The Second Workshop and Lecture Series on: "Cognitive neuroscience of auditory and cross-modal perception"

20 – 24 APRIL 2015, KOŠICE, SLOVAKIA

# Program and abstracts book

# Objectives

- This workshop and lecture series will include introductory lectures and advanced research talks on a range of topics related to the ***neural processes of auditory, visual and cross-modal perception***.
- The talks will illustrate the ***multidisciplinary character of cognitive neuroscience research***, covering behavioral, neuroimaging, and modeling approaches, as well as applications of the research in auditory prosthetic devices.
- The workshop is aimed at ***early-stage and advanced students and young researchers***, and it will provide ample opportunities for direct interactions between the lecturers and the attendees.

# Themes

Spatial hearing, vision and crossmodal perception, neural modeling, methods in cognitive neuroscience: behavioral experiments, EEG and fMRI imaging, modeling, applications: cochlear implants, hearing aids.

# Format

Lectures 20 – 22 April, Consultations 23 – 24 April

# Venue

Historicka aula, P. J. Safarik University, Srobarova 2, 040 11 Kosice, Slovakia

# Organizers

Norbert Kopco, PhD. (norbert.kopco@upjs.sk)
Frederick Gallun, PhD. (Frederick.Gallun@va.gov)

# Organizing team and contact

Beata Tomoriova, Lubos Hladek, Perception and Cognition Lab, kogneuro@gmail.com

# Plenary lecture and panel discussion

**Virginia Best, Frederick Gallun, and Norbert Kopco**
Audibility and spatial hearing
Wednesday, 22 April 17:30-19:00 (Uhorsky dvor restaurant, directions)

## Program

| MONDAY 20.04.2015 | EXPERT LECTURES, CONTRIBUTED TALKS |
|---|---|
| Time / Location | Historická Aula UPJŠ (room #100) |
| 8:30 - 9:15 | Frederick Gallun (US Dept. of Veterans Affairs and Oregon Health & Science University)<br>**Learning From Nature's Experiments: What Clinical Research Can Mean for Sensory Scientists** |
| 9:25 - 10:10 | Arash Yazdanbakhsh (Boston University)<br>**Pursuit eye movements and perceived object velocity, potential clinical applications.** |
| 10:20 - 11:05 | Simon Carlile (University of Sydney)<br>**Active listening: Speech intelligibility in cocktail party listening.** |
| 11:15 - 12:00 | Aaron Seitz (University of California, Riverside)<br>**Perceptual Learning; specificity, transfer and how learning is a distributed process.** |
| | Lunch |
| 13:30 - 14:15 | Virginia Best (Boston University)<br>**Spatial Hearing: Effect of hearing loss and hearing aids.** |
| 14:25 - 15:10 | Pierre Divenyi (Stanford University)<br>**Toward an evolutionary theory of speech: how and why did it develop the way it did.** |
| 15:20 - 16:05 | Petr Maršálek (Charles University of Prague)<br>**Coincidence detection in the MSO – computational approaches** |
| 16:05 - 17:00 | Contributed talks |

| TUESDAY 21.04.2015 | EXPERT LECTURES, CONTRIBUTED POSTERS |
|---|---|
| Time / Location | Historická Aula UPJŠ (room #100) |
| 8:30 - 9:15 | Christopher Stecker (Vanderbilt University)<br>**RESTART theory: discrete sampling of binaural information during envelope fluctuations is a fundamental constraint on binaural processing.** |
| 9:25 - 10:10 | Bernhard Laback (Austrian Academy of Sciences)<br>**Sound Localization Cues and Perceptual Grouping in Electric Hearing** |
| 10:20 - 11:05 | Aaron Seitz (University of California, Riverside)<br>**Brain Training; How to train cognition to yield transfer to real world contexts** |
| 11:15 - 12:00 | Petr Maršálek (Charles University in Prague)<br>**On the single neuron computation** |
| | Lunch |
| 13:30 - 14:15 | Frederick Gallun (US Dept. of Veterans Affairs and Oregon Health & Science University)<br>**Auditory Processing After mild Traumatic Brain Injury: New Findings and Next Steps** |
| 14:25 - 15:10 | Simon Carlile (University of Sydney)<br>**Hearing motion in motion** |

| | |
|---|---|
| 15:20 - 16:05 | Istvan Winkler (Hungarian Academy of Science)<br>**Auditory processing capabilities supporting communication in preverbal infants** |
| 16:05 - 17:00 | Contributed posters |

| **WEDNESDAY 22.04.2015** | **EXPERT LECTURES** |
|---|---|
| Time / Location | Historická Aula UPJŠ (room #100) |
| 8:30 - 9:15 | Arash Yazdanbakhsh (Boston University)<br>**Visuospatial memory and where eyes look when the percept changes.** |
| 9:25 - 10:10 | Volker Hohmann (University of Oldenburg)<br>**Modeling Auditory Scene Analysis by multidimensional statistical filtering** |
| 10:20 - 11:05 | Istvan Winkler (Hungarian Academy of Science)<br>**Modeling auditory stream segregation by predictive processes** |
| 11:15 - 12:00 | Pierre Divenyi (Stanford University)<br>**What is the cost of simultaneously listening to the "what" and the "when" in speech?** |
| | Lunch |
| 13:30 - 14:15 | Christopher Stecker (Vanderbilt University)<br>**Neuroimaging of task-dependent spatial processing in human auditory cortex.** |
| 14:25 - 15:10 | Bernhard Laback (Austrian Academy of Sciences)<br>**Temporal effects in the perception of interaural level differences: Data and model predictions** |
| 15:20 - 16:05 | Volker Hohmann (University of Oldenburg)<br>**Modeling Cocktail Party Processing in a Multitalker Mixture using Harmonicity and Binaural Features** |
| | PLENARY LECTURE |
| Time / Location | Penzión Uhorský Dvor, Bočná 10 |
| 17:30 - 19:00 | Virginia Best (Boston University), Frederick Gallun (US Dept. of Veterans Affairs and Oregon Health & Science University), and Norbert Kopčo (P. J. Šafárik University)<br>**Audibility and spatial hearing** |

| **THURSDAY 23.04.2015** | **CONSULTATIONS & WORK ON ASSIGNMENTS** | |
|---|---|---|
| Time / Location | Dekanátna zasadačka PF UPJŠ (room #332) | Historická Aula UPJŠ (room #100) |
| 8:30 - 9:30 | Virginia Best<br>**MATLAB assignment: simulating the effect of hearing loss on spatial cues** | **(independently organized) Symposium on university spin-offs and start-up companies** |
| 9:30 - 10:30 | Pierre Divenyi<br>**Speech synthesis toolbox – Distinctive Region Models (DRM)** | |
| 10:30 - 11:30 | Petr Maršálek<br>**MATLAB assignment – numerical solving of ordinary differential equations,**<br>**with focus on neuronal simulation** | |

| FRIDAY 24.04.2015 | CONSULTATIONS & WORK ON ASSIGNMENTS | |
|---|---|---|
| Time / Location | Dekanátna zasadačka PF UPJŠ (room #332) | Veľká rektorátna zasadačka PF UPJŠ, 2nd floor (room #116) |
| 8:30 - 9:30 | Frederick Gallun<br>**Establishing normative ranges of performance using non-linear functions** | Volker Hohmann<br>**Implementation of a statistical estimator (particle filter) that tracks a (simulated) pitch track partially masked by noise.** |
| 9:30 - 10:30 | Bernhard Laback<br>**Acoustic simulation of cochlear implant perception with low-frequency residual hearing** | Aaron Seitz, Simon Carlile, Volker Hohmann<br>**Master Class:**<br>**Developing computer games for brain training on Symposium on university spin-offs and start-up companies** |
| 10:30 - 11:30 | Christopher Stecker<br>**1) Psychophysical exploration of binaural cues synchronized to envelope fluctuations: testing the RESTART theory with synthetic and naturalistic sounds. (hackathon type assignment)**<br>**2) Analysis of an fMRI data set combining task and binaural manipulations in a factorial manner.** | Arash Yazdanbakhsh<br>**A simulation assignment to replicate the gain of eye pursuit in following a target** |

| CONTRIBUTED TALKS (MONDAY) | |
|---|---|
| 16:05-17:00 | Robert Baumgartner, Piotr Majdak, and Bernhard Laback<br>**How spectral information triggers sound localization in sagittal planes** |
| | Andrew Abel and Amir Hussain<br>**Cognitively Inspired Speech Processing For Multimodal Hearing Technology** |
| | Jana Eštočinová, Jyrki Ahveninen, Samantha Huang, Stephanie Rossi, and Norbert Kopčo **Auditory Distance Perception and DRR-ILD Cues Weighting** |

| CONTRIBUTED POSTERS (TUESDAY) | |
|---|---|
| 16:05-17:00 | Aleksandras Voicikas, Ieva Niciute, Osvaldas Ruksenas, Inga Griskova-Bulanova<br>**Chirp stimuli for entrainment: chirp up, chirp down and task effects** |
| | Olena Markaryan<br>**Suggestion of rehabilitative treatment for patients subjected to sight restorative surgery** |
| | Simon Júlia, Csifcsák Gábor<br>**Early electophysiological correlates of susceptibility to the double-flash illusion** |

| | |
|---|---|
| | Gábor Csifcsák, Viktória Balla, Szilvia Szalóki, Tünde Kilencz, Vera Dalos<br>**Prediction processes in the visual modality – an EEG study** |
| | Barbora Cimrová, Zdenko Kohút<br>**Cross-modal interaction in spatial attention** |
| | Peter Tóth, Norbert Kopčo<br>**Speech Localization in a Multitalker Reverberant Environment** |
| | Gabriela Andrejková, Virginia Best, Barbara G. Shinn-Cunningham, and Norbert Kopčo **Streaming and sound localization with a preceding distractor** |
| | Beáta Tomoriová, Ľuboš Marcinek, Ľuboš Hládek, Norbert Kopčo<br>**Contextual plasticity in sound localization: characterization of spatial properties and neural locus** |
| | Peter Lokša, Norbert Kopčo<br>**Visual Adaptation And Spatial Auditory Processing** |
| | Norbert Kopčo, Eleni Vlahou, Kanako Ueno, Barbara Shinn-Cunningham<br>**Exposure to Consistent Room Reverberation Facilitates Consonant Perception** |
| | Hládek Ľuboš, Aaron Seitz, Norbert Kopčo<br>**Learning of auditory distance with intensity and reverberation cues** |

