(Supervisors: Tamás Roska and Gábor Szederkényi)
lombai.ferenc@itk.ppke.hu

***Abstract*—This paper contains plans and achievements considered during the investigation of the usage of multimodal sensory signals for catching and grasping objects with a 6 degree of freedom (DOF) robot manipulator. The early building steps of a test environment is proposed for catching flying objects while the sensorial input is aimed to be used in a behavioral manner. The arm is build up from industrial type PID controlled chain of electrical motors. For later system integration the presence of 3D tactile sensors and stereo vision is assumed. The paper gives basic insight to theoretical and technical issues arise during the implementation of everyday motion tasks which are 'easy for human' but difficult to reproduce by robots.**

I. Planned control scheme

On Figure 1 the data flow diagram can be seen where all subsystems perform real-time functionality. The task divisor which switches the system between approaching and caching modes based on the perceived tactile information. If no contact observed the path planning algorithm modulate its output curve which fits to the estimated target trajectory. To implement a visual-behavioral based approaching functionality slight further modifications required in the control scheme. When a contact occurred the trajectory planning algorithm will take the roll of the computations. The Path planning algorithm holds the last estimated path and the trajectory planner modulates the desired joint motion based on the measured tactile properties. During this last phase the trajectory generator drives the robot to a static pose with a stable grasping of the cached object.

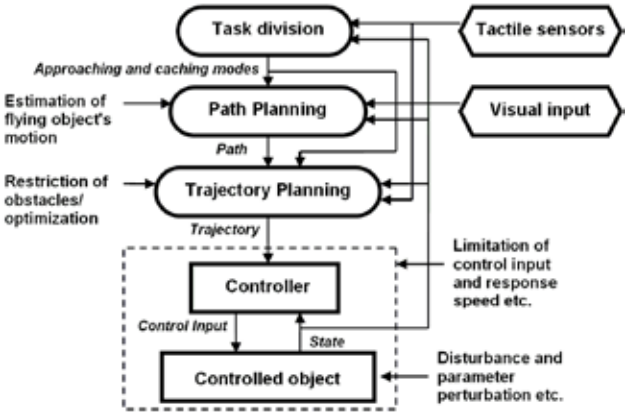


Figure 1. Planned real-time control shame.

II. Achievements

The first steps focused on the robot arm control to provide manipulation of a point with its orientation in space and time. The robot arm was built according to [1]. The DH and MDH parameters were obtained and visualized in different software environments (MATLAB, Mathematica, RobotBuilder). The inverse kinematic solution was derived and implemented in C++ language. The dynamic relationships $\mathbf{M}(\theta)$, $\mathbf{C}(\theta, \dot{\theta})$ and $\mathbf{G}(\theta)$ are symbolically given with the usage of [2] based on approximated inertial matrices obtained from the RobotBuilder model. The mentioned expressions should be used for power consumption optimization during path planning.

III. Further plans

There are further plans to complete trajectory tracking and path generating algorithms. During the control process velocity and acceleration constrains are also considered and early simulations show appropriate results. Later on the manipulator should be used and measured as a peace of a hierarchical control system. Important issues of the mounting and proper coverage for tactile sensors also have to be solved. An investigation is planned in plenty of grasping tasks for reconstruct object related information similar to type of surface, slippage, rotational movement or shape. The implementation of a behavioral based visual feedback, tracking, sensor identification, human-machine interfacing and teleoperation are parts of the future plans.

References

[1] C. Smith and H. I. Christensen, "Using cots to construct a high performance robot arm," in *Proc. IEEE ICRA'07*, 2007, pp. 4056–4063.
[2] P. I. Corke. (1996, Aug.) Robot symbolic dynamics package for maple. [Online]. Available: http://www.cat.csiro.au/ict/staff/pic/download.html

Partial Synchronization Regimes in Oscillator Arrays with Asymmetric Coupling

Dániel Hillier
(Supervisors: Tamás Roska and Joos Vandewalle)
hillier.daniel@itk.ppke.hu

***Abstract*—A new phenomenon is presented where highly asymmetric interaction weights can give rise to cluster synchronization regimes with partial synchronization. Cluster or partial synchronization regimes corresponding to asymmetric interaction patterns can break the underlying symmetries of the network topology and boundary conditions at the expense of some residual synchronization error.**

I. Introduction

Research on synchronization in coupled chaotic oscillators was so far mainly focusing on deriving conditions for synchronization. Putting no restriction on the interaction pattern, the possible spatial layouts of cluster synchronization regimes in a square array are investigated. The reason to allow any interaction pattern is that new kinds of cooperative behavior may be possible that were not observed in previous studies when coupling was kept identical for all cells.

II. Methods and Experiments

A square array composed of diffusively coupled, autonomous nonlinear oscillators with spatially varying interaction weights and bifurcation parameters is considered. Cells are four-connected to their nearest neighbors. The problem of finding cluster synchronization regimes was cast into an optimization problem where the cost function is constructed in such a way that a specified set of cells in the array synchronize with each other but not with the rest of the array.

Fig. 1 shows array configurations giving rise to cluster synchronization regimes. Dashed rectangles indicate cell pairs that were learnt to synchronize during optimization. These pairs may be part of a larger cluster synchronization regime following the spatial symmetry related rules of [1]. The sole exception is the pair of cells with coordinates $(1, 2)$ and $(3, 1)$ that is not symmetric to any axis. This case is not shown on Fig. 1 since no array configuration was found that could synchronize this cell pair. The error between synchronizing cells is not zero. Such type of synchronization was already studied extensively in [2] where it was shown that qualitatively different dynamical systems can synchronize with some small residual error. Our examples confirm this phenomenon extending it into the more complex context of cluster synchronization occurring in an array of oscillators.

III. Conclusions

In [1] symmetries of the coupling topology were exploited to find cluster synchronization regimes of identical oscillators with identical coupling weights. In this study, the oscillators were not identical and the coupling coefficients were highly asymmetric, still the symmetries similar to those shown in [1] can be observed in the resulting cluster synchronization regimes. However, at the expense of some residual synchronization error, the asymmetry of the interaction pattern between cells can give rise to partial synchronization regimes that break the underlying symmetries of the network topology and boundary conditions. This is a new phenomenon to be analyzed more in detail in future studies.

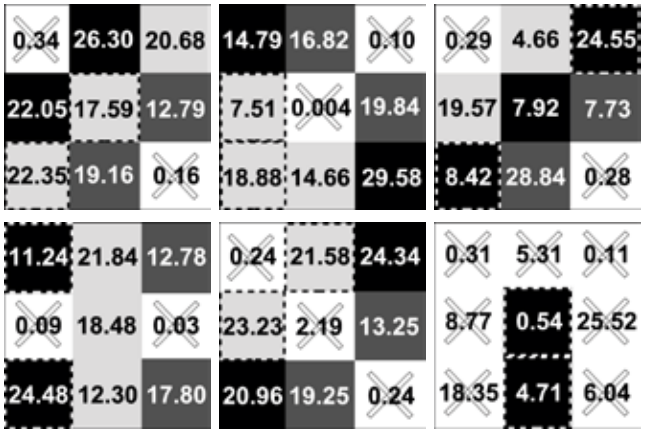


Fig. 1. Cluster/partial synchronization regimes learnt using the proposed method. Cells belonging to the same synchronizing cluster are marked with the same grayscale level. White cells marked with X do not synchronize to any other cell. Dashed rectangles indicate cell pairs that were learnt to synchronize during optimization.

Acknowledgements. Supported by the Research Council KUL: GOA-AMBioRICS, CoE EF/05/006, Flemish Gov: FWO: G.0211.05, G.0226.06, Belgian FSPO IUAP P5/22, The Scientific and Tech. Research Council of Turkey (TUBITAK).

References

[1] I. Belykh, V. Belykh, K. Nevidin, and M. Hasler. Persistent clusters in lattices of coupled nonidentical chaotic systems. *Chaos*, 13:165–178, 2003.

[2] J. A. K. Suykens, P.F. Curran, and L.O. Chua. Robust synthesis for master-slave synchronization of Lur'e systems'. *IEEE T. on CAS-I*, 46(7):841–850, 1999.

Elastic Grid Based Motion Field Analysis in UAV/UGV Applications

Gergely Balázs Soós
(Supervisor: Csaba Rekeczky)
soos@digitus.itk.ppke.hu

***Abstract*— Unmanned vehicles, like small planes (UAVs) or ground platforms (UGVs) can capture and transmit valuable visual information to a human observer from distant or dangerous location. Moreover this sensor modality can be used to extend navigation system of the platform.**

I present an automatic focusing method for detecting independently moving ground objects form a UAV, and give some ideas for extending it for UGV applications.

I. Introduction

The aim of our recent project ALFA was to develop a visual navigation system combined with a commercial autopilot for reconnaissance and surveillance tasks. During the field experiments the path of the flight is modified when an interesting event is detected on the ground in order to collect more detailed information.

For a ground platform inspection field is similar to the field of motion. Detection of independently moving object is important for recognizing other agents acting in the environment for avoiding collision during operation.

II. Model of Image Motion (OF)

If the camera is steady then detecting moving objects can be based on building up background statistics over time and marking the changing areas as foreground or moving objects. The process of detecting independently moving objects in the flow of a moving camera is called *independent motion analysis*.

When the observed 3D structure can be considered smooth in distance then 2D images can be registered using a parametric transform, and residual errors are labeled as moving objects. Registration based techniques can be extended to sparse 3D case, using multiple planes and layered representation.

A projective transformation can be used to model the transformation of the camera mapping of the 3D environment points into 2D screen points. When the platform is moving the camera is undergoing 3D translation and rotation. The projection itself can not be inverted although projections of a 3D plane can be aligned exactly with a single projective transformation. It implies that points from a plane forming a line will be mapped to collinear configuration.

I am using a multi-adaptive computational approach designed for UAV applications. It incorporates correlation based sparse optical flow estimation for all grid points in a regular sparse grid.

III. Elastic Grid

The elastic grid approach is an extension of the deformable contour paradigm. A grid is defined with control points and forces. Each point is connected to its closest neighbors in "4-connected" way. Points are deformed by external forces moving control points for best correlation matching and internal forces are representing the collinearity constraint. After evolution for optimization remaining forces are highlighting interesting locations: object boundaries, moving objects or outliers in distance. These locations are the output of the focusing mechanism, and selected for thorough analysis.

Results

The focusing mechanism was compared to random selection – fig 1. More details are in recent publication [1]. The method can be extended to sparse 3D using layered representation.

Acknowledgment

The author acknowledge the support from the ALFA - "Title" NKFP 2/046/04 project funded by NKTH.

References

[1] G.B. Soós, Cs. Rekeczky, "Elastic Grid Based Analysis of Motion Field for Object-Motion Detection in Airbone Video Flows ", In Proc. of the IEEE International Symposium on Circuits and Systems, 2007.

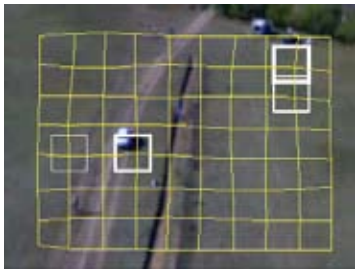
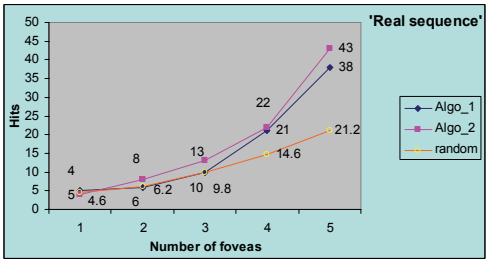
a,)  b,)  c,) 

Figure 1. Result on a 'Real UAV sequence': snapshot of OF calculation (a), illustration of foveal windows (b). Number of hits with different number of foveas for Algo_1 and Algo_2 – (different window size) compared with Random selection. There were 203 frames in the flow were the car was present.

Deep Brain Stimulation in Epilepsy Research

László Grand
(Supervisors: István Ulbert, György Karmos, Daryl R. Kipke)
grand@cogpsyphy.hu

***Abstract*— Today, most of the clinical electrodes are macroelectrodes. Recording and/or stimulating small brain structures is usually impossible or very complicated with those because of their relatively large size and disadvantageous stimulation characteristics. The goal of the Human Hippocampal Epilepsy Electrode (H2E2) project is to develop polymer based electrodes that are more convenient for epilepsy research with their small size and good quality.**

We would like to use these acute and sub-chronic probes for electrical stimulation of the hippocampus in order to reduce or suppress the epileptic paroxysmal activity of the Temporal Lobe.

