P. J. ŠAFÁRIK UNIVERSITY IN KOŠICE

FACULTY OF MEDICINE

Crossmodal automatic attention and spatial hearing

Diploma thesis

2017 Bernadeta HREBEŇÁROVÁ

P. J. ŠAFÁRIK UNIVERSITY IN KOŠICE

FACULTY OF MEDICINE

CROSSMODAL AUTOMATIC ATTENTION AND SPATIAL HEARING

DIPLOMA THESIS

Study programme:	General Medicine (Single degree study, doctor I.II. deg., full time form)
Study branch:	7.1.1. General Medicine
Supervising unit:	Department of Medical and Clinical Biophysics (ULBF)
Supervisor:	doc. Ing. Norbert Kopčo, PhD.
Consultant:	doc. RNDr. Ján Sabo, CSc., mim.prof.

Košice 2017 **Bernadeta HREBEŇÁROVÁ**

Acknowledgement

I would like to express my sincere thanks to my supervisor doc. Ing. Norbert Kopčo, PhD., the main Supervisor. Special mention should go to doc. RNDr. Ján Sabo, CSc., mim.prof., the Consultant. To all others who gave a hand, and to all participants of experiments I say thank you very much.

3856034733

P. J. Šafárik University in Košice
Faculty of Medicine

THESIS ASSIGNMENT

Name and Surname:	Bernadeta Hrebeňárová
Study programme:	General Medicine (Single degree study, doctor I.II. deg., full time form)
Field of Study:	7.1.1. General Medicine
Type of Thesis:	Diploma thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: Crossmodal automatic attention and spatial hearing

Title SK: Krosmodálna automatická pozornosť a priestorový sluch

References: Kopčo N, Best V, Carlile S (2010). Speech localization in a multitalker mixture. Journal of the Acoustical Society of America , 127, 1450-1457
Best, V, Ozmeral, EJ, Kopčo, N, and BG Shinn-Cunningham (2008) Object continuity enhances selective auditory attention, Proceedings of the National Academy of Sciences of the USA, 105, 13174-13178.
Steven J. Luck: An Introduction to the Event-Related Potential Technique, Second edition. Cambridge, Mass.: The MIT Press, 2014
Maddox RK, Pospisil DA, Stecker GC, Lee AKC (2014). Directing eye gaze enhances auditory spatial cue discrimination.Curr Biol. 24(7):748–52. doi: 10.1016/j.cub.2014.02.021

Aims: Study of brain mechanisms of spatial auditory processing and influence of unimodal crossmodal automatic attention on spatial hearing sensitivity.

Supervisor:	doc. Ing. Norbert Kopčo, PhD.
Consultant:	doc. RNDr. Ján Sabo, CSc., mim.prof.
Institutes :	ULBF - Department of Medical and Clinical Biophysics
Head of Institute:	doc. RNDr. Ján Sabo, CSc., mim.prof.

Approved: 10.04.2015 doc. RNDr. Ján Sabo, CSc., mim.prof.
Head of Department

Abstract in English

A previous study found an enhancement of auditory spatial discrimination ability when the listener´s gaze was directed towards the auditory stimulus (Maddox et al., 2014). In this thesis, we examined whether directing spatial auditory attention also affects this cross-modal enhancement when using realistic spatial simulation. Listeners made a judgment about the relative positions of two click-trains following a visual or auditory cue, while fixating on a neutral location. Results show that 1) subjects performed better when visual cue was used, and 2) auditory cue presented from incongruent location resulted in deteriorating performance. These results suggest a complex interaction between attentional and eye-gaze control mechanisms in auditory spatial representation.