## Contributed presentations

Participant's presentations will have a dedicated slot in program on Monday and Tuesday from 16:05 to 17:00. Oral presentations will be on Monday and each presenter will have a 12+3 minutes dedicated for a presentation and a short discussion. If you were selected for a talk, please bring your powerpoint / pdf presentation (compatible with Windows 8.1 and MS Office 2013) on a usb stick on Monday at 12:00 (after morning lectures) to the organizers (Lubos Hladek, Beata Tomoriova). Poster sessions will take place on Tuesday at the foyer of the lecture hall. Poster boards (114 cm x 142 cm; width x height) with pins will be provided. Posters should be installed at 12:00 (during lunch break). The posters should be removed on the next day.

## Social Events

- Mon 18:00 dinner at Mediterran restaurant, then local beer at Madrid (for lecturers, participants welcome to join)
- Tue 17:15 – 18:45 guided tour of Kosice (if interested, sign up with Lubos Hladek at lubos (dot) hladek (at) upjs (dot) sk ). We are meeting outside of Srobarova 2, rector's office.
- Wed 19:00 Reception at Uhorsky dvor (after the plenary lecture, directions)

# Abstracts

## Monday, 20 April 2015

### Learning From Nature's Experiments: What Clinical Research Can Mean for Sensory Scientists

Frederick (Erick) Gallun
US Dept. of Veterans Affairs and Oregon Health & Science University

This presentation will focus on exploring the many ways that knowledge gained from studies of patient populations can not only improve the diagnosis and rehabilitation of sensory system declines but can also lead to insight into the function of the system itself. Data collected from "nature's experiments" are often harder to analyze than when the groups and interventions are assigned randomly, but the payoff is that it is possible to learn things that would be impossible to know if we were only willing to study in phenomena that can be carefully controlled. Furthermore, there are often insights to be gained from observational or correlational data that can then be examined using a more systematic approach. Examples will be drawn primarily from studies of aging and hearing loss and methods of analyzing complex data sets will be described. The assignment "Establishing normative ranges of performance using linear functions" will be introducted.

### Pursuit eye movements and perceived object velocity, potential clinical applications

Arash Yazdanbakhsh
Boston University

Our eyes are constantly moving. Visual information is very important for eye movement planning, while the ongoing eye movements affect the visual motion perception. Our brain is capable of integrating constant eye movements with the visual information from a changing external world, while maintaining a stable visual perception. In the talk, I explain a series of psychophysical and eye tracking experiments we conducted, in which healthy human observers pursue a moving target on a moving background. We systematically vary the background and target velocities, and analyze the initial and later stages of the pursuit eye movement, before and after the eyes are settled on the target. The percept of the target's velocity depends on the motion of the target itself as well as the motion of the background. For example, when the background moves in the opposite direction of the target, the target appears to move faster subjectively. However, little is known whether the pursuit eye movement, especially during the initialization of the pursuit, is affected by the background motion or not. If the pursuit is subject to the background motion, does the pursuit show a similar pattern with the percept, that is the eyes tend to move faster than the target when the background moves opposite to the target? What is the effect of the background motion on the pursuit eye movement, when only one eye sees the target and the other eye sees the background? In the talk, I report our findings that address these questions to understand the relation between the eye movement planning and motion perception.

### Active listening: Speech intelligibility in cocktail party listening.

Simon Carlile
Auditory Neuroscience Laboratory, School of Medical Science and Bosch Institute, University of Sydney, Australia 2006

This talk will briefly review the interaction between energetic and informational masking in the cocktail party problem. Informational masking has been attributed to a failure of attention in selecting or sustaining the focus of attention on the target talker. We will consider some work from our lab demonstrating that information masking need not be dependent on intelligibility in the maskers and may also be operating in bottom-up processing channels. Notwithstanding this, recent neuroimaging has demonstrated that the focus of attention plays a key role in the neural cortical representation and perception of successfully attended speech. In that context we will also review our more recent work

where we have shown that a deficit in temporal switching of attention also profoundly reduces success in cocktail party listening. Motivated by this finding we have developed a method to examine the costs of switching attention between talkers as in a dynamic conversational environment. Our recent data has demonstrated that, under conditions of high information masking, even in a group of high functioning adult listeners, individual differences in performance appear to be related to individual differences in working memory capacity and that the costs of switching is also related to concurrent language processing tasks.

## Perceptual Learning; specificity, transfer and how learning is a distributed process

Aaron Seitz
Department of Psychology, University of California, Riverside, USA

In this talk, I review recent research addressing mechanisms by which the visual system acquires knowledge about the world. In particular, I suggest that learning on any task involves a broad network of brain regions undergoing changes in representations, readout weights, decision rules, feedback processes, etc. However, importantly, that the distribution of learning across the neural system depends upon the fine details of the training procedure. I conclude with the suggestion that to advance our understanding of perceptual learning, that the field must move towards understanding individual, and procedurally induced, differences in learning and how multiple mechanisms may together underlie behavioral learning effects.

## Spatial hearing: Effect of hearing loss and hearing aids

Virginia Best
Boston University

Spatial hearing is critical for locating sounds and for optimal detection and recognition of sounds in complex environments. Hearing loss can disrupt the acoustic cues listeners rely on for spatial hearing, leading to disabilities related to sound localisation and exacerbating difficulties with speech understanding. Hearing aids and their features can in some instances restore spatial information, but in other instances can cause further disruption. This talk will review the acoustic cues that allow listeners with normal hearing to locate sounds. It will also discuss how sound localisation is affected by hearing loss and by hearing aids, using examples from the scientific literature. The talk will also review how spatial cues can enhance speech understanding in noisy environments, and discuss the effect hearing loss and hearing aids on this process. Finally, the talk will consider the implications of bilateral coordination between hearing aids.

## Toward an evolutionary theory of speech: how and why did it develop the way it did

Pierre Divenyi,
Center for Computer Research for Music and Acoustics, Stanford University, Stanford CA 94305, U.S.A.

Humans are endowed with a tube of ~18 cm length, the vocal tract (VT), that they have discovered (probably about 100-200 thousand years ago) that they can use for communication. With the vocal folds vibrating in their larynx and their mouth open, they could utter sounds – vowels – and by making them sufficiently different from one another to be audible by their fellow humans even in the omnipresent noise of their environment, they made up a code of communication. The criterion for making up the different sounds was Darwinian: reach the maximum contrast between sounds at the cost of minimum effort. The contrasts were achieved by deforming the VT at specific points and to specific degrees, directed by the above criterion applied to the physical constraints imposed by tube acoustics. Because constriction/expansion of the tube at those points could not be arbitrary, they needed to be learned, acquired and accepted within a group, and eventually transmitted genetically, according to Lamarck. This talk will be a quick introduction into tube acoustics and into a model of speech production dynamics (the Distinctive Region Model) to show just how ultimately simple speech communication is and how simply it could have evolved into what we deal with, day by day over our whole lifespan.

## On the single neuron computation

Petr Marsalek
Charles University in Prague

It has been described that neurons are capable of performing several arithmetic and logical operations with the use of spike trains. Viewing neurons as logical gates does not bring that much insight since the operations of neural computation are stochastic. We review approaches we used in past for description of stochastic nature of neuronal firing.

## How spectral information triggers sound localization in sagittal planes

Robert Baumgartner, Piotr Majdak, and Bernhard Laback
Acoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14, A-1040 Vienna, Austria

Monaural spectral information is important for human sound-source localization in sagittal planes, including front-back discrimination and elevation perception. Directional spectral information results from the acoustic filtering of incoming sounds by the listener's morphology and can be described by listener-specific head-related transfer functions (HRTFs). We propose a probabilistic, functional model of sagittal-plane localization that compares an incoming sound with internal templates derived from the listener's HRTFs. The model approximates spectral auditory processing, accounts for acoustic and non-acoustic listener specificity, and directly predicts psychoacoustic measures of localization performance. The predictive power of the modeling approach was evaluated under various experimental conditions, namely, band limitation, spectral warping, non-individualized HRTFs, spectral resolution, spectral ripples, and high-frequency attenuation in speech. A vital model component, inspired by the functionality of the dorsal cochlear nucleus, seems to be the extraction of positive spectral gradients, as it explains the listeners robustness in localization performance to macroscopic variations of the source spectrum.