I. Introduction

About 60 percent of people with epilepsy have focal seizures and the Temporal Lobe Epilepsy is a very common type of focal epilepsy. Presumably, the hippocampus plays a significant role in paroxysmal activity generation. It is very hard, usually impossible, to control the TLE with medications. That is why eligible applicants usually accept hippocampectomy as the only existing recourse.

Clinical experience shows that ictal and inter-ictal epileptiform EEG activity first occurs in the hippocampus. Hippocampal stimulation may reduce both complex and partial secondarily generalized tonic-clonic seizures. Additionally, inter-ictal spikes may be either blocked completely or significantly reduced by electrical stimulation. The literature lacks extended discussion of hippocampal stimulation and human hippocampal in vivo layer recording is very rare.

The goal of the Human Hippocampal Epilepsy Electrode (H2E2) project is to build microelectrodes and use them instead of macroelectrodes that are so prevelent in human epilepsy research. Other goals are to investigate the exact role of the hippocampus in TLE generation and to assess the electrical stimulation with different parameters on the TLE activity in different species.

L. G. Author is with the Peter Pázmány Catholic University, 50/A Práter St., Budapest, 1083, Hungary. He spent one year at NeuroNexus Technologies Inc. and at the Neural Engineering Laboratory (NEL) of the University of Michigan in Ann Arbor, MI, USA.

U.I. and G.K. authors are with the Peter Pázmány Catholic University and also with the Institute of Psychology of the Hungarian Academy of Sciences, 83-85 Szondi St., Budapest, 1068 Hungary.

D.R.K. author is with NeuroNexus Technologies Inc., 3985 Research Park Dr. Suite 100, Ann Arbor, MI 48108 USA. and also with NEL

II. Results

Last year, I joined the Deep Brain Stimulator development team at NeuroNexus Technologies Inc. The goal of this group is to develop probes that can be used both for electric recording and stimulation. Details about these probes may be accessible after they are commercially available.

Besides the development of the H2E2 probe, an important part of my work was to design benchmark tests that make it possible for us to compare electrical properties of different probes that were made with different technology. These include impedance, shorting, cross-talk, soak and artificial spike tests.

III. Future Plans

I would like to improve the performance of the acute H2E2 next year based on in vitro tests. Besides, in vivo animal hippocampal studies and hopefully human trials will be made.

Acknowledgement

I thank the generous one-year undergraduate fellowship of the Hungarian-American Enterprise Scholarship Fund that made it possible for me to spend one year in Ann Arbor, Michigan.

References

[1] D.R. CHABOLLA, A. M. MURRO, R. R. GOODMAN, G. L. BARKLEY, G. A. WORRELL, J. F. DRAZKOWSKI, D. R. LABAR, M.C. SMITH, G. RYDER, F. SUN, R. ESTELLER, M. J. MORRELL, " Treatment of Mesial Temporal Lobe Epilepsy with Responsive Hippocampal Stimulation by the RNS Neurostimulator," NeuroPace Inc., Mountain View, CA, Neuropace Publication, 2006.

[2] W. H. THEODORE, "Brain stimulation for epilepsy," Nature Clinical Practice Neurology, vol. 1, pp. 64-65, 2005.

[3] M. VELASCO, F. VELASCO, A. L. VELASCO, B. BOLEAGA, F. JIMENEZ, F. BRITO, I. MARQUEZ, "Subacute electrical stimulation of the hippocampus blocks intractable temporal lobe seizures and paroxysmal EEG activities," Epilepsia, vol. 41, pp. 158–169, 2000.

[4] K. VONCK, P. BOON, E. ACHTEN, J. DE REUCK, J. CAEMAERT, "Long-term amygdalo-hippocampal stimulation for refractory temporal lobe epilepsy," Annals of Neurology, vol. 52, pp. 556–565, 2002.

[5] K. VONCK, P. BOON, P. CLAEYS, S. DEDEURWAERDERE, R. ACHTEN, D. VAN ROOST, "Long- term deep brain stimulation for refractory temporal lobe epilepsy," Epilepsia, vol. 46(suppl 5), pp. 98–99, 2005.

Non-standard Cellular Neural Networks in Physics

Mária Ercsey-Ravasz
(Supervisors: Tamás Roska and Zoltán Néda)
ravasz@digitus.itk.ppke.hu

As a last part of a longer project concerning statistical physics applications on the CNN-UM this year a non-standard locally variant CNN was simulated and used for developing optimization methods on frustrated spin-glass type models. A new project was also started including experiments and computational studies of a very special cellular neural network built up by pulse-coupled neurons (oscillators) with global and inhibitory type interactions. This is a still ongoing work showing very promising opportunities.

For the Cellular Neural Network Universal Machine (CNN-UM) hardwares realized and used until now the templates $\{A, B, z\}$ describing the connection parameters are identical for all cells. This means that the strength of the connections between neighboring cells can be changed, but it varies simultaneously and identically for all cells. We studied a locally variant CNN, on which these connection-parameters could be controlled separately for each cell. This CNN was used to study and optimize frustrated models like spin-glasses. It can be shown that this CNN is an analog correspondent of locally coupled spin-glasses. The only difference between the two systems is that for CNN the variables are continuous in the [-1,1] range and not discrete ones (± 1) like for the usual Ising-type spin models. Starting from an initial condition the final steady state of the template - meaning the result of an operation - will be always a local minimum of the generalized Ising type spin model with connections defined in matrix A. The fact that one single operation is needed for finding a local minimum of the energy, gave us the opportunity for developing very fast optimization algorithms. Spin-glasses are an important representative of frustrated systems, and they are used in many different applications, for example it has been shown that using spin-glass models as error-correcting codes, their cost-performance is excellent. The results of this project were presented in a contributed talk on the conference: "Unconventional computing: Quo vadis?", in Santa Fe, march 2007, and were submitted for publication in the special issue of Physica D Nonlinear Phenomena: "Novel computing paradigms".

In our second project a simple system composed of electronic oscillators capable of emitting and detecting light-pulses was studied. Many interesting synchronization phenomena occur in systems composed of simple units. These simple units in large number and coupled by simple rules can produce extraordinary complex behavior. Nature is also full with many examples in this sense, one phenomena which inspired us in this work is the collective behavior of fireflies. Although our aim was not to model fireflies, the oscillators ("electronic fireflies") considered in our system (Fig.1.) are somehow similar to them: they are capable of emitting light-pulses and detecting the light-pulse of the others. Of course the units can also be considered as pulse-coupled neurons with global, inhibitory type interactions, which means that the firing of one neuron delays the firing of all the others and all neurons have a firing period between a minimal and a maximal value. The system will not necessarily favor synchronization, it is rather designed to keep a desired light intensity in the system. This light intensity is controlled by a firing threshold parameter imposed globally on the oscillators. Interestingly, as a co-product of these simple rules for certain region of the firing threshold parameter phase-locking appears producing very complex patterns in the flashing sequence of the oscillators.

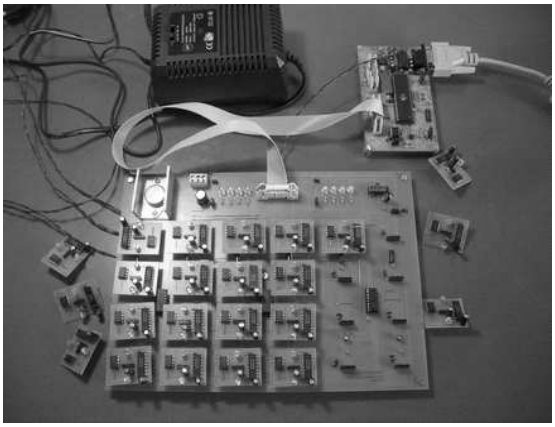


Fig. 1. Experimental setup.

Our further goal is to study a system in which the threshold parameter can be separately controlled for each oscillator. This would be equivalent with introducing different inputs in the system. The behavior of this system also inspired us for studying a CNN model built up by pulse-coupled, firing neurons with local - and not necessarily inhibitory type - interactions. A CNN model like this would have the advantage that a continuous flow can be used for input, and the cells (neurons) are not supposed to be identical or perfect - noise is allowed in the system. This project shows very great opportunities in developing algorithms for detection, pattern recognition etc.

Modeling Visual Attention and Stabilizing its' Input

Anna Lázár
(Supervisors: Tamás Roska, Zoltán Vidnyánszky and Marc Van Hulle[1])
lazar@digitus.itk.ppke.hu

Abstract— **This summary, from one hand, contains the final adjustment and validation results of the bottom-up visual attentional model that we started during the previous years, and, from the other hand, includes a pre-processing method aiming to stabilize the camera input for real applications, namely the 'Bionic Eyeglass Project'.**

Index Terms—**visual attention, retina channels, modeling, Cellular Neural/Nonlinear Networks, CNN, bottom-up attention, video stabilization, area based method, optic flow**

Visual attention is an ability of living creatures to filter out the actually relevant information from the visual scene in real time. From an engineering viewpoint, a system like that can save enormous processing capacity, thus increasing the *quality*, meanwhile decreasing the *time* necessary for the process.

Human visual attentional system is compiled of two different, but closely interconnected methods: the "bottom-up" and the "top-down". Former is a fast, unconscious method, responsible for reacting on unexpected, sudden events, while the latter is slower, and enables voluntary search and focal visual attention. Our aim was to model the former one, which - summarized very briefly -, works as next:

First the input scene dissolves into ten topographic maps, according to the ten retina channels (discovered recently). Then each retina channel creates its' own 'saliency map' with a given receptive field size. Once this is done, these maps are aggregated into one final or 'master' saliency map, which thus already feature-independently codes that how much the different locations are 'conspicuous', "worthy for being attendant to".

During the last years, a neuromorph skeleton was created with a few unknown parameters: the *receptive field size* for the certain retina channels (with which the channel-based saliency maps are created) and the *weights* during the formation of the final (or "master") saliency map.

With the purpose of determining these unknown values, and later to validate the model, we made human gaze direction measurements on two different video sets (both contained moving natural scenes, because this kind of stimulus primarily trigger bottom-up attentional control, according to the literature). The evaluation of these measurements finished in the first semester of this year, namely determinating the optimal receptive field size for each retina channel and defining the weights of the channel-based saliency maps which by they form the master saliency map. The model refined with these parameters was tested as next: for every frame of the test video set (on which we afterwards made gaze-direction measurements too), we made *predictions:* where we expect a human observer to attend to. We calculated four predictions for every frame: the first most probable location, the second, etc. Then we calculated the ratio of the "hits". (We defined "hit", if the distance between the predicted location and the measured fixation was less then 5 degree)

The model performed surprisingly well: the fixation locations (measured on humans), was among the first four predicted locations in around 70%. The accidental chance for this is a bit less then 20%.

During the second semester – which I spent in Leuven, (Belgium) – we concentrated on the practical applicability of the above model. More specifically, we made a *pre-processing step* to adopt it to the "Bionic Eyeglass Project", namely to *stabilize* the input coming from a camera of a mobile phone held by a (blind) person during walking (or traveling, etc.). This is necessary because the above described model works on inputs recorded with *motionless* camera. This is a result of the retina-channel decomposition, because seven channel of the ten responses more or less to motion and change.

The best choice proved to be an 'optic flow' technique (an area-based image-registration method) with a linear affine mapping transformation model.

In the future we plan to complete this stabilization method with a *fixation*-stabilization technique (stabilizing only the attendant area independently from other motions present in the input) as a first step of scene analyzing and object recognition methods.

[1] K.U.Leuven, Leuven, Belgium – Laboratorium Neuro- en Psychofysiologie

Epileptic Seizure Detection and Prediction based on Hurst Exponent Estimation

Béla Weiss

(Supervisors: Tamás Roska, György Karmos and Ronald Tetzlaff)
weiss@itk.ppke.hu

***Abstract*—Long-range dependence of brain electrical activity in epilepsy has been analyzed by estimation of the Hurst exponent (*H*). It was found that *H* drastically changes during seizures. In the preictal state gradual increase, in the postictal period gradual decrease of *H* can be observed. We hope that these properties could be used for detection and prediction of epileptic seizures.**

I. Introduction

Epilepsy is the second most common neurological disorder, and affects 1% of the population. For 25% of patients, no sufficient treatment is currently available. Method capable of detecting, and predicting the occurrence of sudden, unforeseen seizures could significantly improve the therapeutic possibilities [1].