Keywords: crossmodal automatic attention, attentional cueing, sound localization, crossmodal audio-visual interaction, attentional auditory ERP

Abstrakt v slovenskom jazyku

Nedávna štúdia odhalila zlepšenie schopnosti sluchovo priestorového rozlišovania, pokiaľ bol pohľad počúvajúceho nasmerovaný na sluchový podnet (Maddox a kol., 2014). V tejto diplomovej práci sme pozorovali, či nasmerovanie sluchovo priestorovej pozornosti tiež ovplyvňuje toto krosmodálne (medzi zmyslovými orgánmi) zlepšenie výkonu pri využití reálnej sluchovej simulácie. Počúvajúci posudzovali relatívne pozície zvukového „cieľa“ (dvojkliku) nasledujúceho po vizuálnej alebo sluchovej nápovede počas fixácie pohľadu na neutrálne miesto. Výsledky ukazujú, že 1) subjekty výskumu vykonali úlohu lepšie pri použitej vizuálnej nápovede a 2) sluchová nápoveda prezentovaná z nezhodnej pozície oproti cieľu spôsobila zhoršenie výkonu poslucháčov. Tieto výsledky predpokladajú komplexnú interakciu medzi kontrolnými mechanizmami pozornosti a očného pohľadu v sluchovo priestorovej reprezentácii.

Kľúčové slová: krosmodálna automatická pozornosť, navádzanie pozornosti, lokalizácia zvuku, krosmodálna audio-vizuálna interakcia, pozornostné sluchové ERP

Table of Content

List of Abbreviations

DRR	Direct to Reverberant Ratio
EEG	Electroencephalography
EOG	Electrooculography
ERP	Event-Related Potential
GERP	Grand average Event-Related Potentials
HRIR	Head Related Impulse Response
HRTF	Head Related Transfer Function
IID	Interaural Intensity Difference
ILD	Interaural Level Difference
IPD	Interaural Phase Difference
ITD	Interaural Time Difference
MSO	Medial Superior Olive
RT	Reaction Time
SOA	Stimulus Onset Asynchrony
TMS	Transcranial Magnetic Stimulation

Introduction

Objects and events in the real world are made up of multimodal sensory attributes. Our nervous systems process information from different sensory modalities independently, and this information from our senses is at some point combined into one perceptual experience. Perception is a multisensory process where sensory information is integrated both within and across different sensory modalities. Some studies have shown that auditory and visual stimuli can be integrated by bimodal cells, exhibiting spatially overlapping auditory and visual receptive fields. Such neurons have been found in the early sensory cortical areas such superior colliculus (e.g. Lakatos et al., 2007, 2008; Kayser et al., 2009) and recent study found that multi-sensory effects have been shown to occur in primary sensory areas as well (Lemus et al, 2010).

Multimodal activation has also been found in the human parietal cortex (Bremmer et al. 2001; Bushara et al. 1999, 2003; Cusack et al. 2000; Warren et al. 2002) and Intraparietal sulcus in the areas commonly referred to as LIP (lateral intraparietal sulcus bank) and MIP (medial intraparietal sulcus bank). Neuron cells in this area have been found to be sensitive to the locations of both visual and auditory stimuli (O'Dhaniel et al., 2005; Ben Hamed et al. 2001, 2002; Cohen et al. 2004; Gifford and Cohen 2004; Cohen and Andersen 2000).

Information from one sense has the potential to influence how we perceive information from another. For example irrelevant visual stimulus can affect the detection of an auditory stimulus (Lovelace et al., 2003) as well as the perceived loudness (Odgaard et al., 2004).

Attention facilitates selection of objects, events, or spatial regions in complex scenes. Very few studies focused on the effect of attention on sound localization. Even fewer studies looked at whether the effect is modality-dependent. Only a few previous studies asked whether directing automatic (exogenous, involuntary, stimulus-driven) or strategic (endogenous, voluntary, goal-driven) attention by an auditory cue can improve sound localization (Spence & Driver, 1994; Sach et al., 2000; Kopco et al., 2001). The result showed that cueing caused improvements in reaction times (Spence & Driver, 1994), but small (Sach et al., 2000) or no (Kopco et al., 2001) improvements in localization accuracy. Possible reasons were that tested SOAs were too short to orient attention and that auditory cue is not efficient because audition is not primarily a spatial modality.