## Cognitively Inspired Speech Processing For Multimodal Hearing Technology

Dr Andrew Abel, Computing Science and Mathematics, University of Stirling, Stirling, FK9 4LA, Scotland, aka@cs.stir.ac.uk, http://www.cs.stir.ac.uk/~aka/
Prof. Amir Hussain, Computing Science and Mathematics, University of Stirling, Stirling, FK9 4LA, Scotland, ahu@cs.stir.ac.uk, http://www.cs.stir.ac.uk/~ahu/

In recent years, the established link between the various human communication production domains has become more widely utilised in the field of speech processing. Work by the authors and others has demonstrated that intelligently integrated audio and visual information can have a vital role to play in speech enhancement. In addition to this, it is also worth considering the possibility of environments where multimodal information may be sporadic and of varying quality and so one single speech filtering approach may produce inadequate results when applied. We present a preliminary fuzzy logic based multi-modal speech filtering system that considers audio noise level and visual signal quality in order to carry out more intelligent, automated, speech filtering, making use of audio only beamforming, automatic lip tracking, and visually derived speech filtering. We propose to further experiment with the connections between audio and visual aspects of speech.

## Auditory Distance Perception and DRR-ILD Cues Weighting

Jana Eštočinová [1], Jyrki Ahveninen [2], Samantha Huang [2], Stephanie Rossi [2], and Norbert Kopčo [1,2,3]
[1] Institute of Computer Science, P. J. Šafárik University, Košice, 04001, Slovakia

[2] Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Harvard Medical School/Massachusetts General Hospital, Charlestown MA 02129

[3] Center for Computational Neuroscience and Neural Technology, Boston University, Boston MA 02215

**Introduction**

The estimates of auditory distance are typically dominated by the overall received stimulus intensity. However, distance processing can also be guided by intensity-independent cues. Specifically, the *interaural level differences* (ILDs) provide distance information for lateral stimuli and, in reverberant space, the *direct-to-reverberant energy ratio* (DRR) cue provides distance information for sources from all directions. In the absence of the intensity cue, listeners use these cues to estimate nearby-source distance [Kopčo et al. (2012) PNAS, 109, 11019-11024]. In the current study, we examined how the ILD and DRR cues are combined and weighted to create an auditory distance percept, and how previous experience influences this weighting.

**Methods**

We performed a series of behavioral experiments in a virtual reverberant environment in which we simulated sound sources presented at a varying distance (15-100 cm) from directly in front or to the side of the listener. To explore the listeners' weighting of the cues, we manipulated the availability and congruency of the cues. Specifically, we compared performance with the ILD or DRR eliminated, presented congruently, or presented incongruently. Stimuli were either binaural, monaural, or diotic. We also examined the effect of the preceding listening experience on cue weighting.

**Results**

Incongruent DRR-ILD stimulation caused a weaker distance percept (inaccurate estimates of source distance) compared to the congruent presentation. Individual cue weighting critically depended on previous experience. Very low weight was put on DRR after the subject was exposed to stimuli with congruent ILD and intensity cues. On the contrary, after stimulation with DRR-based performance, DRR weighting increased dramatically. Finally, diotic DRR-based performance was found to be better than monaural DRR-based performance even though the directional percept was more consistent with a realistic listening situation in the latter condition.

**Conclusions**

The weighting of ILD and DRR cues in judging distance of nearby sources is strongly adaptive, depending the previous room exposure. This result is consistent with the hypothesis that the brain dynamically updates its model of the acoustic environment, preferring the most reliable cue combination in each room. Future studies will need to examine the properties of this process and the underlying neural mechanisms.

[Work supported by APVV-0452-12, by the TECHNICOM project (ITMS: 26220220182) supported by the European Research and Development Fund, and by the NIH awards R01HD040712, R21DC014134, and R01MH083744]

## Tuesday, 21 April 2015

### RESTART theory: discrete sampling of binaural information during envelope fluctuations is a fundamental constraint on binaural processing.

G. Christopher Stecker, PhD

Vanderbilt University School of Medicine, Nashville TN USA

Effective localization of real sound sources requires neural mechanisms to accurately extract and represent binaural cues, including interaural time and level differences (ITD and ILD) in the sound arriving at the ears. Many studies have explored the relative effectiveness of these cues, and how that effectiveness varies with the acoustical features of a sound such as spectral frequency and modulation characteristics.

In particular, several classic and recent studies have demonstrated greater sensitivity to ITD and ILD present at sound onsets and during other positive-going fluctuations of the sound envelope. The results of those studies have clear implications for how spatial cues are extracted from naturally fluctuating sounds such as human speech, and how that process is altered by echoes, reverberation, and competing sources in real auditory scenes. In fact, they dramatically change our view of how the brain tracks objects in a spatial scene: rather than continuous processing of spatial information, it appears that sound envelopes form the basis for discrete and temporally sparse sampling the locations of sound sources.

In this talk, I review (at least) three decades' worth of results to summarize and critique the evidence for envelope-triggered extraction of ITD and ILD across a wide range of spectral frequencies. In sum, we find strong support for that view, which I will present in a theoretical framework termed "RESTART theory." I will describe several variants of this idea that have been introduced in the classic literature, and also the potential physiological and psychophysical mechanisms that underlie it.

## Sound Localization Cues and Perceptual Grouping in Electric Hearing

Bernhard Laback
Austrian Academy of Sciences

Cochlear implants (CIs) are auditory prosthetic devices that have originally been designed to restore speech understanding in easy listening situations, i.e., in quiet. It is well known that understanding of speech in complex listening situations, e.g., involving interfering sound sources, critically relies on monaural as well as spatial auditory grouping mechanisms. Thus, the performance of CI listeners in such situations likely depends on their access to respective grouping cues. This talk gives an overview of basic studies on the potential of providing spatial (binaural and monaural spectral) localization and monaural grouping cues in electric stimulation with current and future CI systems. With respect to binaural cues, bilateral CI listeners have been shown to be sensitive to interaural level differences but show a strong dependence on stimulation parameters in perceiving interaural time differences. The sensitivity to spectral cues for vertical-plane localization appears to be susceptible to level variation, showing that coding in the temporal domain may be required. CI listeners show relatively low sensitivity to monaural grouping cues and strong dependence on stimulation parameters. Overall, studies show the need to improve the salience of both monaural grouping cues and spatial localization cues in electric stimulation in order to better handle difficult challenging listening situations.

## Brain Training; How to train cognition to yield transfer to real world contexts

Aaron Seitz
Department of Psychology, University of California, Riverside, USA

Imagine if you could see better, hear better, have improved memory, and even become more intelligent through simple training done on your own computer, smartphone, or tablet. Just as physical fitness underwent a revolution in the 20th century, brain fitness is being transformed through innovations in psychology, neuroscience and computer science. This talk discusses recent research that begins to unlock this potential with consideration of the strengths and limitations of extant research.

## Coincidence detection in the MSO - computational approaches

Petr Marsalek

Charles University in Prague

Neural circuit of the medial superior olive (MSO) in mammals calculates azimuth of sound source direction. Several mechansisms have been proposed to describe the time precision of the interaural time difference (ITD) in the range of tens of microseconds used in the azimuth computation. Coincidence detection has a central role in achieving such high time precision. Using a particular description of post-synaptic potential interactions in the MSO neuron, we describe analytically and we further numerically explore the parameter space of these postsynaptic potential interactions.

## Auditory Processing After mild Traumatic Brain Injury: New Findings and Next Steps

Frederick (Erick) Gallun

US Dept. of Veterans Affairs and Oregon Health & Science University)

While traumatic brain injury (TBI) has long been known among clinicians to have potential implications for the auditory system, only in the past few years have neuroscientists started to focus on this issue. One of the reasons for this is that the main focus of most clinical evaluation of the auditory system is on damage to the peripheral system. One of the potential effects of TBI, however, is damage to the central auditory system. Such damage would be evident as difficulty understanding speech in background noise, making temporal discriminations among those auditory patterns necessary for speech perception, and potential difficulty with sound localization and lateralization. Studies of current members and Veterans of the United States Armed Services who have experienced mild TBI are beginning to suggest that auditory dysfunction of this type may be widespread. A recently completed study of civilians with multiple mild TBIs using similar research methods has also found evidence of auditory dysfunction that would not be obvious using only standard audiological evaluation techniques. Implications of these results for future diagnostical and rehabilitative approaches will be discussed.

## Hearing motion in motion

Carlile, S, Leung J, Locke, S, and Burgess, M.

Auditory Neuroscience Laboratory, School of Medical Science and Bosch Institute, University of Sydney, Australia 2006

The acoustic cues to auditory space are referenced to the head which moves through the world, itself composed of moving sources so that our sensation convolves source with self- motion. With the head still, velocity discrimination, a perceptual process, is related to static acuity via the minimum audible movement angle (MAMA). Yet, while we can accurately localise static sounds, we are much less sensitive to velocity, resorting to distance and time cues where available. Interestingly, when velocity changes as a step function, discrimination thresholds and the amount of post-transition stimulus required for detection is greater than the corresponding MAMA. This suggests that when the head is stationary, the window of temporal integration may vary according to the sound's velocity characteristics. We also have evidence that auditory representational momentum scales with velocity, not duration or distance.