II. Methods, Data and Results

Considering spike trains it was found that there are long-term correlations among interspike intervals [2]. A fractal spike train process is statistically self-similar, which means that fluctuations and other properties over brief times are proportional to those measured over a longer period. Development of brain electrical activity is hierarchical so it is assumed that signals recorded by intracranial macro electrodes keep the self-similar property. The self-similarity parameter - *H* - describes the long-range dependence of the process. A method based on the rescaled adjusted range has been implemented for estimation of *H* [3].

Temporal and frontal seizures have been assessed. Long-term recordings were provided by Epilepsy Center Freiburg (ECF) and National Institute of Neurosurgery (NIN), Budapest, Hungary.

On Fig. 1 drops of *H* (blue curve) can be observed approximately 30s after seizures' onset (vertical red lines). Applying H_{dth} = 0.6 threshold value, seizures can be detected. *H* gradually increases (green curve) before seizures, and decreases in the postictal period. For used parameter values one can set H_{pth} = 0.8 threshold value to predict impending seizures without false positive alarms.

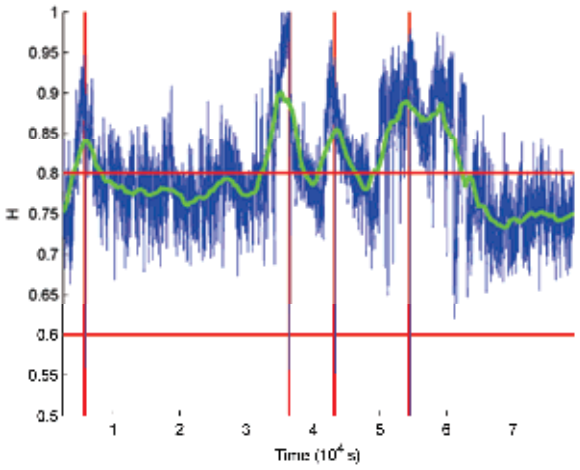


Fig. 1. Patient G.J. (NIN), 22h 33m long recording, 4 seizures. Blue curve – *H* values of 33s long data segments, 8 channels averaged; Green curve – *H* values of 1 hour long data segments, 8 channels averaged; Vertical red lines – seizures' onset time; Horizontal red lines – applied detection (H_{dth}) and prediction (H_{pth}) thresholds.

III. Conclusion

Analyzed seizures can be detected and predicted by estimation of *H*.

Acknowledgment

I would like to thank to Zsuzsanna Vágó for helpful comments, to ECF and to NIN for provided recordings.

References

[1] F. Mormann, C. E. Elger, K. Lehnertz, "Seizure anticipation: from algorithms to clinical practice," Current Opinion in Neurology 2006, 19:187-193.

[2] C. D. Lewis, G. K. Gebber, P. D. Larsen, S. M. Barman, "Long-Term Correlation in the Spike Trains of Medullary Sympathetic Neurons," The Journal of Neurophysiology Vol. 85 No. 4 April 2001, pp. 1614-1622.

[3] B. B. Mandelbrot, M. S. Taqqu, "Robust R/S analysis of long-term serial correlation," in Proceedings of the 42nd Session of the International Statistical Institute, Manila, 1979. Bulletin of the International Statistical Institute. Vol. 48, Book 2, pp. 69-104.

Towards More Natural Hearing with Cochlear Implants: O*verview, News and Outlook*

Tamás Harczos
(Supervisors: Tamás Roska and Karlheinz Brandenburg)
harczos.tamas@itk.ppke.hu

***Abstract* — Modern auditory implants are undoubtedly the most successful prostheses ever. Yet, current cochlear implants do not truly mimic the human auditory system. A neurophysiologically parameterized auditory model will be discussed, which could substantially enhance, or even substitute, current "quick and dirty" calculation methods. The proposed model was evaluated in several ways. Some of these will be presented in this work, which is meant to provide summary on the last year's work.**

***Index Terms* — cochlear implant, hearing aid, auditory processing.**

I. Introduction

This document is a very short summary on last year's work, summarizing only the major issues of the period. Aims, realization and outcomes will be shortly presented. Then a glimpse into ongoing work and an outlook will be outlined. Corresponding publications will be referred to during the text.

II. Recent Research Aims

We have shown that cochlear delay trajectories (CDTs) are noise-robust features for speech recognition, which means that cochlear implant strategies of the future should not ignore this unique feature of the human auditory processing chain. Our statement has been verified by automatic speech recognition (ASR) tests on a significant part of the well-known TIMIT continuous speech database [1].

Furthermore, we have used our ASR framework to simulate speech recognition possibilities of CI patients. Three different CI strategies were tested with various numbers of electrodes. We found that results will not get better by increasing the number of CI channels above twelve. This is consistent with speech intelligibility tests obtained with CI patients [4]. We presume that by employing temporally finer CI strategies and by placing the stimulating electrodes more accurately into the cochlea, a new dimension of hearing sensation could be achieved [2].

We believe that current CI strategies will soon be replaced by biologically motivated ones. We examined various cochlea modeling levels and found that in unforeseen noise situations the best choice is always the exhaustive simulation, i.e., simulation up to the auditory nerve populations. We have also found that a good compromise (with decreased complexity) may be to simulate up to the level of neurotransmitter concentration in synaptic cleft of the inner hair cells, and to use these values as features for further processing [3]. This finding may play an important role for designing future cochlear implants.

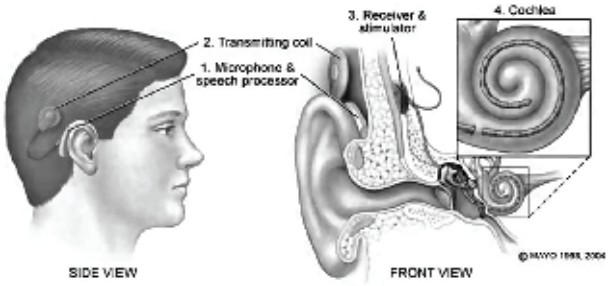


Fig. 1. Typical cochlear implant system with behind-the-ear processor.

References

[1] T. Harczos, G. Szepannek, and F. Klefenz, "Towards Automatic Speech Recognition based on Cochlear Traveling Wave Delay Trajectories," *Int. Symposium on Auditory and Audiological Research ISAAR*, Helsingør, Denmark, 2007. **(accepted)**

[2] W. Nogueira, A. Kátai, T. Harczos, F. Klefenz, A. Büchner, and B. Edler, "An Auditory Model based Strategy for Cochlear Implants," *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBC*, Lyon, France, 2007. **(accepted)**

[3] T. Harczos, W. Nogueira, G. Szepannek, and F. Klefenz, "Comparative evaluation of successive cochlear modeling stages as possible front-ends for automatic speech recognition," *19th Int. Congress on Acoustics ICA*, Madrid, Spain, 2007. **(accepted)**

[4] W. Nogueira, T. Harczos, B. Edler, J. Ostermann, and A. Büchner, "Automatic Speech Recognition with a Cochlear Implant Front-End," *8th Annual Conf. Interspeech*, Antwerp, Belgium, 2007. **(accepted)**

T. H. is with the Fraunhofer Institute for Digital Media Technology (Fraunhofer IDMT), Ehrenbergstrasse 29, 98693 Ilmenau, Germany (phone: +49 3677 467225, fax: +49 3677 467467; e-mail: hzs@idmt.fraunhofer.de); and with the Faculty of Information Technology, Péter Pázmány Catholic University, Práter u. 50/a, 1083 Budapest, Hungary.

T. R. is with the Faculty of Information Technology, Péter Pázmány Catholic University, Práter u. 50/a, 1083 Budapest, Hungary; and Regular Member of the Hungarian Academy of Sciences (e-mail: roska@sztaki.hu).

K. B. is with the Fraunhofer IDMT, Ehrenbergstrasse 29, 98693 Ilmenau, Germany; and full professor at the Technical University of Ilmenau (e-mail: bdg@idmt.fraunhofer.de).

Simulating Small Peptides Using Discrete Molecular Dynamics

Gergely Gyimesi
(Supervisors: András Szilágyi and Péter Závodszky)
ggyimesi@enzim.hu

***Abstract*—Discrete (or discontinuous) molecular dynamics (DMD) is a method for molecular dynamics simulations which uses step potentials with discrete energy values. Combining this method with coarse-grained protein models yields a speedup of several orders of magnitude relative to conventional molecular dynamics. We implemented a coarse-grained protein model to study the folding and association properties of the Trp-cage miniprotein. In our folding and unfolding simulations, we found the results to be strongly dependent on forcefield parameters including atomic radii. After correction using feedback from our simulations we were able to fold Trp-cage to within 2 Å C_α-RMSD of the native structure.**

***Index Terms*—discrete, molecular dynamics, Trp-cage, protein**

I. Introduction

Discrete (or discontinuous) molecular dynamics (DMD) is an old but significantly underutilized method to simulate macromolecular systems. DMD uses a step potential with discrete energy values. Instead of integrating the equations of motion using a finite time step, an event-driven algorithm is employed, and the dynamics simplifies to a series of collision events separated by ballistic runs. This results in a speedup of several orders of magnitude relative to conventional molecular dynamics [1]. We have implemented an extensible and scalable molecular dynamics framework using the DMD approach.

II. Applications

We applied the DMD method to study the folding and association properties of Trp-cage, a 20 residue miniprotein. It exhibits cooperative folding with a melting point of 3 °C and a folding time of approx. 4 μs, which makes it one of the fastest folding proteins known [2]. A point mutant variant of the Trp-cage has been shown to form amyloid-like fibrils when exposed to heat treatment. These structures show high β-sheet content but their exact atomic structure and mechanism of formation remains unknown. In order to study the molecular mechanism of association and propose an atomic-level structure of the fibrils we applied the methodology of discrete molecular dynamics together with a suitable coarse-grained model adopted from the literature.

G. Gyimesi is with Péter Pázmány Catholic University, Department of Information Technology, Budapest, Hungary and the Institute of Enzymology, BRC HAS, Budapest, Hungary (e-mail: ggyimesi@enzim.hu).

A. Szilágyi and P. Závodszky were with the Institute of Enzymology, BRC HAS, Budapest, Hungary.

III. Coarse-grained Model

We adopted a coarse-grained model developed to describe the folding of the Trp-cage miniprotein [3]. The model employs 4 backbone atoms and up to three side-chain beads to represent the geometry of the protein. Side-chain interactions include the hydrophobic interaction, aromatic interaction between aromatic amino acids, aromatic-proline interaction between aromatic and proline side-chains, and salt-bridge interaction between oppositely charged side-chains. Only pairwise interactions between side-chain beads are considered in this model, and the potential functions are stepwise.

The original published model was missing several important parameters and required significant corrections. Using our corrected model, we are now able to consistently fold the protein to within 2 Å C_α-RMSD of the native structure (Fig. 1). The model is ready to be applied for the study of the oligomerization and association properties of the Trp-cage miniprotein.

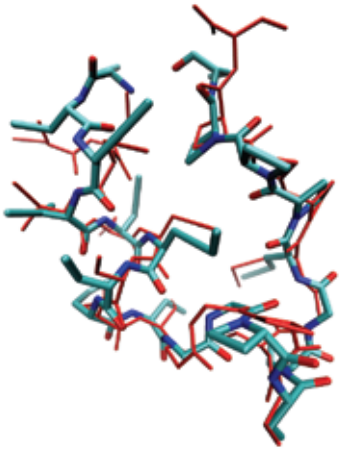


Fig. 1. Superposition of a low-RMSD structure (colored according to atom type) with the native structure (red). Because the model is coarse-grained, only the reduced side-chains are shown.