A recent behavioral study demonstrated enhancement of auditory spatial cue discrimination ability when the listener’s gaze was directed towards the auditory stimulus (Maddox et al., 2014). However, such an effect has only been demonstrated for simplistic binaural cues (interaural time and level differences).

In this thesis, we are expanding the findings of the study by utilizing head related transfer functions (HRTFs) and by examining whether spatial auditory attention also affects this crossmodal enhancement.

Our experiment has two parts, behavioural and electrophysiological. In the behavioural one we try to evaluate localization ability of participants with cued attention and to study various phenomena observed in their accuracy results. In the electrophysiological one we search for electrophysiological correlates of these processes in recorded EEG data.

1 Psychoacoustics

The audible sound consists of rapid changes of air pressure which can be produced in different ways (for example…). Sound transmits through the air in normal temperatures at sea level at a velocity of approximately 335 m/s., for all amplitudes except for very great ones and for all waveforms. Especially interesting are periodic sounds, i.e. sounds having fixed waveforms that repeat at a fixed frequency. The frequency unit is Hertz (Hz), which means number of repetitions of a waveform per second (Warren, 1999). Period is the time that is required for one complete statement of a repeated waveform. Periodic sounds between about 20 and 16000 Hz are able to produce a sensation of a pitch and they are called tones. Waveforms can be described as time-domain representations of amplitude. By using Fourier procedure they can also be depicted "in terms of frequency-domain or spectral analysis, in which sound is described in terms of harmonic sequence of sinusoidal components having appropriate frequency, amplitude, and phase relations (Phase describes the portion of the period through which a sinusoidal waveform has advanced relative to an arbitrarily fixed reference time)" (Warren, 1999, p.2). Nonperiodic sounds spectrum is either continuous or band-shaped rather than line-shaped, as linear sounds do have. Frequency analysis is very important in hearing for both periodic and nonperiodic sounds, especially because the spectral analysis is performed by each ear before stimulation of the auditory nerve fibers occurs.

The range of audible amplitude is changing very largely. A sound that produces discomfort might be as much as 10^6 times the amplitude level at threshold. We can measure the sound level by its power or by amplitude as well as pressure at a particular time point. The term "sound intensity" can be, strictly speaking, defined as the sound power which arrives from a specific direction and passes through a unit area that is perpendicular to the direction (Warren, 1999).

In order to include a large range of values which are needed for describing levels of sound we normally encounter, a logarithmic scale has been developed and the units of the scale were named Bels. One Bel is a large unit, therefore it is conventional using 1/10 of this size, called the decibel (dB). In order to give a feeling to the sound pressure levels expressed in dB, the normal listener threshold for sinusoidal tones with certain frequencies was set. In between 1000 and 4000 it is about 0 dB (this is the standard reference level), for a background noise (the ambient level) in radio or TV it is about 30 dB, for a normal conversational speech the threshold is about 55 dB, and the level we can perceive inside a bus is about 90 dB. The sound level at some concerts can achieve values of 110 or 120 dB. This level approaches the pain threshold and can cause a permanent damage to the fine structures of inner ear leading to hearing impairment and it usually follows relatively brief exposures.

2 Sound localization

Sound localization is a listener's ability to identify the location or origin of a detected sound in direction and distance. With progress made in acoustic simulation techniques a sound localization technique appeared. This enables us to localize the sound sources by modeling a sound field which contains one or several sources. Using such technique, people can obtain the hearing sense from any place in the sound field.

Animals with the ability to localize sound have a clear evolutionary advantage. Especially the mammalian sound localization processes have intensively been studied. Their auditory system uses several cues for sound source localization, including time- and level- (intensity-) interaural differences (between both ears), spectral information, timing analysis, correlation analysis, and pattern matching. These cues are also used by other animals, but in slightly different usage, and there are also cues which are absent in the human auditory system, such as the effects of ear movements.