Facing and following a moving auditory source is an ecologically important behaviour. Tracking exhibits on-line velocity correction for slow to moderate velocities (< 80 °/s) but at higher velocities reflects a more predictive mechanism. Patients with schizophrenia are impaired in their ability to track a moving auditory target, when compared with controls, despite having normal velocity perception when the head is not moving. The presence of other efference copy dysfunctions in schizophrenia suggests a key role for motor efference copy in the disambiguation of self and target motion.

## Auditory processing capabilities supporting communication in preverbal infants

István Winkler

Research Centre for Natural Sciences, Hungarian Academy of Sciences

It is well-known that infants can enter dialogues using sounds well before they learn to speak and that communication with adults is essential for socio-cognitive development. The theory of *natural pedagogy* postulates the existence of a human-specific form of teaching in which adults convey to infants precise information about objects in the environment and solutions to various problems (ostensive communication). This type of information transfer contrasts the goal-oriented learning observed in animals (with the possible exception of dogs, but not non-human primates). Ostensive communication requires infants to carry dialogues with adults, detecting when being addressed, marking understanding, consent, etc. In the talk, I suggest that the advanced auditory processing capabilities observed in preverbal infants are precisely the ones needed to support dialoguing by allowing infants to a) segregate concurrently active sound sources, b) represent speaker identity, c) categorize messages (speech vs. non-speech), d) detect whether or not a message has been directed to them, e) determine whether some action or answer is required, and f) reply or interject with correct timing.

## Chirp stimuli for entrainment: chirp up, chirp down and task effects

Aleksandras Voicikas, Ieva Niciute, Osvaldas Ruksenas, Inga Griskova-Bulanova
Vilnius University, Department of Neurobiology and Biophysics.

Auditory steady-state response (ASSR) is an electrophysiological response recorded to periodically presented auditory stimuli. According to theory, the frequency of the ASSR is close to the frequency of stimulation being maximal at around 40Hz with standard AM or FM modulated tones and clicks. ASSRs reflect the ability of neural networks to synchronize and recently were proposed to serve as a biomarker of schizophrenia. However, the optimal type and condition of stimulation is not estimated. In this study we investigate 1-120 Hz and 120-1Hz chirp elicited ASSRs. Experiment consisted of 3 tasks performed by 22 male subjects: counting stimuli, reading and sitting with closed eyes. 62 channels EEG was recorded and wavelet-extracted intertrial phase locking factor (PLF) was computed as the main measure of entrainment. PLF was lowest during reading, as expected. Topographical properties of the responses were the same for both chirps and similar to those as elicited by clicks - maximum in Fz location.  Maximum entrainment for both chirp stimuli was reached at 46 - 48 Hz on a group level, in contrast to standardly elicited ASSR that are largest at 40Hz.

## Cross-modal interaction in spatial attention

Marián Špajdel (1, 2), Zdenko Kohút (3), Barbora Cimrová (1, 3), Stanislav Budáč (1), Igor Riečanský (1, 4)

(1) Laboratory of Cognitive Neuroscience, Institute of Normal and Pathological Physiology,
Slovak Academy of Sciences, Sienkiewiczova 1, 813 71 Bratislava, Slovakia

(2) Department of Psychology, Faculty of Philosophy and Arts, University of Trnava, Hornopotočná 23, 918 43 Trnava, Slovakia

(3) Centre for Cognitive Science, Department of Applied Informatics, Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Mlynská dolina, 842 48 Bratislava, Slovakia

(4) SCAN Unit, Institute of Clinical, Biological and Differential Psychology, Faculty of Psychology, University of Vienna, Liebiggasse 5, 1010 Vienna, Austria

mspajdel@gmail.com, zcohut@gmail.com, barbora.cimrova@savba.sk, igor.riecansky@savba.sk

Spatial attention is a form of attention that directs perceptual processes to a location in space. The interaction of simultaneous processing of auditory and visual cues in spatial attention is not well-known. In our study, we used reflexive saccadic movements to selectively focus visual spatial attention during the dichotic stimulation task in healthy individuals. Our results suggest bias in auditory attention in the congruent direction to the evoked shift in visual attention. Furthermore, if the auditory attention shift is congruent to the visual attention shift, the visual processing of saccades is delayed. This delay is more pronounced if the attention is shifted to the right. These differences do not take place if the presentation of both modality cues is asynchronous. The results are discussed with respect to hemispheric asymmetry.

This study was supported by VEGA Grant No. 1/0083/15

## Prediction processes in the visual modality – an EEG study

Gábor Csifcsák, Viktória Balla, Szilvia Szalóki, Tünde Kilencz, Vera Dalos

The ability to feel agency has been associated with internal forward modeling. This phenomenon is considered to rely primarily on sensory predictions, and have been increasingly investigated in the auditory modality. Only a few studies were using visual stimuli, none of them focused on stimuli relevant in social interactions. Our aim was to examine prediction-related modification of visual processing by analyzing ERPs elicited by images of hands. Twenty-three adults participated in our experiment. In the ACTIVE condition, predictable stimuli appeared after the participants pressed a button with either right or left hand while in the PASSIVE condition, they were only observing the images. In a predominantly right-handed group, an amplitude decrease over the occipital cortex was observed in the early, 80-120 ms post-stimulus time interval, (i.e. around the visual N1 component) in the ACTIVE vs. PASSIVE condition much stronger for movements made by the dominant hand. We found that dominant hand stimuli and stimuli shown in the visual field of the dominant side evoke greater neural activity. Considering that these effects appear very early, it can be assumed that the modulation of the activity of the striatal/extrastriatal visual cortex contributes to the feeling of agency connected to hand movements.

## Early electophysiological correlates of susceptibility to the double-flash illusion

Simon Júlia, Csifcsák Gábor
Institute of Psychology University of Szeged

A double-flash illusion can occur when a single flash of light is accompanied by two brief auditory stimuli leading to the perception of two flashes. Although susceptibility to this illusion shows great inter-individual variability, which might even be used to differentiate between certain special conditions, reliable experimental protocol for its measurement are still needed. To this end, we recorded electroencephalography (EEG) in 17 participants in order to identify early neural responses, which are less influenced by decisional strategies. Two types of illusory conditions with different inter-sound intervals were used: 67 ms (short condition, SC) and 133 ms (long condition, LC). The strength of the illusion was 75% and 43% in SC and LC, respectively. In line with previous results, moderate to strong correlations between behavioural data and EEG difference waveforms amplitudes were observed in the early 120-140 ms post-stimulus interval. We also found another EEG marker peaking at 30 ms, presumably reflecting stimulus expectancy. In conclusion, we can say, that our experimental protocol is a good candidate for further examination of the relevant factors influencing the susceptibility of this illusion.

## Suggestion of rehabilitative treatment for patients subjected to sight restorative surgery.

Olena Markaryan
Independent researcher

Studies of patients who underwent the visual restoring surgery do not give encouraging results. Particularly, Dormal et al. (2014) recently informed that although the post-surgical improvement of

patient's visual performance takes place, it is still not complete and interferes by auditory crossmodal responses. This case is not the sole and other researchers also observed just partial vision function recovery after visual system repairing (Fine et al., 2003; Ostrovsky et al., 2009). Those observations refer not only to early-onset blindness but also to late-onset one, when blindness is acquired far beyond the critical (sensitive) period of development (Sikl et al., 2013).

All those cases sound quite pessimistic, especially from a perspective of newly developed sight-restoring surgical techniques. However is it that much pessimistic in point of fact? Apparently, it is not if to come to understanding that visual function loss is not restricted solely to the local tissue damage (Bola et al., 2014). While observing alpha band oscillations in both visually impaired and sighted subjects, Bola et al. concluded that visual function loss is accompanied with the disturbance of brain networks synchronization (BNS). More importantly, BNS may be adjusted with noninvasive repetitive transorbital alternating current stimulation (rtACS). Treatment with rtACS leads to improvement of patients' visual function performance. The success of rtACS was ascertained, particularly, by clinical investigations (Fedorov et al., 2011), where patients with optic nerve damage exhibited significant visual field and acuity improvements (by 9,3% and 0.02 correspondingly) after treatment. The explanation of such phenomenon was proposed by Sabel et al. (2011) within the "residual vision activation theory". One of the theory aspects is that the cellular mechanisms of vision functional recovering are ultimately similar to ones involved in the cognitive processes.

Finally, taking into consideration that rtACS improves visual performance in patients with visual system damage; rtACS approach could be suggested for using in rehabilitation of patients subjected to surgical restoration of visual system.

References:

**Bola** M., Gall C., Moewes C., Fedorov A., Hinrichs H., Sabel B.A.(2014).Brain functional connectivity network breakdown and restoration in blindness. Neurology 6, 542–551.doi:10.1212/WNL.0000000000000672

**Dormal** G, Lepore F, Harissi-Dagher M, Albouy G, Bertone A, Rossion B, Collignon O (2014). Tracking the evolution of crossmodal plasticity and visual functions before and after sight-restoration. Journal of Neurophysiology, 113, 1727-1742. doi: 10.1152/jn.00420.2014.