References

[1] F. Ding, N. V. Dokholyan, "Simple but predictive protein models," *Trends in Biotechnology* 2005 Sep;23(9):450-5

[2] L. Qiu, S. A. Pabit, A. E. Roitberg, S. J. Hagen, "Smaller and faster: the 20-residue Trp-cage protein folds in 4 micros," *J Am Chem Soc* 2002 Nov 6;124(44):12952-3

[3] F. Ding, S. V. Buldyrev, N. V. Dokholyan, "Folding Trp-cage to NMR resolution native structure using a coarse-grained protein model," *Biophys J.* 2005 Jan;88(1):147-55

Perceptual and Neural Mechanisms of Decision Making about Motion Direction

István Kóbor
(Supervisor: Zoltán Vidnyánszky)
kobor.istvan@itk.ppke.hu

Abstract—**Perceptual decision making in a visual discrimination task depends on how strong and unambiguous is the task-relevant sensory signal. Our goal was to investigate the neural correlates and dynamics of decision making using electroencephalography (EEG). Observers performed a motion direction discrimination task. Discrimination difficulty was regulated by changing the percentage of coherently moving dots in the random dot motion display used as stimulus (there were six coherence levels). EEG (64 channels) data were analyzed using the average-based method. Average event related response (ERP) amplitudes on the occipito-temporal (PO7) electrodes were modulated by the motion coherence level within two intervals: 250-380 ms after stimulus onset amplitudes were more negative whereas between 400-600 ms they were more positive with increasing motion coherence (Fig.1). On the parietal (Pz) and frontal (Fcz) electrodes, however, motion coherence-dependent modulation started later: amplitudes become more positive starting from 350 ms as coherence was increased (Fig.2,3). The onset delay of the later component on the occipito-temporal electrodes and the onset of the modulation on the parietal and frontal electrodes was inversely correlated with motion coherence. These results suggest that decision difficulty is reflected in ERP responses and that in the motion discrimination task used in our study decision difficulty component arises ~ 350 ms after stimulus presentation.**

Index Terms—**perceptual decision making, visual discrimination, electroencephalography**

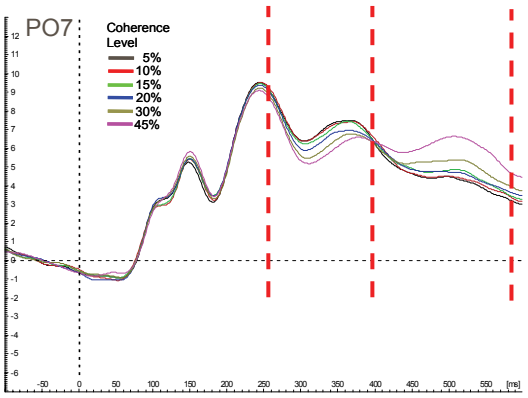


Fig.1 Grand-Averaged waveforms on the PO7 electrode

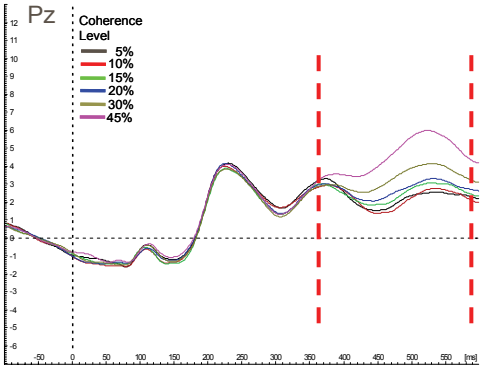


Fig.2 Grand-Averaged waveforms on the Pz electrode

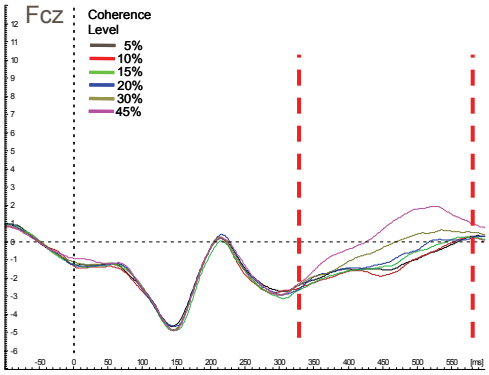


Fig.3 Grand-Averaged waveforms on the Fcz electrode

This work was supported by grant from the Hungarian Scientific Research Fund (T048949) to Z.V.

I. Kóbor is with the Faculty of Information Technology, Pázmány Péter Catholic University, Budapest, H-1083, Hungary (email: kobor@itk.digitus.itk.ppke.hu)

Z. Vidnyánszky is with Neurobionics Research Group, Hungarian Academy of Sciences –Pázmány Péter Catholic University – Semmelweis University, Budapest, H-1094, Hungary (email: vidnyanszky@digitus.itk.ppke.hu).

Neural Dynamics of Shape-specific Visual Information Processing in Amblyopia

Judit Körtvélyes
(Supervisor: Zoltán Vidnyánszky)
kortvelyes.judit@itk.ppke.hu

***Abstract*— Amblyopia is a visual disorder starting at early childhood. It has been shown recently that in amblyopia - in addition to the well known deficits in contour integration, encoding of visual position and motion perception – there is also an impairment of high-level shape processing. It was also suggested that the impairment of shape perception is primarily confined to the fovea and that it might be specific for the processing of visual faces. We used event related potentials (ERP) to investigate the dynamics of shape processing in amblyopia in case of foveal as well as perifoveal presentation of the visual stimuli. We also aimed at uncovering the category specificity of the possible impairment of the shape processing in amblyopia by using human faces as well as human hands as stimuli. Our preliminary results - in accordance with previous findings – revealed that deficits in shape processing are indeed confined to the central, foveal vision in amplyops. Furthermore, our findings also suggest that the deficit in amblyopia is not restricted to processing of visual faces but it extends to other categories, including processing of human hands.**

REFERENCES

[1] Y. Lerner, P. Pianka, B. Azmon, H. Leiba, C. Stolovitch, A. Loewenstein, M. Harel, T. Hendler, and R. Malach, "Area-Specific Amblyopic Effects in Human Occipitotemporal Object Representations," *Neuron*, vol. 40, pp. 1023–1029, 2003.

[2] Y. Lerner, T. Hendler, R. Malach, M. Harel, H. Leiba, C. Stolovitch, and P. Pianka, "Selective fovea-related deprived activation in retinotopic and high-order visual cortex of human amblyopes," *NeuroImage*, vol. 33, pp. 169–179, 2006.

[3] S. P. McKee, D. M. Levi, J. A. Movshon, "The pattern of visual deficits in amblyopia," *Journal of Vision*, vol. 3, pp. 380-405, 2003.

[4] J. M. Holmes, M. P. Clarke, "Amblyopia," *Lancet* vol. 367. pp. 1343-51, 2006.

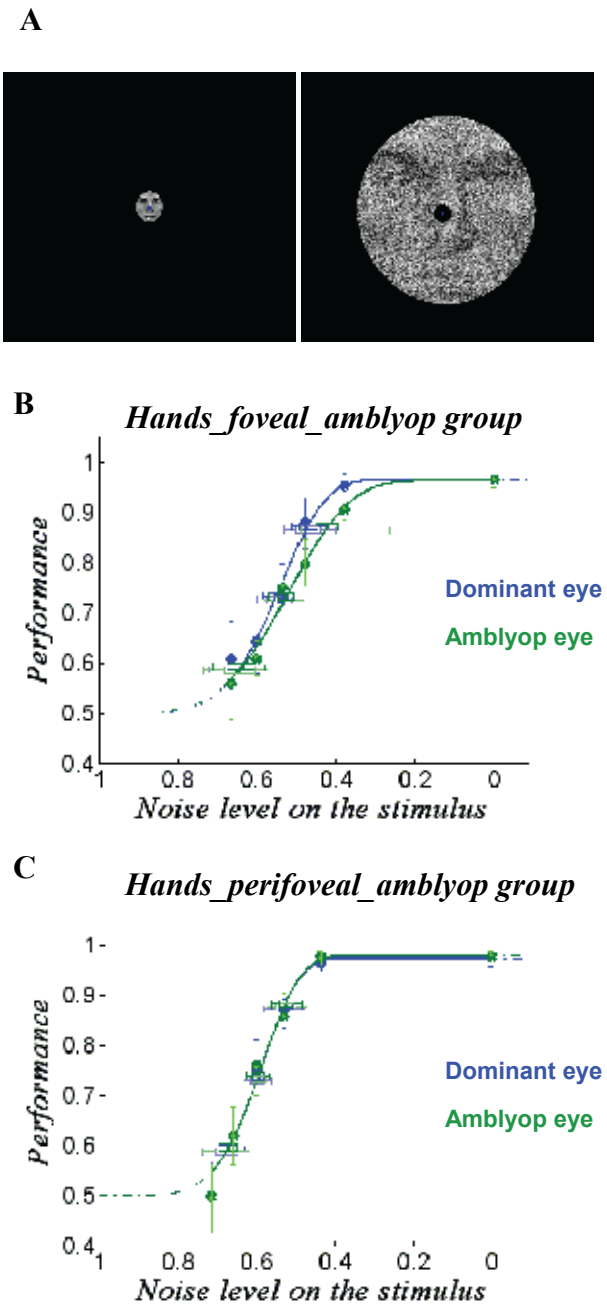


Fig. 1.: In the face hand discrimination task we used stimuli in two different sizes: foveal (visual angle was 2 degree) and perifoveal (visual angle was 15 degree) with 1,5 degree hole placed on the fixation spot for better isolation of the foveal activation (A). The amblyops performed worst with their amblyopic eye as compared to the fellow eye when the stimuli were presented foveally (B). Whereas in the case of perifoveal stimuli there was no difference between the amblyopic and fellow eyes of the amblyops (C).

Multiple Object Tracking: a Test for Amblyops and Eye Dominancy in Healthy Controls

Patrícia Domsa
(Supervisor: Zoltán Vidnyánszky)
domsa.patricia@itk.ppke.hu

Abstract—**The goal of our base study was to investigate the efficacy of visual attentional selection in amblyopia using the multiple object tracking (MOT) task, which require selection and tracking of a subset of visual objects in a visual display containing moving identical objects. Fifteen adult patients with unilateral amblyopia and 15 controls were tested monocularly on the MOT task (three conditions, which differed in the speed of the moving objects). It was found that amblyopic patients performed better with their fellow eyes than with their amblyopic eyes at all three speed levels tested. In the patients with amblyopia, there was a significant positive correlation between the Amblyopia Index (visual acuity of the amblyopic eye divided by the visual acuity of the dominant eye) and the MOT performance (performance of the amblyopic eye divided by the performance of the dominant eye in the MOT task) at the low ($p<0,04$) and intermediate ($p<0,01$), but not at the highest speed tested ($p>0,347$).**

Surprisingly, at the highest speed, performance of the control group was also better when the dominant eye was tested as compared to the non-dominant eye. It was also found that amblyopic patients performed worst with their amblyopic eye than controls with their non-dominant eye. The results provide evidence that in amblyopic patients attentional tracking is less efficient in the case of visual information conveyed by the amblyopic eye as compared to that originating from the fellow eye. We suggest, that there is a difference in processing of visual information originating from the dominant and non-dominant eye, which can be revealed both in controls and patients with amblyopia using the MOT task.

In our first control experiment we tested 15 healthy adult patients with artificially deteriorated vision, to explore the effect of impaired vision on MOT task performance. We implemented artificial vision deterioration with glasses – spherical and cylindrical plus lenses. We matched visual acuities to our amblyopic patients. We tested eye dominancy, the patients were tested monocularly, and we used the same MOT task as in the base experiment. To avoid learning effect, we randomized the 'with glasses' and 'without glasses' blocks.

Patients performed MOT task worse with artificially deteriorated vision at all of the three speed levels. At the highest speed, with artificially deteriorated vision, the performance of the non-dominant eye was better. Performance on the MOT task was worse with artificially deteriorated, than with normal vision, but in the healthy patients, there was no correlation between the Amblyopia Index and the MOT performance at none of the three speeds tested.

Artificially deteriorated vision control implies, that vision deterioration can affect MOT performance but the mechanism differs from underlying amblyopia. We suggest, that randomization affected eye dominancy in MOT task. Dominancy is not a wired status in healthy controls, but adjustment is contiguous in the visual system. If visual conditions necessitate eye dominancy can shift.

Index Terms—**amblyopia, multiple object tracking, eye dominancy**

Figure 1. MOT task

REFRENCES

[1] Buckingham, T., Watkins, R., Bansal, P., & Bamford, K. (1991). Hyperacuity thresholds for oscillatory movement are abnormal in strabismic and anisometropic amblyopes. Optometry and Vision Science, 68, 351–356.

[2] Hess, R. F., Demanins, R., & Bex, P. J. (1997). A reduced motion aftereffect in strabismic amblyopia. Vision Research, 37, 1303–1311.

[3] Ho, C. S. & Giaschi, D. E. (2006). Deficient low-level and high-level maximum motion displacement in amblyopic children. Vision Research, submitted, under revision.

[4] Ho, C. S., Giaschi, D. E., Boden, C., Dougherty, R., Cline, R., & Lyons, C. (2005). Deficient motion perception in the fellow eye of amblyopic children. Vision Research, 45, 1615–1627.

[5] Levi, D. M. (1991). Spatial vision in amblyopia. In D. Regan (Ed.), Spatial Vision (pp. 212–238). London: MacMillan.S. Magnussen, "Low-level memory processes in vision," *Trends Neurosci.*, vol. 23, pp. 247-251, 2000.