The localization can be described in terms of three-dimensional properties: the azimuth or horizontal angle, the elevation or vertical angle, and the distance (for static sounds) or velocity (for moving sounds). (Roads, 2007)

The azimuth of a sound is signaled by the difference in arrival times between the ears, by the relative amplitude of high-frequency sounds (the shadow effect), and by the asymmetrical spectral reflections from various parts of our bodies, including torso, shoulders, and pinnae. (Roads, 2007)

The distance cues are the sound level (loss of amplitude), spectrum changes (the loss of high frequencies), interaural level differences, and the ratio of the direct signal to the reverberated signal (DRR). Additional cues include vocal effort, vision and other cues related to DRR.

2.1 Cues of sound localization

2.1.1 Binaural cues of sound localization

When a sound stimulates equally or produces sensation identically at both ears, we call it ‘diotic’ stimulation. On the contrary, when the inputs arriving into the two ears are different, we call the stimulus ‘dichotic’.

Only dichotic hearing exists in natural environments, which enables binaural processing to occur. If we received the same information to both ears, in noisy situations we would not be able to follow conversations, locate sound sources or define our sound environment accurately.

Depending on where the source is located, our head acts as a barrier to change the timbre, intensity, and spectral qualities of the sound, helping the brain orient where the sound emanated from. These minute differences between the two ears are known as interaural cues (Thompson, 2005).

For sound localization, the most important parts of human head are ears. Human ears are located on the different sides of the head, therefore they have different coordinates in space. Since the distances between the sound source and individual ears are different, a time difference and an intensity difference between the sound signals of the two ears arise. These differences are called Interaural Time Difference (ITD) and Interaural Intensity Difference (IID) respectively.

2.1.1.1 Interaural Time Difference (ITD)

As noticeable in the Figure 1, the sound signal from the speaker has a shorter trajectory to reach the right ear comparing to the left ear. The auditory system evaluates ITD from:

(a) Phase delay at low frequencies. This is the time delay of the phase of the waveform and it is referred Interaural Phase Difference (IPD).

(b) Group delay at high frequencies. This is the time delay of the amplitude envelope of the waveform.

Many experiments demonstrate that ITD is frequency dependent. The function of ITD can be enumerated from the equation: $ITD = 300 * r * \sin\theta / c$, if $f \leq$4000 Hz; and $ITD = 200 * r * \sin\theta / c$, if $f \geq$4000 Hz (Zhou, 1996); where f is signal frequency, θ is the angular position of the sound source (0 degree is right ahead of the head, counter-clockwise is positive), r is the radius of the head, and c is acoustic velocity.

When the sound source is equally distant from both ears (at 0° or 180° azimuth), the ITD equals 0. On the contrary, the maximum ITD is generated when the sound source lies at ±90°. For a human head of an average size, it equals around 0.7 ms at this position (Campbell, 2006).

The Just Noticeable Difference (JND) of the ITD is smaller than 20 μs for pure tones between 500 Hz and 1 kHz (Campbell, 2006).

Above a certain frequency, when the sound wavelength becomes smaller than the diameter of the head, the ITD information gets ambiguous. There are numbers of locations forming an imaginatory cone which all elicit the same ITD. This is generally called the cone of confusion (Paukner, 2014).