**Fedorov** A, Jobke S, Bersnev V, Chibisova  A., Chibisova Y., Gall C., Sabel B. A. (2011). Restoration of vision after optic nerve lesions with noninvasive transorbital alternating current stimulation: a clinical observational study. Brain Stimul.4:189-201. DOI:10.1016/j.brs.2011.07.007

**Fine** I, Wade AR, Brewer AA, May MG, Goodman DF, Boynton GM, Wandell BA, MacLeod DIA (2003). Long-term deprivation affects visual perception and cortex. Nat Neurosci 6: 915–916. DOI:10.1038/nn1102

**Ostrovsky**, Y., Meyers, E., Ganesh, S., Mathur, U., and Sinha, P. (2009). Visual parsing after recovery from blindness. Psychol. Sci. 20, 1484–1491. doi: 10.1111/j.1467-9280.2009.02471.x

**Sabel** B.A, Henrich-Noack P., Fedorov A., Gall C. (2011). Vision restoration after brain and retina damage: The "Residual Vision Activation Theory". Prog Brain Res, 192, 199-262. DOI: 10.1016/B978-0-444-53355-5.00013-0

**Šikl** R, Šimeček M, Porubanová-Norquist M, Bezdíček O, Kremláček J, Stodůlka P, Fine I, Ostrovsky Y (2013). Vision after 53 years of blindness. i-Perception 4(8) 498–507; doi:10.1068/i0611

## Learning of auditory distance with intensity and reverberation cues

Hladek Lubos[1], Seitz Aaron[2], Kopco Norbert[1]

[1] Institute of Computer Science, P. J. Safarik University in Kosice, Slovakia

[2] Department of Psychology, University of California Riverside, USA

The overall intensity and reverberation are major cues for auditory distance perception in rooms. When exposed to a new room, the auditory system has to adapt to accurately interpret the incoming stimuli. How this adaptation affects distance perception and whether it is subject to room learning remains unclear. Here, a learning experiment was performed investigating whether removing the intensity cue would result in enhancement of the reverberation cue learning [also see Kopčo et al., (2011) "Learning of reverberation cues for auditory distance perception in rooms." J. Acoust. Soc. Am. **129** , 2487]. Seven training sessions were performed on consecutive days, each consisting of 12 runs in which the subject localized a broadband sound in distance. The presented intensity was either fixed (A condition), or roved on trial-by-trial basis (R condition). Testing was performed at the beginning of the first, fourth, or seventh session, with testing condition (R or A) changing after each run. Results suggest that, compared to the A-training, the R-training caused a larger improvement in the R-test condition (re. A-test condition). Thus, learning the room-specific reverberation distance cues can be enhanced by eliminating the overall intensity cue. However, this learning doesn't generalize to the stimuli for which the overall intensity cue is available. [Acknowledgement: APVV-0452-12, TECHNICOM ERDF, ITMS: 26220220182]

## Streaming and sound localization with a preceding distractor

Gabriela Andrejková[1], Virginia Best[3], Barbara G. Shinn-Cunningham[3], and Norbert Kopčo[1,2,3]

[1] Institute of Computer Science, P. J. Šafárik University, Košice, 04001, Slovakia

[2] Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Harvard Medical School/Massachusetts General Hospital, Charlestown MA 02129

[3] Center for Computational Neuroscience and Neural Technology, Boston University, Boston MA 02215

A previous study of sound localization with a preceding distractor showed that (1) the distractor affects response bias and response variance for distractor-target inter-stimulus-intervals of up to 400 ms, and that (2) localization responses are biased away from the distractor even on interleaved control trials in which the target is presented alone [Kopco et al., JASA, 121, 420-432, 2007]. Neural mechanisms operating on time scales of milliseconds to tens of seconds are likely to cause to these effects. The current study examined how perceptual organization affects target localization performance. Sound localization was examined for 2-ms click target stimuli. On 80% of trials the target was preceded by a distractor, designed either to be grouped with the target (distractor was an identical 2-ms click) or to be perceived in a separate stream (an isochronous train of 8 clicks whose inter-click-interval was different from the distractor-target inter-stimulus-interval). As hypothesized, the single-click distractor affected target localization more than the 8-click distractor. On the other hand, the biases in the control trials were greater for the 8-click distractor. [Acknowledgement: APVV-0452-12, TECHNICOM ERDF, ITMS: 26220220182]

## Exposure to Consistent Room Reverberation Facilitates Consonant Perception

Norbert Kopčo[1], Eleni Vlahou[2], Kanako Ueno[3] & Barbara Shinn-Cunningham[4]

1 Institute of Computer Science, P. J. Šafárik University

2 Department of Psychology, University of California, Riverside

3 School of Science and Technology, Meiji University

4 Center for Computational Neuroscience and Neural Technology, Boston University

**Background**

An important aspect of real-world speech communication is the ability to adapt to reverberant listening environments that distort the speech signal. A few past studies show that consistent exposure to a particular room facilitates speech understanding, at least for a limited set of speech sounds [A. J. Watkins, 2005, J. Acoust. Soc. Am. 118:249–262.] and for sentences with rich lexical information [N. Srinivasan and P. Zahorik, 2013, J. Acoust. Soc. Am. 133: EL33-9]. Here, we present the results of two experiments investigating the effects of room consistency on phoneme perception using a wide range of consonants. Stimuli were nonsense syllables, allowing us to factor out lexical influences on perceptual compensation for reverberation.

**Methods**

Stimuli were VC syllables consisting of 16 consonants that were preceded by the vowel /a/ spoken by three different talkers, presented over headphones. Using room-related transfer functions, we simulated two different reverberant environments or anechoic space. On each trial, listeners heard an initial "carrier" phrase consisting of 0, 2 or 4 VC syllables, followed by a single target VC syllable. Listeners had to identify the consonant in the final target syllable. In some trials, the target and carrier had the same reverberation (matching), while in others the carrier syllables were simulated with either a different reverberant room (non-matching) or in anechoic space (anechoic). In Exp. 1 the carrier length was randomly varied from trial to trial; in Exp. 2 it was fixed within each block. We hypothesized that phoneme identification would be best in the matching condition, and that the benefit would increase with the length of the carrier phrase.

**Results**

Qualitatively, there was no difference between randomized and blocked trials. Reverberation hampered phoneme perception, with some consonants being particularly affected. Consistent with our hypotheses, exposure to consistent reverberation improved target consonant identification accuracy compared to when the carrier was anechoic. Relative to non-matching reverberation, the benefit of matching reverberation was less consistent: the benefit was significant for both 2-and 4-VC carrier phrases in Exp. 2 (blocked trials), but only reached significance for the 2-VC carrier phrase in Exp. 1 (randomized trials).

**Conclusions**

Short-term exposure to a consistent acoustic environment mitigates the detrimental effects of reverberation on phoneme perception, facilitating speech understanding in adverse conditions.

**Research Funding.**

Work supported by CELEST, a National Science Foundation Science of Learning Center (NSF SMA-0835976), APVV-0452-12, and the TECHNICOM project (ITMS: 26220220182) of the European Research and Development Fund

## Contextual plasticity in sound localization: characterization of spatial properties and neural locus

Beáta Tomoriová[1], Ľuboš Marcinek[2], Ľuboš Hládek[1], Norbert Kopčo[1]

[1] Pavol Jozef Šafárik University in Košice, Jesenná 5, Košice, Slovakia.

[2] Technical University of Košice, Letná 9, Košice, Slovakia.

Sound localization can be affected by context in which the task is performed. A previous study [Kopco et al., JASA, 121, 420-432, 2007] observed biases in responses on target-only trials, when they were interleaved with trials in which the target was preceded by distractor from fixed location. Here we performed three experiments in which we examined how the effect depends on 1) the spatial configuration of contextual stimuli, 2) the examined locations of the target-only stimuli and on 3) availability of visual signals and on the response method used, in order to understand the nature of its underlying neural representation. The context biased responses away from the distractor. The effect generalized also to locations at the same side of the distractor as the contextual stimuli but not to locations on the other side. In case the contextual stimuli were restricted to only a subregion of the target-only test region, the context also reduced response variability. This suggests that the repeated presentation of the contextual stimuli might increase spatial perceptual sensitivity for targets

presented in the region covered by the context. Effect of the context on biases and variance in responses was roughly independent of the response method or of availability of the visual signals, suggesting that contextual plasticity is caused by changes in auditory spatial processing only.