High-fidelity Short-term Memory for Changeable but not for Invariant Facial Attributes

Éva M. Bankó,
(Supervisor: Zoltán Vidnyánszky)
banko.eva@itk.ppke.hu

***Abstract*— Processing of invariant (e.g. identity) and changeable (e.g. emotional expressions) facial attributes takes place to a large extent independently, in different visual pathways [1,2].. However, the way in which coding in these two processing routes differs is still an unresolved question [3]. Here we investigated how efficiently can humans compare facial emotions and facial identity as a function of the delay between the presentation of the two faces. It was found that separating the face stimuli by several seconds impairs observers' ability to discriminate their identity but not their emotional expressions (happiness or fear). We also show that emotion discrimination is just as good when the observers perform the task in the first time with novel face stimuli as it is after several hours of practice, whereas fine grained facial identity discrimination needs practice. By revealing high-fidelity visual short-term memory for facial emotions but not for facial identity, these results provide evidence for a clear functional dissociation in the processing of changeable and invariant facial attributes. Our findings also imply that face processing in the human brain has adapted to the statistics of the visual input associated with the different facial attributes.**

REFERENCES

[1] J. V. Haxby, E. A. Hoffman and M. I. Gobbini, "The distributed human neural system for face perception," *Trends Cogn. Sci.*, vol. 4, pp. 223-233, 2000.

[2] A. J. Calder and A.W. Young, "Understanding the recognition of facial identity and facial expression," *Nat. Rev. Neurosci.*, vol. 6, pp. 641-651, 2005.

[3] S. Magnussen, "Low-level memory processes in vision," *Trends Neurosci.*, vol. 23, pp. 247-251, 2000.

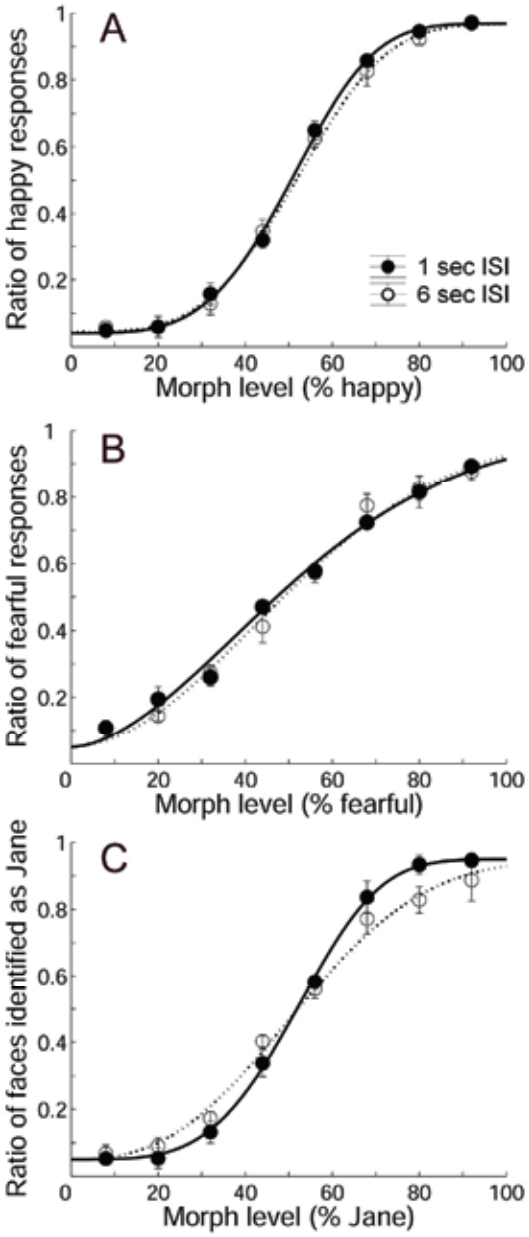


Fig. 1. Effect of ISI on the performance of facial emotion and identity discrimination. Insertion of a long delay of 6 sec between the faces also had a significant effect on identity (C) but not on emotion (A and B) discrimination performance which manifested itself in the change of slope of the psychometric functions. This may imply better storage capacity for changeable versus invariant facial attributes. Data are presented as mean of six subjects ± SEM.

Energy Balancing Cooperative Communications for Wireless Sensor Networks with Biomedical Implants

Barnabás Hegyi
(Supervisor: János Levendovszky)
hegyi.barnabas@itk.ppke.hu

***Abstract*— The paper is concerned with developing cooperative communication schemes with energy balancing for wireless sensor networks (WSNs) in a fading environment. Our objective is to increase the lifespan of a bottleneck node (BN) for which energy consumption is of crucial importance, while maintaining energy balance between the BN and the rest of the network composed of less energy constrained nodes (distributed subject to a 2D Poission distribution). This network scenario is primarily motivated by biomedical applications where, for instance, recharging of a wireless implant is out of reach (e.g. pacemakers, esophageal pH sensor, etc.).**

The gain achieved by the different collaborative strategies and power allocations has also been evaluated and compared via extensive simulations, which has demonstrated that the lifetime of the BN can considerably be increased by the proposed energy balancing, cooperative communication schemes.

"*Denise and her husband Mitch are at opposite ends of a living room at a crowded party. Denise tries to attract Mitch's attention and shouts something at him. All Mitch can hear is the word 'Let's.' Celine, in the middle of the room, who overhears Denise and notices their predicament, repeats to Mitch the part she hears: 'Go home.' This time, all Mitch hears is the word 'home.' Mitch finally figures out that his wife wants to go home.*" "This analogy from everyday life vividly portrays the essential element of cooperative wireless communications, namely, utilizing information overheard by neighboring nodes to provide robust communication between a source and its destination." (Liu et. al. 2006).

In our work we consider anycast cooperative communication schemes with a random (2D Poisson) node distribution and energy balancing between the source and relay nodes. According to the best of our knowledge this is the first contribution in the field of collaborative communications treating these (three) considerations at the same time.

In the proposed communication schemes the BN packet is simultanenously received by a couple of neighboring nodes in the proximity of the BN. Then a couple of these nodes are selected for collaboration in the course of which they may amplify and forward their received analogue signal to each other. In case of the repetition based version of the schemes, the collaborating nodes relay the packet *one after the other*, while for the simultaneous transmission based version, the packet is relayed *at the same time*. The scenario is termed as anycast since the BN packet is not intended for a specific node in the network, i.e. any of the cooperating nodes can be considered as destination.

The network scenario considered in the paper is primarily motivated by biomedical applications where recharging the battery-operated wireless implants is out of reach. By applying the cooperative communication scheme described above, one can significantly reduce the transmission power of the wireless implant (the BN) and thus its lifetime can be prolonged.

We investigated the problem of optimal power allocation among relay nodes and derived the corresponding exact (if applicable) optimal power allocation formulas along with some high SNR ratio approximations.

Figure 1 shows the lifetime gain as a function of the number of relay nodes both for the repetition- and the simultaneous transmission based schemes with optimal power allocation. The value given in dB in the legend is the total transmission power of the relays expressed in terms of the transmission power of the BN. The curve labelled as MRC corresponds to the hypothetic case, when the relay nodes have unlimited transmission power.

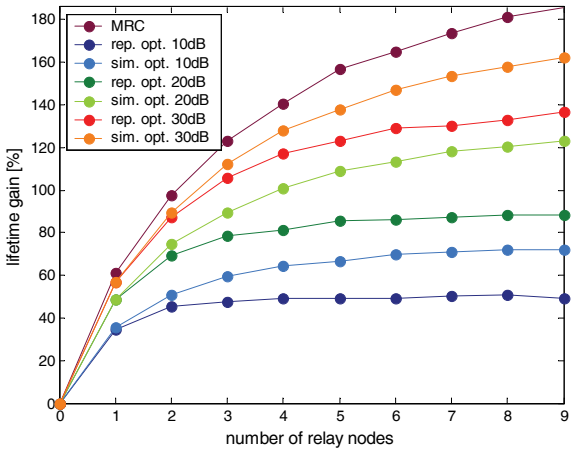


Fig. 1. Lifetime gain as a function of the number of relays.

Challenges of In-body Data Collecting Using Body Sensor Networks

András Bojárszky
(Supervisor: János Levendovszky)
bojarszky.andras@itk.ppke.hu

***Abstract*— Through development of the wireless sensor technologies, there are new possibilities for their usage. One of these possibilities is to implant sensors into bodies, improving the efficiency of information gathering. With the help of biosensor networks doctors can receive information about the status of the patient even if they are at home, not in the hospital. Thisway a significant decrease of hospital costs can be reached while the accuracy of information gathering increases. However, we are facing many questions in this area that needs to be resolved, which hinders the daily usage of implented sensors.**

I. Introduction

THE implented sensor forwards the measured data to a processing unit (for example PDA), which sends them over to the doctor with the help of already known technological channels (for example internet). The PDA can already perform pre-processing, in case of danger it can contact the medical supply. In the developed countries for instance, the cause of death in about one third of the population is a cardiovascular illness. A sensor network of a pace-controller/observer and a sphygmomanometer can inform the patient or the doctor about the coming heart-attack when the patient has no symptoms yet. It is a big problem at the large scale of medical measurements that a patient has to be monitored through several days or weeks, until enough information is gathered for a diagnosis. We can reach this with implanted sensors without changes in the life quality of the patient.

II. Challenges in BSN technology

In case of implented wireless sensors we are facing new challenges, which is also highlighted by the 802.15.4 IEEE standard.

- Scale – the functional distance is of the order of mm/cm in stead of m/km distances
- Number and function – less number of nodes inside the body, but a node performes more measurements (multiple task)
- Dynamics – the smaller sensors are facing new challanges within the calculable environment of the human body
- Event detection – while measurement of the biological signs of the body neither mistakes (false measurement, false information forwarding) nor bad timing is acceptable
- Data protection – the forwarded data needs to be high security level confidential because of the protection of personal data
- Power supply – in case of implanted sensors the phisical change of the exhausted batteries is rather problematic
- Energy Scavenging – kinetic or heat-energy can serve as a renewable energy resource inside the body
- Data transfer – the false information forwarding can be offset with the big number of nudes in the traditional WBS networks. In case of BSN it is not possible, therefore the successful information forwarding is of primal importance

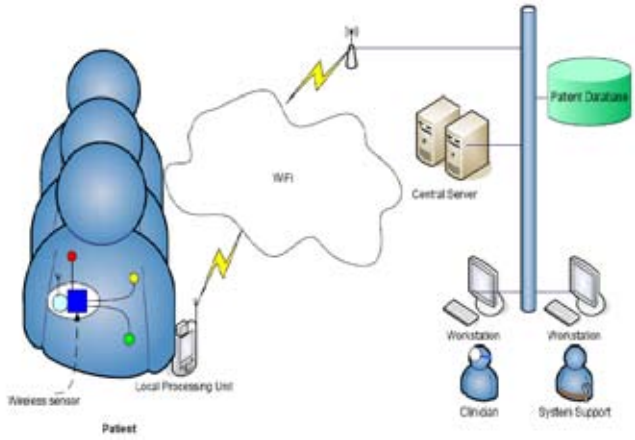


Figure 1 A typical BSN network

We need to consider the followings while planning the communication protocol for BSN:

- number of nodes
- tissue around the nodes
- required energy for transmitting
- occurancy of measurement
- compression of data sent
- securing the data sent

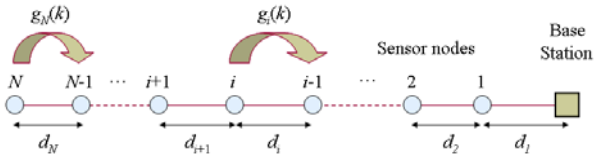
Based on these information we can precisely predict the lifetime of the network.