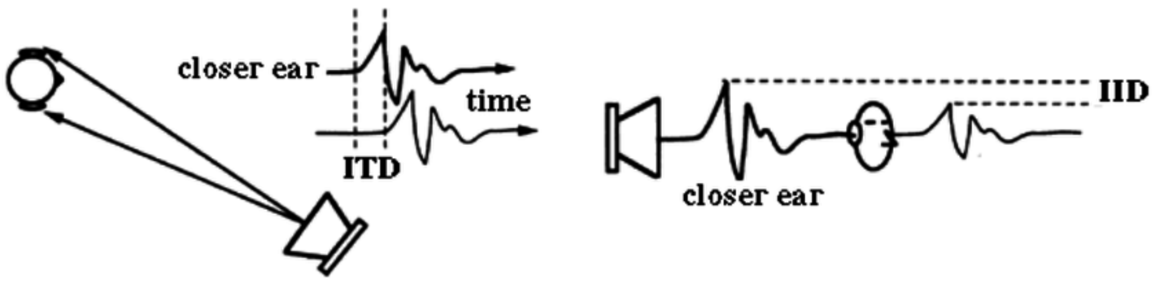

Figure 1. Interaural Time Difference (ITD) and Interaural Intensity Difference (IID) (Virtual Reality - History, Applications, Technology and Future - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/2617390_fig27_Figure-2531-Duplex-theory-spatial-cues-a-interaural-time-difference-b [accessed 7 Jun, 2017])

2.1.1.2 Interaural Intensity Difference (IID) or Interaural Level Difference (ILD)

As noticeable in Figure 1, sound from the off-centred sound source has a higher level at the closer (right) ear than at the further (left) ear. It is caused by the acoustic shadow that the head produces towards the left ear. ILD is highly frequency dependent and it increases with increasing frequency. Many theoretical researches demonstrate that ILD is frequency and angular position dependent, as given by following function: $ILD = 1 + (f/1000)^{0,8} * \sin\theta$ (Zhou, 1996). JND of ILD varies very much.

Taking these characteristics into account, a Duplex theory has been formed. It suggests that for frequencies lower than 1000 Hz, mainly ITDs are evaluated (phase delays), for frequencies higher than 1500 Hz mainly ILDs are evaluated. “For an object the size of the human head, a wave travelling at the speed of sound in air will not create a useful ILD below about 1–2 kHz. This cut-off frequency will be higher for heads of a smaller size. Thus, in humans, below this cut-off frequency binaural information for sound location is derived from ITD cues alone” (Campbell, 2006, p.24).

2.1.1.3 Evaluation of low frequencies

For frequencies lower than 800 Hz, the proportions of the head (ear distance average 21.5 cm, corresponding to the interaural time delay of 625 μs) are smaller than the half wavelength of the acoustic waves. Therefore the auditory system is able to determine phase delays between the both ears without distraction. Because the wavelength is very long in this frequency range, the waves diffract around the head and ILDs are therefore negligible, so an exact evaluation of the input direction on the basis of intensity differences alone is almost impossible. For frequencies below 80 Hz it becomes impossible to use either time differences or intensity differences to determine the sound source, because the phase difference between the ears is too small for evaluation of the direction (Blauert, 1997).

2.1.1.4 Evaluation of high frequencies

For frequencies higher than 1600 Hz the proportions of the head are bigger than the wavelength of the acoustic waves. At these frequencies, a definite determination of the input direction using interaural phase delay alone is not possible. The interaural intensity differences become greater, and the auditory system is able to evaluate them. Also interaural group delays can be evaluated, and they are more distinct at higher frequencies. This mechanism plays especially important role in reverberant environments.

As stated above, the duplex theory suggested that sounds of low frequency are localised on the basis of ITDs and sounds of higher frequency on the basis of ILDs. „However, the situation is now known to be more complex. For example, high frequency sounds can be localised using ITDs if the envelope of the stimulus has a amplitude modulation of sufficiently low frequency. Even more important is the direction-dependent spectral filtering of sound by the external ear" (Campbell, 2006, p.25).

After a sound onset a short time frame follows where the direct sound reaches the ears, but the reflected sound not yet. The auditory system evaluates the sound source direction based on this time frame, and it keeps this detected direction during the time when reflections and reverberation prevent a definite direction estimation (Wallach, 1949).

Binaural cues described above can be only used to evaluate lateral sound source positions, because two sound sources which are symmetrically placed in the front and in the back of the head generate equal ITDs and ILDs. This is called cone model effect. For differentiating whether the sound source lays ahead of the listener of behind him, additional cues must be evaluated.