## Visual Adaptation And Spatial Auditory Processing

Peter Lokša, Norbert Kopčo
Institute of Computer Science, P. J. Šafárik University in Košice, Slovakia

Sensory information from one modality (e.g. audition) can be affected by stimuli from other modalities (e.g. vision.), a phenomenon known as cross-modal interaction. The most famous cross-modal interactions include the ventriloquism effect and the ventriloquism aftereffect in which the perceived location of an auditory target is shifted when the stimulus is presented simultaneously with a spatially displaced visual stimulus. A previous study of the reference frame of the ventriloquism aftereffect showed that: 1) locally induced ventriloquism effect can be induced, corresponding to 80% of the AV displacement, that 2) ventriloquism aftereffect corresponding to 50% of the AV displacement is observed for auditory-only stimuli, and that 3) the reference frame of the aftereffect is a mixture of eye-centered and head-centered coordinate frames [Kopčo, Lin, Shinn-Cunningham, Groh (2009). Reference frame of the ventriloquism aftereffect. Journal of Neuroscience, 29(44):13809-13814]. Another recent study [Wozny, Shams (2011). Recalibration of auditory space following milliseconds of cross-modal discrepancy. Journal of Neuroscience, 31(12): 4607-4612] observed ultra-fast adaptation effect which has a very quick onset and which fades away after few seconds. Here, a dissertation project is presented that will use the data of Kopco et al. (2011) to analyze a new visually-induced auditory adaptation phenomenon, and the reference frame of the ultra-fast adaptation.
Work supported by: VEGA-1/0492/12, APVV-0452-12

## Speech Localization in a Multitalker Reverberant Environment

Peter Toth, Norbert Kopco
Charles University in Prague

One way to deal with various aspects of sound source localization in complex situations is to select only cues that represent actual source location. Selection mechanism can be based on measurements of interaural similarity such as Interaural Coherence (IC) or Interaural Vector Strength (IVS).

Here, we use cue-selecting models in complex situations from recent psychoacoustic study where the task was to localize speech target in the presence of four masker talkers in reverberant room [Kopco et al., JASA 127, 2010, 1450-7]. Distributions of temporal differences in lower frequency bands and level differences in higher frequency bands were analyzed separately.

We confirm that cue-selecting models show good performance in several situations involving multiple talkers or reverberations. However, it seems that for more complex situations models that use information from distorted cues are more suitable than cue-selecting models.

# Wednesday, 22 April 2015

## Visuospatial memory and where eyes look when the percept changes

Arash Yazdanbakhsh
Boston University

A bistable visual stimulus, such as the Necker Cube or Rubin's Vase, can be perceived in two different ways which compete against each other and alternate spontaneously. Percept switch rates have been recorded in past psychophysical experiments, but few experiments have measured percept switches while tracking eye movements in human participants. We used an eye tracking system to track eye gaze position during spontaneous percept switches of a bistable, structure-from-motion (SFM) cylinder that could be perceived as rotating clockwise or counterclockwise. Participants

reported the perceived direction of rotation of the SFM cylinder by key presses. Reliability of participant's reports was ensured by including unambiguous rotations that were generated by assigning depth using binocular disparity. Gaze positions were analyzed up to 2000 ms before and after key presses. Our results showed that eye gaze positions 1000 ms before and 1500 ms after percept reports clustered in separate neighborhoods depending on the percept reported, but no such clustering was found beyond that timeframe. Direction of eye movements before percept report also depended on which percept was being reported. These findings suggest that percept switches of ambiguous stimuli can be correlated with prior eye gaze positions and velocities, and the visual hemifield where the ambiguous stimulus is located. In the talk, I will mention the future directions such as defining a spatial and temporal clustering metric to quantitatively measure the correlation between percept and eye position before, during, and after switches and investigating whether one precedes the other or if there is a bidirectional relationship.

## Modeling Auditory Scene Analysis by multidimensional statistical filtering

Volker Hohmann
Medical Physics, University of Oldenburg, D-26111 Oldenburg, Germany

Auditory Scene Analysis' (ASA) denotes the ability of the human auditory system to decode information on sound sources from a superposition of sounds in an extremely robust way. ASA is closely related to the 'Cocktail-Party-Effect' (CPE), i.e., the ability of a listener to perceive speech in adverse conditions at low signal-to-noise ratios. This contribution discusses theoretical and empirical evidence suggesting that robustness of source decoding is partly achieved by exploiting redundancies that are present in the source signals. Redundancies reflect the restricted spectro-temporal dynamics of real source signals, e.g., of speech, and limit the number of possible states of a sound source. In order to exploit them, prior knowledge on the characteristics of a sound source needs to be represented in the decoder/classifier ('expectation-driven processing'). In a proof-of-concept approach, novel multidimensional statistical filtering algorithms such as 'particle filters' have been shown to successfully incorporate prior knowledge on the characteristics of speech and to estimate the dynamics of a speech source from a superposition of speech sounds (Nix and Hohmann, 2007).

## Modeling auditory stream segregation by predictive processes

István Winkler
Research Centre for Natural Sciences, Hungarian Academy of Sciences

In most everyday situations, multiple sound sources are concurrently active in the environment. Based on the mixture of sounds arriving at the ears, the auditory system must group together the sounds riginating from the same source while separating them from the emission patterns of other sound sources (auditory stream segregation). Because this inverse problem has no unique solution, the auditory system uses heuristic principles to achieve stable and veridical perception. The talk describes a cognitive-level computational model of auditory stream segregation that qualitatively mirrors the way the human auditory system solves this difficult perceptual problem. The model is based on cognitive principles revealed in experiments with human listeners, based both on subjective perceptual and electrophysiological data. The resulting perceptually inspired model is based on predictive processingnd competition between alternative perceptual hypotheses (termed auditory proto objects). The modeling as well as the experimental results suggest that predictive processing is a powerful principle in organizing the auditory input originating from multiple sources.

## What is the cost of simultaneously listening to the "what" and the "when" in speech?

Pierre Divenyi,
Center for Computer Research for Music and Acoustics, Stanford University, Stanford CA 94305, U.S.A.

Speech is an approximately 4-Hz amplitude-modulated flow of units – syllables – each of which carries information that uniquely qualifies the syllabic flow in time. That information derives from the talker moving his/her tongue/mouth in a way that frequency-modulates, i.e., dynamically changes the spectral shape of, the speech sounds. A logical question is: are these two kinds of modulation processed by the listener independently of each other or, if not, do they reinforce or interfere with each other, when the information is embedded in interfering sounds. This question was addressed in a set of experiments in which speech-spectrum nonspeech sounds were used first to tests listeners' ability to discriminate simple AM-bound or FM-bound signals presented in AM or FM interference. Results of these experiments were compared to those of an experiment in which the listener had to make simultaneous AM and FM discrimination. The outcome showed that the two processes interfered with each other to a moderate degree, suggesting that attention of the listener is divided between two different sensory processes.

## Neuroimaging of task-dependent spatial processing in human auditory cortex.

G. Christopher Stecker, PhD
Vanderbilt University School of Medicine, Nashville TN USA

The importance of spatial processing in the auditory cortex (AC) is highlighted by two key findings. First, binaural tuning has been found in a majority of AC neurons (Kitzes 2008). Second, sound localization performance is profoundly disrupted by AC lesions in both human and animal listeners. Together, these observations suggest that populations of AC neurons participate in the representation of auditory space and link such representations to behavior.

However, the specific nature of spatial processing in AC–and especially human AC–is not well understood. Current debates revolve around the nature of spatial coding by neural populations, the degree to which spatial processing varies across AC regions, the sensitivity to specific cues such as interaural time and level differences (ITD and ILD), and the potential impacts of task-related factors.

This presentation highlights several results of recent studies using functional MRI to investigate these questions in human listeners. In the first set of studies, we attempted to quantify changes in sound-evoked responses as a function of ITD and ILD in presented sounds. Clear tuning to ILD could be observed throughout large regions of human AC including Heschl's gyrus (HG) and posterior regions in planum temporale (PT) and superior temporal gyrus (STG). Consistent with animal data, more robust ILD tuning was observed in these regions than anterior to HG. ILD tuning functions typically favored contralateral ILD values but demonstrated clear non-monotonicity consistent with opponent-channel representations of auditory space. Across studies, tuning to ITD was much weaker, limited in both the degree of response modulation and the extent of AC in which it could be observed.

Other studies focused on the influence of task features on spatial processing, contrasting auditory spatial, auditory non-spatial, and non-auditory tasks involving sensory discrimination and sensory memory judgments. Across studies, sound-evoked responses were enhanced during auditory as compared to visual tasks, especially in regions of posterolateral STG implicated previously in studies of non-spatial auditory attention. However, we observed little to no evidence for feature-specific modulations such as sharper ITD or ILD tuning, or response enhancements during spatial vs. non-spatial tasks, in AC.

## Temporal Effects in the Perception of Interaural Level Differences: Data and Model Predictions

Bernhard Laback
Austrian Academy of Sciences

While it is often assumed that interaural level difference (ILD) processing is integrative, i.e., not dependent on temporal stimulus properties, there exist data showing a limitation in ILD perception for stimuli with high rates of modulation. It has been proposed that this limitation is due to so-called binaural adaptation, a limit in perceiving binaural cues after the stimulus onset at high-rate modulation rates. In this study the origin of this limitation is studied by measuring ILD thresholds as a function of modulation rate for high-frequency-filtered pulse trains and comparing them to diotic sequential level discrimination thresholds measured in the same listeners. Overall, the ILD data show a non-monotonic dependence on modulation rate, with best performance at 400 pulses/s. In comparison, the sequential level discrimination thresholds show a monotonic improvement from low to mid rates and are overall higher than the ILD thresholds. An ILD model based on an auditory periphery front-end is shown to predict the non-monotonic ILD data and several other ILD data from the literature. Overall, the results suggest that rate effects in ILD perception can be predicted based on known properties of the auditory periphery, while there appears to be a more central limitation in sequential level discrimination performance.