Energy Balancing by Combinatorial Optimization for Wireless Sensor Networks

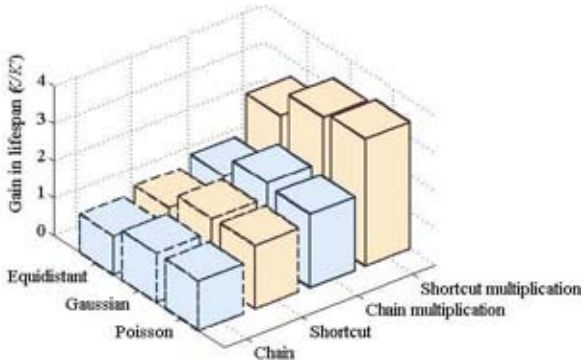
Balázs Karlócai
(Supervisor: János Levendovszky)
karlocai.balazs@itk.ppke.hu

***Abstract*— In this paper some novel protocols are developed for wireless sensor networks (WSNs) in order to ensure reliable packet transmission and maximize lifespan at the same time.**

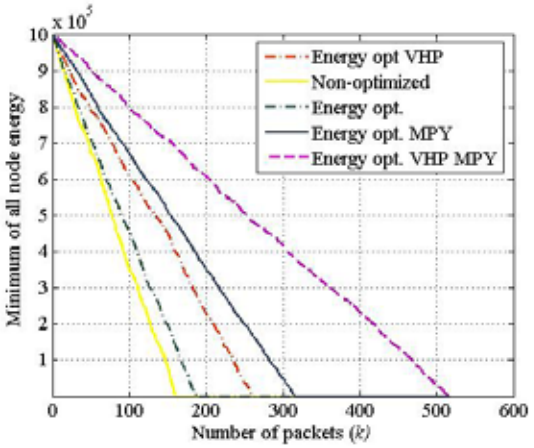
The optimal transmission energies are derived which guarantee that the packets are received by the Base Station (BS) with a given probability subject to achieving the longest possible lifespan.



The optimization has been carried out for the chain protocol (when nodes are forwarding the packets toward the BS via the neighboring nodes) and for the shortcut type of protocols (when nodes can transmit packets for further nodes not only to their nearest neighbors). The new results have been tested by extensive simulations which demonstrated that the lifespan of WSN can significantly be increased by the new protocols.



In this paper, novel energy balancing packet forwarding methods have been developed to maximize the lifespan of WSNs and to ensure reliable packet transfer at the same time. We have optimized the transmission energies of the nodes depending on the source of the packet in order to minimize the energy consumption of the bottleneck node (the node with the lowest available energy) subject to satisfying a given the reliability constraint . Three scenario has been studied extensively: (i) the traditional chain protocol (nodes are passing the incoming packet to their neighbors closer to the BS) ; (ii) the random shortcut protocol (nodes make random choices subject to an optimized probability mass function) whether to forward the packet to the neighboring node or sending it directly to the BS); (iii) the hopping-ahead-in-any-possible-manner protocol (nodes can send the packet to any possible node ahead in the chain towards the BS). Each of this protocol has been extended to include multiple packet sending in order to further lower the energy consumption.
The underlying protocol optimization was reduced to a constrained optimization problem which has been solved by a stochastic search algorithm.



The performance of the protocols have been analyzed and compared to each other in the case of a 10-node WSN where the nodes were distributed in equidistant or random fashion subject to Gaussian and Poisson distribution. The reliability constraint was set 0.961. Form the performance analysis one can infer, that the HHAPW and random shortcut protocol provides the longest lifespan. Furthermore, the protocols with packet multiplication achieve better longevity than the protocols without packet multiplication (the gain is approximately 3,4x). Finally, it has turned out that the highest performance in prolonging the lifetime was always obtained in the case nodes distributed subject to Poisson distribution.

Vision Based Human-Machine Interface via Hand Gestures

Norbert Bérci
(Supervisors: Péter Szolgay and Tamás Roska)
berci.norbert@itk.ppke.hu

***Abstract*—In the current paper we outline the first stage of a hand gesture recognition system with the primary purpose of replacing classical input peripherals. The system works by visual input and the algorithm has been implemented on the CNN [1] based Bi-i visual processing architecture [2], [3]. Its properties allow it to be used for security, military, medical, surgery and public media applications as well.**

A. Building Blocks of the Algorithm

As explained in [4], there is a huge difference between a static posture and a dynamic gesture. We have been dealing with gestures only, postures were not considered, but the system may be able to recognize some types of them also.

The implemented algorithm is built up from a preprocessing stage, a feature extraction stage and an object tracking stage. Its flowchart is shown on Fig. 1.

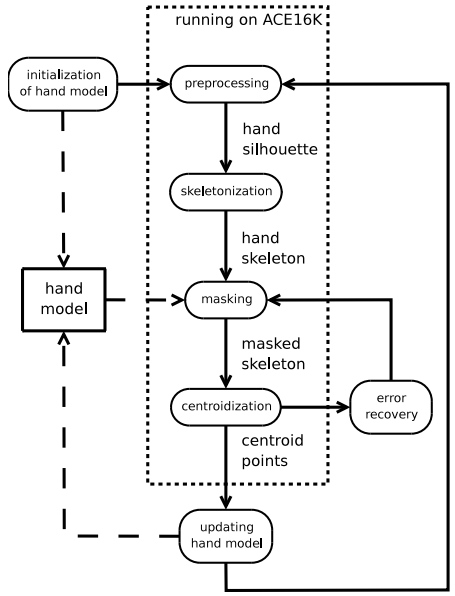


Fig. 1. The flowchart of the implemented algorithm. Blocks inside the dashed line run on the ACE16K visual processor, all the others have been implemented on the DSP.

B. Features and the Hand Model, Initialization

Considering the abilities of the used architecture and the properties of the anatomy of the hand, our implementation currently stores fingertip position and finger direction which will be extended to finger joints, and to a wireframe hand model. Recognized features can be easily and quickly mapped to the hand model, and it also efficiently supports the recognition.

By initializing the hand model appropriately in the start position, valuable information is collected. Importantly, this makes it possible to transform the recognition process to a tracking task: no need to find the hand in every frame, but be able to follow it while moving. Partial or defective information is enough later.

C. Feature Extraction

Only the fingertip carries valuable information, so after masking it out, its direction and center position is calculated from the centroid point of the masked skeleton and the mask itself, and the hand model is updated accordingly. The arisen difficulties have been mainly avoided by the smart placement of the mask which is supported by the hand model.

D. Preprocessing

In order for the skeletonization to work properly, a well separated silhouette of the hand should be available. This preprocessing is currently composed of a static image correction to remove the uniformity artifacts of the cells of the ACE16k chip [5] caused by the manufacturing process, then a diffusion to make the contour smooth and last a threshold function to convert the image to binary.

E. Results, Comments

The system is able to process optical flow with human-hand real-time performance in a robust way. Further research needed to extended its capabilities, but a mouse replacement device is almost available.

The original paper has been accepted for the 18^{th} European Conference on Circuit Theory and Design, ECCTD'07 (paper number 5213).

REFERENCES

[1] L. O. Chua and T. Roska, *Cellular Neural Networks and Visual Computing*. Cambridge, UK: Cambridge University Pess, 2002.

[2] Á. Zarándy and C. Rekeczky, "Bi-i: a standalone cellular vision system, Part I. Architecture and ultra high frame rate processing examples," in *Proc. 8th Int. Workshop on CNNs and their Appl. (CNNA)*, Budapest, Hungary, Jul. 22–24, 2004, pp. 4–9.

[3] ——, "Bi-i: a standalone cellular vision system, Part II. Topographic and non-topographic algorithms and related applications," in *Proc. Int. Workshop on CNNs and their Appl. (CNNA)*, Budapest, Hungary, Jul. 22–24, 2004, pp. 10–15.

[4] J. J. LaViola, "A survey of hand posture and gesture recognition tehniques and technology," Department of Computer Science, Brown University, Providence, Rhode Island, USA, Tech. Rep. CS-99-11, 1999.

[5] G. Liñán, R. Domínguez-Castro, S. Espejo, and Á. Rodríguez-Vázquez, "ACE16k: A programmable focal plane vision processor with 128x128 resolution," in *Proc. Eur. Conf. on Circ. Theory and Design (ECCTD)*, Espoo, Finland, Aug. 28–31, 2001, pp. 345–348.

Statistical Connections Between Image Features

Dániel Szolgay
(Supervisors: Tamás Szirányi and Tamás Roska)
szolgay.daniel@itk.ppke.hu

I. Introduction

The algorithm presented here aims to help the existing content based image retrieval (CBIR) methods by giving a new description of the images, based on the eye movement information.

There are three common types of eye movement: fixations, pursuit and saccades. For us the most interesting type of the eye movements is the fixation, because we receive practically all the visual information during the fixations.

To track the eye movements of a human observer, we used the system of the SensoMotoric Instruments® (SMI) [1], called iView X. This system can register the position of the eye with the precision of 0.5 degree in every 50 milisecundum. After we have the eye positions we can calculate the coordinates of the fixation points.

II. Our Approach

A. The CBIR and the Eye Movement Information

Our aim is to give a new description of the picture based on the low level features of the picture and the eye movement information belonging to the image. Of course we don't have this eye movement information for all the images, that is why we want to collect statistics from the available information.
First the low level features of the neighborhood of the fixation points are calculated, and these parameters are collected into a feature vector. Then the feature vectors are put into different clusters and the transition probabilities between the clusters are calculated. In the future we want to use these probabilities to help making a good description of the picture.

B. Feature Extraction and Clustering

To describe the neighborhood of the fixation points, low level image features were used. These features were extracted from a 120 x 120 sized square neighborhood of each fixation points. In this territory color and texture features were calculated.

1) Color Features: The color features were extracted using the HLS color space, where L encodes the luminance information and H (hue) and S (saturation) encodes the color information. As a color feature the average value for each of the three color coordinates were calculated.

2) Texture Features: For describing the texture properties of the given area, we used the three most significant of the six parameters presented in [2]. These parameters were experimentally tested and were found to be very important for the human observer. The three chosen parameters are the coarseness, the contrast and the directionality.

The *coarseness* of a texture gives information about the size of the texture element. Higher coarseness value means rougher texture. There are many definitions for the *contrast*of a picture. In our case the Tamura feature for contrast seemed to be the most adequate for the task. For describing *directionality*, which means the presence of the orientation, but not the orientation itself, a modified version of the Tamura feature [3] was used. Beside these three Tamura feature, the edge intensity of the ROI was calculated too.

Many other features are used in the field of CBIR. Later I would like to complete the set of features with an elemental and well known feature, the orientation.

For clustering the received and normed feature vectors, the k-means [4] algorithm was used, which is a very simple unsupervised learning algorithm.

After all the feature vectors are clustered, the transition probability matrix was calculated for the clusters. This is an NxN matrix, where N is the number of the clusters. An element a_{ij} $(i, j \in 1, 2 \dots N)$ represents the probability that the eye moves from a point, clustered as i to a point, clustered as j. The sum of each column is 1.

C. Image Description

To describe the picture we need a model which considers the properties of the picture along with the eye movement statistics. A possible way to reach this goal is to use the transition probability statistics as weights to compute the transitions' probabilities of the new image. But there are many other possibilities, for example we should consider using Hidden Markov Model (HMM) to solve the task.

III. Conclusion and Future Work

In the future we want to add new features (e.g. orientation) to the feature vector.
We need more eye movement information so we must make additional measurements. Of course we must design a model, which describes the pictures based on the collected information.

References

[1] SensoMotoric Instrument http://www.smi.de/iv/
[2] H. Tamura, S. Mori, T. Yamawaki. "Textural Features Corresponding to Visual Perception". IEEE Transaction on Systems, Man, and Cybernetcs, Vol. SMC-8, No. 6, pp. 460472, June 1978.
[3] Xing-Jian He, Yue Zhang, Tat-Ming Lok, and Michael R. Lyu, "A new feature of uniformity of image texture directions coinciding with the human eyes perception," Lecture Notes in Artificial Intelligence, Springer-Verlag, LNAI 3614, pp. 727 - 730, 2005.
[4]] J. B. MacQueen (1967): "Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, 1:281-297

A Multi-Layer MRF Model for Object-Motion Detection in Unregistered Airborne Image-Pairs

Csaba Benedek
(Supervisor: Tamás Szirányi)
benedek.csaba@itk.ppke.hu

***Abstract*—A probabilistic model has been given for automatic change detection on airborne images taken with moving cameras. A novel three-layer Markov Random Field (MRF) model has been proposed which integrates two measurements, meanwhile it ensures the spatial smoothness of the segmentation.**

***Index Terms*—Change detection, aerial images, camera motion, MRF**

I. Introduction

Change detection is an important early vision task for many image processing application. The present Ph.D. topic attacks three tasks of this problem family [1].

- **Task 1**: Separation of foreground, background and moving shadows in surveillance videos captured by static cameras.
- **Task 2**: Moving object detection in airborne images captured by moving cameras.
- **Task 3**: Structural change detection in registered airborne images captured with significant time difference.

This report deals with the second task, addressing the problem of extracting the accurate silhouettes of moving objects or object-groups in images taken by moving airborne vehicles in consecutive moments.