Naturally hearing, humans are able to localize sound sources very accurately even by using only one ear, it means without ITDs or ILDs. Taking this fact into account, monaural cues provides a satisfactory explanation.

2.1.2 Monaural cues of sound localization

The sound waves coming to the ears are reflected from structures of the body such as the shoulders or pinnae, and these reflections interfere with the sound entering the ear canal. In each frequency zone, these interferences cause spectral modifications, either reinforcements (eliciting spectral peaks) or deteriorations (eliciting spectral gaps) which enable to localize the sound source, especially in the vertical plane.

The human outer ear is the structure of the utmost importance for vertical localization (including structures of the pinna and those of the external ear canal). Human pinna has got a very special shape, it is concave with folds and asymmetrical horizontally and also vertically. All these structures form direction-selective filters. Depending on the sound direction in the vertical plane, different filter resonances become active and they implant direction-specific patterns into the frequency-dependent responses of the ears. All this information is then evaluated by the auditory system, creating directional bands (i.e. frequency bands of different elevation perception, perceived when presenting a narrow-band noise at certain frequencies (Paukner, 2014)).

In other words, the direct waves and the reflected waves generate the acoustic source-related frequency spectrum on the eardrum. Subsequently, the auditory nerve localizes the sound source by this frequency spectrum. Other direction-selective reflections of the head, torso and shoulders together with those of the ears mold the Outer Ear Transfer Function.

Depending on the size and shape of the outer ear, the direction-specific patterns in the frequency responses of ears are very different across individuals. When a listener is trying to use the directional patterns of another person, for example when using headphones individualized for another listener (with differently shaped head and outer ears), it becomes very problematic to evaluate sound direction, especially in the median plane. Similarly, when listening to a dummy head recordings, front–back permutations or perception of inside-the-head-localized sound can appear.

Humans can binaurally localize both high and low frequencies. However, monaural localization is possible only for high frequencies, but not for low ones. This is probably because the pinna is too small to interact with acoustic waves of long wavelength, i.e. of low frequency (Butler, 1992). It seems that in the vertical plane, it is only possible to localize accurately auditory stimuli which are complex and of frequency above 7000 Hz and the pinna must be present (Roffler, 1968).

2.1.2.1 Head-related transfer function (HRTF)

The monaural cues reflect the interaction between the sound source and anatomical structures of a human. Before the sound reaches the ear canal it is modified and this modification encodes the source location. Impulse response captures the source location and the ear location. In this case, the impulse response is named the head-related-impulse-response (HRIR). By convolution of an arbitrary sound source with the HRIR the target sound is converted so that the listener perceives the sound source at the same location as it would have been played at the estimated sound source location. Therefore, HRIR can be used to create virtual sounds. The HRTF originates as the Fourier transform of HRIR.

Head-related transfer function is a response to a sound from a particular location, and it reflects characteristics of transmission of this sound from the location to the auditory system. When a sound strikes a listener, many head properties (the dimensions, shape and density of the head, ears with ear canal, dimensions and shapes of head cavities) are involved in transforming this sound and affecting the perception of the sound, resulting in boosting some frequencies and attenuating others. This happens because the incoming sound contains many various frequencies and therefore many copies of the same sound, each copy of a different frequency, going down the ear canal at different times, depending on their reflection and diffraction at the structures of the ear. Some copies match in their phases, what results in enhancing the signal, while others do not match and this leads to cancelling out each other. In general, HRTF boosts frequencies in range 2 – 5 kHz, with a primary resonance of + 17 dB at 2700 Hz. It is known that brain looks for certain frequency notches in the sound signal which correspond to particular known sound directions.