## Modeling Cocktail Party Processing in a Multitalker Mixture using Harmonicity and Binaural Features

Volker Hohmann

Medical Physics, University of Oldenburg, D-26111 Oldenburg, Germany

### Introduction

In a cocktail party situation with multiple talkers, human listeners are able to attend to one specific talker. This ability involves target talker identification, tracking of the target talker over time, and understanding his or her message in the presence of the distracting sounds. This study proposes an auditory model framework for performing these tasks and evaluates it using a call-sign-based multitalker listening task with spatially separated talkers [Brungart and Simpson, Perception & Psychophysics, 2007, 69 (1), 79-91].

### Methods

The listening task involved (i) recognition of a target talker via a call-sign ("Baron") and (ii) understanding of a target phrase uttered by the target talker that consisted of a color and number word. Our proposed model framework consists of two main steps: First, the location and identity of the target talker is identified using a template matching procedure of harmonicity features [Ewert et al., Proceedings of the International Conference on Acoustics AIA-DAGA 2013, pp. 271-274] and binaural features [Dietz et al., Speech Communication 23, 2011, 592-605]. The template is based on multiple realizations of the target word "Baron". Second, the algorithm estimates a template of the color and number word based on the previously estimated target talker location and identity. Based on this template, the most likely color and number word spoken by the identified target talker is estimated. This contribution focuses on the first step, i.e, estimating the identity and location of the target talker.

### Results

The template matching procedure is compared to an approach using the Ideal Binary Mask, and to the subject results from Brungart et al. (2007). Tests using a small set of sample runs with two spatially separated talkers show that the detection of target talker identity and the estimation of its location is possible with a high accuracy.

### Conclusions

The proposed auditory model is able to detect the identity and location of the target speech token from a multitalker signal. It thus appears to be a good basis for estimating the target color and number tokens. Furthermore, the results achieved here on a small vocabulary are promising towards integrating harmonicity and binaural features into a complete CASA model that is based on a large vocabulary.

### Audibility and spatial release from masking

Virginia Best, Frederick Gallun, Norbert Kopčo

Listeners with hearing loss show poor speech understanding in spatialized multitalker listening situations, leading to the common belief that spatial processing is disrupted by hearing loss. This talk reviews studies from different laboratories that explored the contribution of reduced target audibility to this deficit. In addition, a new measure of target audibility based on the ideal binary mask is introduced, which might be a useful way to estimate audibility-based limits in spatialized speech mixtures. The implications for hearing aids and real world listening will also be discussed.

## Assignments

On-line resources: http://pcllin.ics.upjs.sk/~noro2/

Ask or email for the password.

### A1: Psychophysical exploration of binaural cues synchronized to envelope fluctuations: testing the RESTART theory with synthetic and naturalistic sounds.

G. Christopher Stecker, PhD
Vanderbilt University School of Medicine, Nashville TN USA

[I would like for this to be a programming / hackathon type assignment, where we develop algorithms (in MATLAB) for applying binaural cues selectively as a function of the sound envelope, and then test those algorithms by psychophysical demonstration. Alternatively, a purely psychophysical assignment could involve listening to and collecting brief data on several of the examples from the talk.]

### Assignment 2 (tied to presentation 2, and similar to last year's assignment)

G. Christopher Stecker, PhD
Vanderbilt University School of Medicine, Nashville TN USA

Analysis of an fMRI data set combining task and binaural manipulations in a factorial manner. [Last year, this was a walk-through of the steps involved in extracting stimulus-dependent activations from an fMRI dataset, in MATLAB. This year, students could do the same, or we could spend more time inventing and implementing novel approaches to get at effects in the data, such as event-based averaging, multivoxel pattern analysis, task-dependent functional connectivity, etc.]

### Establishing normative ranges of performance using linear functions

Frederick Gallun
US Dept. of Veterans Affairs and Oregon Health & Science University

The data associated with this assignment are saved in the Matlab file "spatial_release_data.mat" and represent threshold in a spatial release from informational masking task as a function of listener pure-tone threshold. The variable "Thresh0" indicates target/masker ratio at threshold in the condition where the target speech is colocated with the masking speech, while "Thresh45" indicates T/M at threshold for spatially separated target and maskers. The other two variables are named "Age", which is in years, and "PTA_5124_Bilateral" which is the average of the audiometric thresholds at the left and right ears for the frequencies .5, 1, 2, and 4 kHz in dB HL (Hearing Level, which is defined with reference to the average threshold for young normally hearing listeners).

1) Calculate z-scores for each listener in the two conditions and plot the data in z-units as a function of age and again as a function of PTA. How many of the listeners are "abnormal" (z-scores greater than 2)?

Clearly, there is a substantial spread to the data. How then can a clinical decision be made as to whether or not a given listener is behaving abnormally? Transforming performance to Z-scores allows abnormal ranges to be identified, but if the entire set of data is used to create the Z transformation, this can result in such a large range of "normal" performance that identification of abnormal thresholds becomes very difficult. A more useful approach would be to identify abnormal performance for a given age or hearing loss.

2) Now create functions to calculate z-scores. First, use linear regression to predict thresholds in the two conditions as a function of age.
3) For each condition, calculate the residuals from the predicted values for each listener and then calculate the function that predicts how the square of the residual depends on age.
4) Combine these two predictive functions to calculate z-scores for each listener in each condition, given their age and threshold.
5) Repeat the process for PTA.
6) Now how many listeners are abnormal? Are the same listeners abnormal using the z-scores based on PTA as on Age?

Want to take this further? What if you wanted to use both age and PTA?

Or, to take it in another direction, how could this be extended to non-linear functions? The non-linear approach is described in the article below, which is saved on the server.

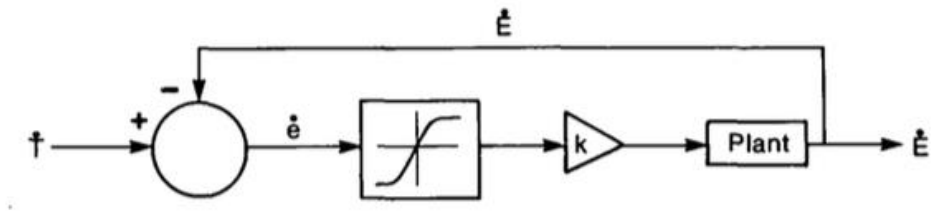
Tomlin, D., Dillon, H., and Kelly, A.S. (2014) "Allowing for Asymmetric Distributions When Comparing Auditory Processing Test Percentage Scores with Normative Data" JAAA 25:541-548 doi:10.3766/jaaa.25.6.4

## A simple model for replicating the pursuit eye movement and its gain

Arash Yazdanbakhsh
Boston University

In this assignment, you are required to simulate a simplified model of smooth pursuit system. You can use MATLAB or any other programming language with which you are comfortable.

The following figure, taken from White et al. 1983, offers a simple framework for the model elements.




$$G = \dot{E}/\dot{T} = \frac{k}{1 + k}$$

Proceeding from the left, eye velocity (Ė, the time derivative of E, eye position) is subtracted from the target velocity ($\dot{T}$, the time derivative of Target position, T) to yield retinal slip (target-fovea offset) velocity (ė, the time derivative of slip, e). This signal is passed through a saturating nonlinearity, depicted by the box containing a sigmoid curve.

You can adapt the following equation for your sigmoid and tweak the free parameters $\alpha$, $m$, and $q$ as needed.

$$S(x) = \frac{m}{(1 + qe^{-\alpha x})}$$

The sigmoid output is then multiplied by a gain factor, k, before being fed to the orbital plant (eye muscles and orbital mechanical properties) resulting in an eye velocity ($\dot{E}$), which is the output of the system.

The negative feedback loop, subtracts the eye velocity ($\dot{E}$) from the registered target velocity $\dot{T}$.

Eye pursuit gain is defined as the ratio of eye to target velocity, $\frac{\dot{E}}{\dot{T}}$ .

Prove that the negative feedback loop causes the overall system gain (G) to be k/(1 + k). Interpret the fraction in terms of k. Is this fraction invariant against $\alpha$, $m$, and $q$, or should $\alpha$, $m$, and $q$ be 1 to yield k/(1 + k)?

Explore the parameter space in your simulation while the simulated eye pursues a target with a fixed velocity. Are there parameter sets which yield:

a) an oscillating relative eye-target offset?
b) a plateauing to zero eye-target offset?
c) an increasing negative eye-target offset (the eye falls more and more behind the target)?
d) an increasing positive eye-target offset (the eye gets more and more ahead of the target)?

## Instructions for using the Distinctive Region Model (DRM) programs

Pierre Divenyi
Stanford University

Greetings!

This note assumes that you have heard the lecture on the evolution of speech and the DRM – it is a model of speech production based on the acoustic laws acting on tubes that are manifest every time you utter any speech sound. We also have to recognize that speech is a dynamic process characterized not by static speech sounds (phonemes) but by the movements between those sounds, the acoustic characteristics of which (mainly the peak frequencies of the first three formants, F1, F2, and F3) continuously change even if, for the listener's mind, those sounds appear as if they were simply beads on a chain. The trajectories of these dynamic, changing formant values are enclosed in the "vowel triangle": the triangle representing the frequency of F2 as a function of that of F1, with the three cardinal vowels, [a], [i], and [u] in the three corners. In natural speech and in any language no speech sounds occur with formant frequencies outside the three sides of that triangle.