II. Methods and experiments

The procedure needs an efficient combination of camera motion compensation and frame differencing. To ensure robustness, an unsupervised coarse (2D projective) matching is used instead of a precise image registration. The challenge of the proposed model is to eliminate the registration errors, noise and the parallax artifacts caused by the static objects having considerable height (buildings, trees, walls etc.) from the difference image. In the error-eliminating step, a Bayesian approach is used: the optimal motion map is obtained as a maximum a posteriori (MAP) estimate like in [2]. We derive features describing the background membership of a given image point in two independent ways, and develop a three-layer Bayesian (Markov Random Field [3]) labeling model to integrate the effect of the different features, meanwhile the spatial smoothness of the segmentation is ensured. The efficiency of the method has been validated through real-world aerial images, and its behavior versus three reference methods has been quantitatively and qualitatively evaluated. Considering the ratio of correct pixel classification, the proposed model outperfoms the older ones with about $5-10\%$, using a database containing 83 image pairs from three different environments.

The introduced work has been done in international cooperation. Collaborators Tamás Szirányi, Zoltan Kato (University of Szeged) and Josiane Zerubia (INRIA, France) are co-authors of this material and the corresponding publications

This work was partially supported by the EU project MUSCLE (FP6-567752), the Hungarian Scientific Research Fund (OTKA #49001) and the Hungarian R&D Project ALFA. The author would like to thank the MUSCLE Shape Modeling E-Team for financial support and to Xavier Descombes from INRIA for his kind remarks and advices.

III. Author's publications and further activities

In this academic year, three refereed international conference publications of the author have been accepted [4] [5][6], while three journal publications are under review at the end of June 2007. A detailed description about this work in topic is given in [7].

The author had a two week visit at ARIANA group (joint research group INRIA/CNRS/UNSA) in France, and hold a seminar at the Ramon Llull University, Barcelona.

References

[1] **Cs. Benedek**, "Markovian approaches regardig change detection on video sequences and airborne images," in *Ph.D. Proceedings of the Pazmany Peter Catholic University*, Budapest, 2006, pp. 132–136.

[2] D. Farin and P. With, "Misregistration errors in change detection algorithms and how to avoid them," in *Proc. ICIP*, Genoa, Italy, 2005, pp. 438–441.

[3] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. PAMI*, vol. 6, no. 6, pp. 721–741, 1984.

[4] **Cs. Benedek** and T. Szirányi, "Markovian framework for structural change detection with application on detecting built-in changes in airborne images," in *Proc. IASTED International Conference on Signal Processing, Pattern Recognition and Applications (SPPRA)*. Innsbruck, Austria: ACTA, February 2007.

[5] **Cs. Benedek**, T. Szirányi, Z. Kato, and J. Zerubia, "A multi-layer MRF model for object-motion detection in unregistered airborne image-pairs," in *Proc. IEEE International Conference on Image Processing (ICIP)*. San Antonio, Texas, USA: IEEE, Sept. 2007.

[6] **Cs. Benedek** and T. Szirányi, "Color models of shadow detection in video scenes," in *Proc. International Conference on Computer Vision Theory and Applications (VISAPP)*, vol. IFP/IA. Barcelona, Spain: INSTICC, March 2007, pp. 225–232.

[7] **Cs. Benedek**, T. Szirányi, Z. Kato, and J. Zerubia, "A three-layer MRF model for object motion detection in airborne images," INRIA Sophia Antipolis, France, Research Report 6208, June 2007. [Online]. Available: https://hal.inria.fr/inria-00150805

Speaker Independent Voice to Animation Conversion

Gergely Feldhoffer
(Supervisor: György Takács)
flugi@digitus.itk.ppke.hu

***Abstract*—A speaker independent training method is presented for continuous voice to facial animation systems. Temporal structure of multimodal speech is discussed. An audiovisual database with multiple voices and only one speaker's video information was created using dynamic time warping. The video information is aligned to more speakers' voice. The fit is measured with subjective and objective tests.**

I. Temporal scope

The synchrony of the audio and video data is checked by word papapa in the beginning and the end of the recording. The first opening of the mouth by this bilabial can be synchronized with the burst in the audio data. This synchronization guaranties that the pairs of audio and video windows were recorded in the same time, which gives an input and an output for a back-propagation neural network.

For the best result the neural network has to be trained on larger temporal scope of audio information. The mutual information between audio and video data was measured for different time shifts. 200 ms delay is advised. This can be explained as the speech process is a predicting mental process which moves the mouth according to not only the actual but the next approximately 200 ms of voice since the facial muscles are slower than the muscles in the tongue.

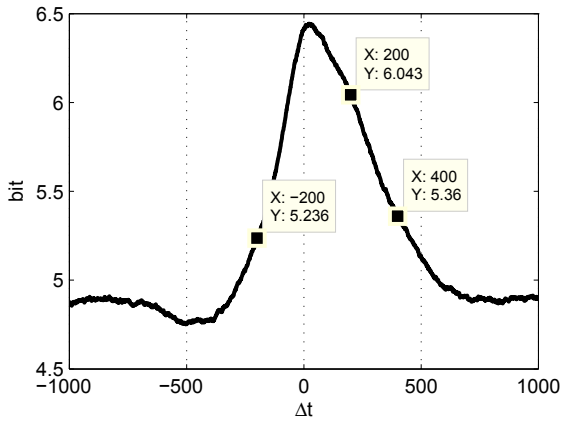


Fig. 1. Sum of the mutual information curves of the independent components. This independent components were taken from the audio and the video information with ICA.

Therefore we are using a 5 element sized queue of audio feature vectors as the input of the neural network, and the corresponding PCA vector from the video data.

II. Speaker independency

The described system works on well defined pairs of audio and video data. This is evident if the database is a single person database. If the video data belongs to a different person, the task is to fit the audio and the video data together.

The text of the database was the same for each person. This allows the aligning of audio data between speakers. We used the Dynamic Time Warping technique for this, which is a widely used method in speech recognition on small dictionaries. Usually for speech recognition purposes this is a distance estimation method using cumulative distance sums. We used it to extract the best match of the windows between the two audio data.

This matching is represented by index arrays which tell that speaker A in the i moment says the same as speaker B in the j moment. As long as the audio and video data of the speakers are synchronized, this gives the information of how speaker B holds his mouth when he says the same as speaker A speaks in the moment i.

With this training data we can have only one person's video information which is from a professional lip-speaker and in the same time the voice characteristics can be covered with multiple speakers' voices.

40 sentences of 5 speakers were used for this experiment. We used the video information of speaker A as output for each speaker, so in the case of speaker B, C, D and E the video information is warped onto the voice. We used speaker E as test reference.

Fig. 2. Training with speaker A, A and B, and so on, and always test by speaker E which is not involved in the training set

NP Extraction Techniques for NP Alignment Purposes

Gábor Pohl
(Supervisor: Gábor Prószéky)
pohl.gabor@itk.ppke.hu

I. INTRODUCTION

In this abstract, automatic noun phrase (NP) extraction techniques are comparatively evaluated considering their adequacy for automatic NP alignment purposes. The main motivation behind research for better NP extraction techniques for alignment purposes is MorphoTM, an experimental English-Hungarian translation memory (TM).

In MorphoTM not only whole sentences are searched in the memory. NPs and the sentence skeleton (derived from the sentence by substituting NPs with symbolic NP slots) are also searched in the database, and their most probable translations are combined to form a possible translation sentence. Hence the recall of the translation memory is increased by suggesting translations built up from NP and sentence skeleton translations looked up separately. Therefore, NPs of the stored sentence pair have to be selected and aligned automatically.

In the past two years, a lexical feature based NP alignment technique was developed for MorphoTM. This article is focused on the evaluation of automatic NP extraction techniques, more precisely NP extraction techniques of the translation sentence, because NPs of the source sentence have to be selected by a full sentence parser in MorphoTM.

II. MEANS OF TRANSLATION SIDE NP EXTRACTION

Available options for translation side NP extraction include deep sentence parsing, translation guided shallow NP parsing, and combinations of the two techniques.

Deep parsing of the whole translation sentence is not a real option in MorphoTM, because this technique is too slow and its recall is not high enough. Therefore, in the past years, the much faster translation guided shallow NP parsing technique for parallel texts was developed. This technique maps the words of a parsed source language NP to the words of the translation sentence using dictionary matching of word stems and cognate matching. Matched words can occur more than once in the target sentence, therefore the shortest span in which all the matching terms are found is selected as an NP skeleton. The NP skeleton is expanded to a full target language NP by a shallow NP grammar and considering the part-of-speech of unmatched words in the source NP.

This technique is fast, and its recall is high enough, but its precision is questionable, therefore we tried to improve it by combining it with a deep parser. We tried two possible combinations: (1) expanding the NP skeleton with a deep parser, (2) checking the result of the translation guided shallow NP parsing technique with a deep parser.

III. MEASUREMENT SETUP

We compared not only the raw NP extraction quality of translation guided shallow NP parsing and its combinations with a deep parser, but NP alignment results were also compared with each technique.

Our NP alignment algorithm extracts various lexical features from the compared NP pair and a corpus trained classifier decides whether the processed NP pair should be stored as a translation pair. In order to avoid classification bias for one of the compared NP extraction methods, the classifier was trained and evaluated separately for each NP extraction technique.

A small parallel corpus of 100 English-Hungarian sentence pairs was selected for the measurements. NPs of the English sentences were extracted by MorphoLogic's MetaMorpho English parser [5]. Then three parallel corpora were created by selecting the possible Hungarian NPs with the three compared techniques. The Hungarian MetaMorpho parser was applied as the deep parser of the translation side. The three parallel corpora were hand-aligned to measure the raw NP extraction performance of the algorithms and in order to create the three separate training and testing sets for the classifier of the NP aligner algorithm. NP alignment results were measured with 10-fold cross validation.

IV. MEASUREMENT RESULTS

Measurement results showed that the translation guided shallow NP parsing algorithm is an appropriate means of NP extraction for alignment purposes. Checking its output with a deep parser, yields higher precision but with reduced recall, both in raw NP extraction and in NP alignment results. (NP alignment results are shown in Table I below.)

The lower recall of the combined technique can be solved if the NP candidates failing the deep parser test are only marked less probable and are not rejected in MorphoTM.

TABLE I: NP PAIRS STORED IN TRANSLATION DATABASE

Method	Correct Pairs	Incorrect Pairs	Precision
TGSNPP	**179**	25	87.7%
NPS+DP	126	19	86.9%
TGSNPP+DP	112	10	**91.8%**

TGSNPP = Translation Guided Shallow NP Parsing, NPS+DP = NP skeleton expansion with deep parser, TGSNPP+DP = NP candidates extracted by TGSNPP are checked by deep parser. TGSNPP+DP has the best precision, but its recall (number of correct NP pairs stored in the database) is lower. NPS+DP performed poorly.

„Mazsola” – a Tool for Investigating Complement Structure of Hungarian Verbs

Bálint Sass
(Supervisor: Gábor Prószéky)
sass.balint@itk.ppke.hu

We should know the complement structure of verbs first, if we want to say anything about the meaning of a sentence. I present a new research tool called „Mazsola” (“raisin” in English) for investigating complement structure of Hungarian verbs. In Hungarian the function of a complement is determined by the case marker at the end of the complement. This allows broadly free complement order. Complements are usually phrases, a complement can be represented as the lemma of the head-word of the phrase plus the case of the phrase. Accordingly, the simple model of a Hungarian sentence looks like this:

verb + phrase(lemma + case)–list

A *complement frame* consists of a verb and a list of complements which occurs with the verb together. *Positions* of frames are morphosyntactically defined positions, characterized by the case marker. The *complement structure* of a given verb consists of the union of all complement frames of that verb. The example in Fig. 1 also shows that different frames often represent different meanings. There are *fix* and *free* positions in frames. If the lemma is specified, then the position is fix, if any lemma can fill in the position, then it is a free position.

To have a picture about how complement frames work in natural language text we should process any natural language text and extract these frames. I worked out a rule based method for completing this task. Frames are extracted on the basis of punctuation and conjunction patterns. Recall and precision of this method is around 85 percent, but it can be said that this is enough for such a tool which I present now.

The „Mazsola” tool is a special corpus query tool, where the typical research question can be: What are the salient lemmas which can fill in a free position of a given verb frame? Saliency is measured by a specially adjusted version of the classical mutual information. According to this measure a lemma is salient if it occurs more frequently with the frame than it is expected, and the lemma itself is frequent.

On the user interface you can specify the corpus, a verb and

‘von -t kétség-bA’	(to question sg)
‘von váll-t’	(to shrug one’s shoulder)
‘von -t’	(to pull sg)
‘von felelősség-rA -t’	(to call sy to account)

Fig. 1. Four main frames of the complement structure of verb ‘von’. ‘-t’, ‘-bA’ and ‘-rA’ are case markers. (Note: the upper case letter signs a vowel alternation point where the exact vowel is determined by vowel harmony.)

Fig. 2. Answer screen of the „Mazsola” tool.

two complements. In the example in Fig. 2 we are interested in, what are the salient lemmas of the direct object position of the frame ‘kér -t -tÓl’ (ask sy sg). We can see these lemmas below the query form with blue. Salient direct objects are: ‘bocsánat’ (forgiveness), ‘segítség’ (help), ‘elnézés’ (also forgiveness), ‘engedély’ (permission) and so on. The „Mazsola” tool is available for free at the following URL: http://corpus.nytud.hu/mazsola. Try it with username: `itk` and temporary password: `dec14`, and if you are interested send me an email to get a permanent password.

The tool was immediately applied in two running projects. During creating the verbal part of the Hungarian WordNet it was useful in determining the synsets where a particular verb gets into, since different frames often represent different meanings. During the manual building of the lexical database of a Hungarian to English machine translation system, an essential task was to collect important lemmas in a position of a given verb frame. From the machine translation point of view important lemmas are the ones, for which the verb frame has a special translation. In most cases they are exactly the same as the salient lemmas provided by the „Mazsola” tool.

Similar complement structure can entail similar meaning. To test this hypothesis I created verb classes based on the salient lemmas filling particular positions of verbs. Clustering was done by the k-means algorithm. Results show that this method is useful in capturing near-synonyms.

Evaluating Ontology Development through Information Extraction and Coreference Resolution

Márton Miháltz
(Supervisor: Gábor Prószéky)
mihaltz.marton@itk.ppke.hu

Representing and using common-sense or domain knowledge of a subject has become increasingly important for natural language processing. This knowledge is commonly referred to as "ontology".

Linguistic ontologies are a special kind of ontologies, where not the knowledge of a domain, but the semantics of natural languages are modeled. The most notable example is WordNet (WN) which consists of nodes called synsets (synonym sets) of nouns, verbs, adjectives or adverbs that are synonymous (interchangeable in a given context without changing the denotational meaning). These synsets represent the various concepts in a language, which are interconnected by various semantic relationships that form the edges of the network. The most important relationship for nominal and verbal concepts is *hypernymy,* the linguistic equivalent of the *is-a* relationship which defines a hierarchical inheritance network. Other types of relationships include *meronymy* (part-whole relationship), *antonymy* (an opposition between concepts over a given domain), *entailment* (among verbs, e.g. *snoring* entails *sleeping*) and many others.

The EuroWordNet (EWN) project extended the WN architecture to a multilingual level, with the synsets of the English WN serving as interlingua (ILI) among the concepts of the various other languages. A common starting set (Common Base Concepts) was implemented in each participating language and then was expanded individually in a top-down manner by each partner. In addition to the 11 European languages covered by EWN, the BalkaNet project several years later introduced connected Wordnets for 5 more Southeast-European languages.

The Hungarian WordNet (HuWN) project follows the BalkaNet project's resources: Princeton WordNet 2.0 as ILI, 8500 base concept synsets as a starting point and the VisDic XML-based ontology/dictionary editor. It now contains 40,000 synsets in the general domain and in the busines/financial domain.

In order to evaluate this new ontology, two experimental applications were implemented that rely on its encoded knowledge.

In the information extraction (IE) evaluation application we process a short business news text and extract structured information that enables future structured storage and querying. The first part of the extraction is the identification of the event type (event frame) that is expressed by the sentence, while the second task is to identify the participants (frame elements) of the event, such as *buyer, seller, price* etc. in a selling event.

In the IE algorithm, a parser first identifies the main syntactic constituents in the input text, then it tries to match these to the elements of the candidate event frames. There are several kinds of constraints that need to be satisfied for a match. Lexical constraints can either be specified by strings, or by synset ids corresponding to hyponym subtrees in the HuWN ontology. Semantic constraints are expressed by so-called semantic meta-features, or basic semantic categories, such as "human", "company", "currency" etc. that are mapped to HuWN synsets and all their hyponyms. There are also syntactic and morphologic constraints, which are checked against the output of the parser and the underlying morphologic analyzer. Finally, the IE engine ranks the candidate event frame matches for the output according to the ratio of event participants matched.

In the anaphora resolution experiment, the task of the system was to process a document and for each non-anaphoric expression referring back, determine which of the previously mentioned entities (possible *antecedents*) it *corefers* with, that is, which denotes the same entity. In the experiment, we focused on types of anaphora that resolvable using knowledge encoded in the HuWN ontology. The resolution algorithm calculates semantic similarity between the anaphoric element and each candidate antecedent by looking at possible paths that connect their corresponding concepts.

In the evaluation, we looked at what the experiences gained from these applications may tell us about ontology development. In short, IE needs more domain knowledge in the form of adding more concepts. On the other hand, coreference resolution needs an ontology more enriched with additional relations between concepts based on our world knowledge.

Studies on Visual Speech and Speech Melody Perception

Tamás Bárdi
(Supervisor: György Takács)
bardi.tamas@itk.ppke.hu

***Abstract*—This paper introduces a psychoacoustic effect of speech prosody perception, which was found in our listening experiments. Uncertainties of the test subjects' pitch perception were measured for different pitch accents: emotional ones and linguistic ones. The experimental investigation was done by measuring just noticeable differences in the pitch contours. Our test subjects were found to be less sensitive for subtle pitch modifications in the case of language driven accents than for emotional stresses. This uncertainty relation is interpreted as an analogous relation with the quantum mechanical one. This paper also summarizes an experimental investigation on the importance of 3D visual information in lip-reading. Our hearing impaired subjects were found to be able to lip-read the visual speech stimuli with reduced depth information just as the original videos.**

***Index Terms*—prosody, pitch, perception, emotion, accent, lip-reading, visual speech**

I. Introduction

Speech prosody and especially intonation is commonly assumed to be suited partly for verbal and partly for emotional communication. This double role makes it difficult to find an intrinsic linkage between prosodic cues and emotions, linguistic and emotional patterns are prone to interfering. The relation of emotions and prosody is in the focus of several recent research studies. In spite of the many efforts in this field, we cannot say what makes intonation really emotive. Natural sounding emotional speech synthesis has not been solved yet.

Modeling F_0 contours for emotional speech synthesis can be done in two ways. Top-down method is starting from psychological evidences, and from biology and physiology in the bottom-up way.

This paper presents a bottom-up strategy to model emotional pitch contours in a novel approach. That is to interpret F_0 generation and pitch perception in the terms of particle/wave duality, exploiting our knowledge on the lateralization phenomena in the cortical level of human speech and music processing. Now, let us consider the emotional stresses in pitch as wave-like objects, and linguistic accents as particle-like object. In quantum mechanics, Heisenberg's uncertainty relation says that energy can be measured more precisely for wave-like objects, and the time can be measured more precisely for particle-like objects. In prosody annotations usually not more than 2 or 3 distinct strengths are defined, while emotional prosody can by tuned much finer by speakers or perceived by the listener. On the other hand, timing of linguistic stresses in pitch is crucial in understanding, while the exact time order of emotional effects is not so important.

Theoretically, if a natural quantization acts under the perception, it can be revealed upon its effect on the expected rates of correct discrimination.

II. Listening Experiments

A special listening experiment was organized to measure uncertainties and find out of the assumed quantization effect in pitch perception. Both were investigated by measuring just noticeable differences in the pitch contours. Pairs of stimuli were played for the subjects, in which a short utterance was repeated two times, but in one of them or both F_0 was modified a little.

Results: the discrimination rates for the neutral samples always get lower than for the same speaker's emotional sample. Whereas 80 cents difference and above in the case of her sad sample was noticed by 100% of the subjects, the discrimination rates for the neutral one are just a random fluctuation around the by-chance rate (50%) up to 100 cents, and the subjects began to notice the difference from 125 cents.

III. Lip-reading Experiments

The importance of 3-dimensional visual information in lip-reading was tested in this experiment. We intended to build up an audiovisual database on the purpose to train facial animation systems and get the animations as lip-readable as we can. This test was destined to decide whether to use 2D or 3D motion tracking in our audiovisual database.

Original and binarized videos were shown to deaf subjects (N=8). 24 short clips were shown with no sound. The task was to recognize what two digit number was said by the speaker. The speaker was a teacher of a secondary school for deaf students; her articulation was clear and favorable to lip-reading.

The result of this experiment was 83.0% correct recognition for videos with natural colours and 82.1% for distorted videos. The difference very, therefore we decided to use only 2D motion tracking in our database.

Ca^{2+} and Voltage Sensitive Imaging in Mouse Visual Cortex

Tamás Fehérvári
(Supervisors: György Karmos and István Ulbert)
fehervari.tamas@itk.ppke.hu

***Abstract*—In April 2007 I began my two-year research scholarship at the Osaka University, Japan. My current activity focuses on Ca^{2+} and voltage sensitive imaging of the mouse visual cortex. The experiment is in its setup phase. The aim of the experiment is the comparison of direct cortical and peripherial stimulation of the visual pathway (retina, optical neuron). The results are intended to be useful for the development of cortical visual prostheses.**

I. Introduction

I spent the first semester of my doctoral studies with acquiring theoretical knowledge and practical experience in electrophysiological experiments. In April 2007 I started my two-year research scholarship at the Graduate School of Engineering, Osaka University, Japan. The host laboratory's activities include *in vivo* and *in vitro* calcium and voltage sensitive imaging in mouse visual cortex. I belong to the *in vivo* group, which is currently in the process of setting up a new experimental environment and practicing the required surgical methods.

It is theoretically possible to restore vision to blind patients who have functional visual cortex by sending the necessary signals to the brain by direct cortical stimulation. The idea of cortical vision prostheses has been proposed decades ago, but practical results have not yet surpassed the evoking of simple phosphenes. Many aspects, like the course of neural response on a cell level and the structure of activated cortical neural micro-networks need to be researched more thoroughly.

Until a few years ago mice were largely neglected among visual neurophysiologists. The species was considered unfit for this purpose because, as opposed to predator mammals (like the cat) and monkeys, both widely used for vision research, mice as nocturnal animals do not rely primarily on their vision, their eyes are positioned laterally instead of in the front, and their visual cortex has been found to have important differences as compared to higher mammals. The most notable cortical differences are generally large receptive fields of neurons, many of the cortical cells respond to stimuli placed anywhere in the larger part of a hemifield and the apparent lack of columnar organization. Despite all, mouse visual cortex can be used for observing the functional properties of neural microcircuits and fundamental functions of the visual cortex. Experiments on the retinotopic mapping of the mouse visual cortex have shown a very small variation in the topical organization, this is a great advantage when processing experimental data. The growing interest in mouse visual cortex was to a large extent stimulated by the advent of transgenic mouse technology. In our opinion the results of the experiments conducted on mice will help to further our understanding of the human vision.

Ca^{2+}-sensitive dyes are capable of displaying neuron action potentials and voltage sensitive dyes transform subthreshold changes to the membrane potential with submillisecond accuracy. These studies enable us to simultaneously visualize the behavior of a large number of neurons, or the signal propagation in neural microcircuits in cortical slices.

II. Methods

The around 8 weeks old adult mice are anesthetized with intraperitoneal urethane. Supplemental injections were used to maintain a stable anesthesia as assessed by heart rate and animal behavior. Airways are kept free by the intubation of the trachea. Animals are positioned in a stereotaxic frame or by using a nose clamp. Craniotomy is performed over the right visual cortex and the dura is removed. If necessary for peripherial stimulation, the contralateral orbita is also opened up, the optical nerve exposed.

The visual cortex is stained for ~45 minutes with RH-795 voltage sensitive dye. Until now we have only been using voltage sensitive dyes, but we plan to do Ca^{2+} sensitive imaging too in the future. A Micam Ultima camera was used for image recording.

Stimulation of the cortex is done either by optical stimulation or by applying electrical impulses to the retina, optic nerve or directly the visual cortex.

III. Progress

In the process of setting up the new experimental environment we try out several surgical techniques, experiment with different kinds of stimulation, altering the camera focus length, etc. Currently our biggest question in surgery is how to remove the dura in a safe way without hurting the cortex.