- The programs are centered on vowel production and on how you perceive them. First, you will be asked to use the program to synthesize single vowels and sequences of 2 (V1-V2, VV_DRM.exe) or 3 vowels (V1-V2-V3 or V1-V2-V1, VVV_DRM.exe), listen to them, and write down your perceptual impressions about them. You should play with the synthesizers; change the duration of the two or the three components one-by-one and listen if their perceived identity has changed, especially that of the middle vowel in VVV_DRM.

- Next, you should do a synthesis of VCV (vowel-consonant-vowel, VCV_DRM.exe) syllables. Here, for sake of simplicity, the consonant C will always be a voiced stop: b, d, or g. You will create these by changing the region (=the place in the vocal tract) at which the vocal tract-tube is constricted, the point in time of its onset, and its duration. You should also write down what parameters you had to select, in order for you to obtain the stop-consonant you chose sound good. Then keep those parameters and change the vowel. Is the consonant still good? If not, what parameters have to be modified to make it good?
- Next, you should see what happens when you locate the F1 and F2 frequency values way outside the vowel triangle. You will notice that the synthesized vowels don't sound natural: they recall the voice of Disney characters. With this vowel-illusion program (V1V2_Illusion.exe) you will be able to change the formant frequencies of two vowels and create eerie-sounding V1-V2-V1 sequences and notice that the slope of the synthesized vowels' F1-F2 frequency trajectory can give the illusion of a natural V1-V2-V1 sequence the slope of which is similar to the one between vowels inside the triangle. Thus, you will understand that it is the trajectory and the speed of motion of the formant frequencies that determine the identity of the vowels you hear, rather than the formant frequency values at the point at which the direction of the movement is reversed.

## MATLAB assignment: Simulating the effect of hearing loss and hearing aids on spatial cues

Virginia Best
Boston University
(**ginbest@bu.edu**)

Locate a soundfile of broadband speech (I can provide one if needed) and download the different sets of HRTFs available here: https://starkeypro.com/resources/starkey-evidence/research-resources/hrtf-sets

1. Basic calculation of spatial cues

Filter the soundfile with the reference HRTFs for a set of locations.

a. Listen to the stimuli to confirm that the perceived locations are correct
b. Calculate the broadband ITD using cross-correlation for each location, and plot ITD as a function of location
c. Calculate the broadband ILD by comparing levels at the two ears, and plot ILD as a function of location
d. Plot the spectral cues from one ear as a function of location

2. Simulation of high-frequency hearing loss

The loss of audibility due to hearing loss can be simulated as a low-pass filter. Choose a low-pass cutoff (e.g. 1500 Hz) and filter the stimuli for each location.

a. Listen to these stimuli in comparison to the broadband stimuli, paying attention to both the quality of the speech and the spatial percept
b. Recalculate the ITD and ILD and plot as a function of location to compare to the broadband stimuli
c. Plot the spectral cues from one ear as a function of location and compare to the broadband stimuli

3. Simulation of different hearing-aid styles

HRTF sets are available for different hearing aid styles. Refilter the stimuli using the "BTE" and "IIC" styles.

a. Listen to these stimuli in comparison to the reference set, paying attention to the spatial percept, especially in vertical and front-back dimensions
b. Recalculate the ITD and ILD and plot as a function of location to compare to the reference set

c. Plot the spectral cues from one ear as a function of location and compare to the reference set

## MATLAB assignment: Numerical solving of ordinary differential equations, with focus on neuronal simulatio

Petr Marsalek, Petr.Marsalek@lf1.cuni.cz
Charles University in Prague

The purpose of this assignment is to learn and improve techniques of numerical solving differential equations, specifically ordinary differential equations, ODE, where the variable to be integrated is typically time, *t*.

In Matlab, ODEs are solved numerically as follows: Matlab contains several ODE solvers (integrators), ode23, ode45, ode113, ode23tb, and others.

The solvers are Matlab functions and they are part of respective function libraries (under MS Windows, or Unix):

C:\Program_Files\MATLAB\R2008a\toolbox\matlab\funfun\

/opt/matlab/toolbox/matlab/funfun

The equation to be solved must be written in a format of a function as well and as such it is passed to the solver. Other inputs have to be specified: Initial conditions, the range of values of the integrated variable, and initial variable difference and initial output tolerance to be used as initial precision of the solution. Some constants of the solvers are pre-set in order to give user typical initialization values.

Current version of Matlab contains only solvers with adaptive, i. e. variable step size. They use Runge-Kutta methods of different orders, Adams method, and others. The simplest Euler method is not available amongst them. Other utilities are available for plotting the solution, for obtaining phase plots or using the event location property. In Matlab, there is no solver with the fixed step, to my knowledge.

Several years ago I have written my own ODE solver with the fixed step size, using Runge-Kutta and Euler methods. Then I also wrote my own version of odeplot, since at that time, when I was writing these, it was not available.

When simulating models of neurons by dynamic ODE, it is advantageous to solve complicated and nonlinear equations, like the Hodgkin-Huxley equations, and others, with the use of fixed time step, simultaneously plotting the solutions and capture specific events, typically action potentials, or synaptic potentials and others. With the use of these techniques one is able to construct for example the f-I curve (firing frequency response to the level of DC current).

I will try to add more documentation into my source codes, since they do not contain many comments. Together with the students, we will reproduce several simulations of neurons.

I will bring with me two books, as an additional reading about ODEs. One is a voluminous survey of DEs and numerical techniques, (Rektorys, 1969) and the other, (Wilson, 1999), is an excellent introduction to the use ODE in neuronal simulation (even though the source codes at the attached 3.5“ floppy disc do not conform to Matlab standards, which we will use).

**References**

K. Rektorys: Survey of applicable mathematics, Iliffe books, London, 1969

H. R. Wilson: Spikes, decisions and actions: dynamical foundations of neuroscience, Oxford University Press, Oxford, 1999

Various User and Reference guides to MATLAB (TM) by Mathworks.

## Acoustic Simulation of Hearing with Cochlear Implants

*Bernhard Laback*
*Acoustics Research Institute, Austrian Academy of Sciences*

This assignment is intended to introduce you to acoustic simulations of electric stimulation of the auditory nerve with cochlear implants (CIs). To complete the assignment load the matlab script CIsim.m and follow the instructions below.

1) An example speech file is provided in the directory. Define the path on your computer to load the file into the vector wavedata.
2) Plot the waveform of the signal in units of seconds, using the "plot" command. The sampling interval is 1/srate, srate being the sampling rate.
3) The vector crnfreqAna contains the crossover frequencies for of a cascade of filters to separate the acoustic input signal into frequency channels. bv and av are the filter coefficients for these filters. Plot the amplitude response (and phase response) of all filters, using the function "freqz(bv,av,N,srate)". N is the number of points used for plotting (typical vale: 512). Restrict the plotted frequency range to 0.1 to 10 kHz by using the "axis([xmin xmax ymin ymax])" function. Use the command "hold on" to plotting the in frequency responses in one figure.
4) Plot the band-pass filtered signal of channel 6, from 0.66 to 0.72 ms. Plot the envelope of the same filtered signal into the same plot. This you can do using the absolute value of the Hilbert transform of the signal. Use a different color for the envelope signal, by adding, for example, the ', r' in the plot command.
5) Plot the band-pass filtered noise carrier for channel 6.
6) Plot the final synthesized signal for channel 6, i.e., the band-pass filtered noise carrier, modulated with the channel signal for channel 6.
7) Modify the script so that the crossover frequencies of the synthesis filterbank are lowered by a factor of 1.5. Listen to the resulting speech sound.
8) Create a binaural (two ear) signal by presenting the speech signal at both ears and introducing an interaural time difference of 600 μs (delay at the right ear).
9) Run the simulation with classical music (file "Classicmusic.wav"). How are different aspects of the music, like rhythm, melody, harmony, and timbre encoded? Think about why different aspects of music are affected differently by the CI processing.
10) Now add a stage simulating low-frequency residual hearing as occurring in many CI listeners. Do this by low-pass filtering the input signal. A common cutoff-frequency of low-frequency residual hearing is at about 500 Hz. Plot the frequency response of the low-pass filter, and finally add the residual hearing signal to the simulation of electric hearing.
11) Compare the simulations for speech and music when either using electric stimulation alone or when combining electric stimulation and residual low-frequency hearing. What do you observe?
12) Now add an interfering speech sound presented from the side of the listeners. Do this by adding an ITD to this interfering speech sound. Add the target and interfering speech sounds at a signal-to-noise ratio of 1. Run the simulation for electric stimulation alone and electric stimulation + low-frequency residual hearing. Attempt to follow the target speech. What effect do you observe when low-frequency hearing is added?

For more details of the method, see:

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," Science 270, 303–304.

Goupell, M. J., Laback, B., Majdak, P., and Baumgartner, W. (2008). "Effects of upper-frequency boundary and spectral warping on speech intelligibility in electrical stimulation," J Acoust Soc Am 123, 2295-2309.

For the effects of simulating low-frequency residual hearing, see e.g., Turner et al. (2004). J Acoust Soc Am. 115, 1729-1735