To synthetize a binaural sound coming from a particular location a pair of HRTFs (of the two ears) can be used. HRTF is a transfer function, i.e. it is a mathematical representation of the relation between the input and output of a linear time-invariant (LTI) system with zero initial conditions and zero-point equilibrium, in terms of spatial or temporal frequency. It describes how a sound from a specific location in space arrives at the ear.

Also HRTF can be described as modifications of the sound moving from the source direction in free air until reaching the eardrum. These modifications include the acoustic characteristics of the space where the sound is played, head´s, body´s, and outer ear´s shape and they play a great role in listener´s accuracy to tell the direction of the target sound.

HRTF describes filtering of the sound source ($x(t)$) before perceiving it. It is evaluated for right and left ear separately, $x_R(t)$ and $x_L(t)$, respectively.

The measured HRTF is not only analyzed but can also be used to reproduce a sound in the 3D virtual sound field, for example when using headphones. Through the headphones a mono-audio signal is presented. For a specific distance and direction, this signal is filtered by the measured HRTF (Paukner, 2014).

The transfer function $H(f)$ of any LTI system at frequency f is: $H(f)$ = Output(f) / Input(f). In other words, it is a ratio between the output signal spectrum and the input signal spectrum as a function of frequency.

2.1.2.1.1 How is HRTF measured?

HRTF can be obtained as the Fourier transform of HRIR. HRIR $h(t)$ can be measured from the eardrum for the sound impulse $\Delta(t)$ placed at the source.

HRTF is usually measured in an anechoic room in order to minimize the impact of early reflections and room reverberation on the measured response. The measurement procedure adds increments of 15° or 30° to the azimuth in horizontal plane and by interpolation it synthetizes HRTF for any horizontal position (azimuth).

2.1.3 Neurophysiology and physiological correlates of sound localization

In real world a sound source in space stimulates both ears. Side on which the source is heard depends on time and sound intensity differences at the two ears.

A large population of neurons at and beyond the Superior olivary complex is responsive to these time and intensity differences at the two ears. Experiments provided evidence that the cells are involved in sound localization. When a speaker is moved round the head and neurons are recorded electrophysiologically, the biggest responses are usually found on the contralateral side from the speaker side, the response area in some neurons covers the whole contralateral hemifield and in others it is limited to certain specific areas. These neurons with narrow response areas tend to respond to high frequencies and tend to be directionally selective aligned with axis of the pinna. If the pinna is manipulated the directional sensitivity of the cells can be altered. (Pickles, 1991) "In this type of cell, therefore, the fundamental directional selectivity is probably derived from the directionally-selective amplification produced by pinna, enhanced and preserved by the sensitivity to interaural intensity differences. On the other hand, the selectivity of low-frequency neurons may be a result of the lower directional selectivity of the pinna in the frequency range." (Middlebrooks and Pettigrew, 1981, p.118). Also timing cues are possibly involved.

Understanding of the basis of sound localization remains still unclear, since many cells tend to respond optimally to ITDs and ILDs which are bigger than could be produced by any real source of sound in space. For example, the optimal or characteristic interaural delays in 88% of MSO neurons of a kangaroo rat are bigger than the maximal delay calculated from separation of both ears (Crow at al., 1978). In chinchillas, the maximal interaural intensity difference for frequencies below 2 kHz is expected to be 4-5 dB but the cortical neurons of these frequencies are optimally sensitive to intensity differences bigger than 20 dB" (Benson and Teas, 1976). Such neurons may have a role, not in representing the direction of a sound source, but in discrimination between the directions of sound sources." (Pickles, 1991, p.285).

Many behavioral experiments were done to clarify the role of the brainstem in sound localization. For instance, experiment with cats showed that only cutting the crossing fibers in the trapezoid body would affect the sound localization skills (Moore at al., 1974). These are the fibers which transfer binaural interactions to the Superior olivary complex. Other transections were without effect. Other experiment showed that if there was a unilateral lesion at the level above the Superior olivary complex, deficit in sound localization occurred on the contralateral side (Jenkins and Masterton, 1982).

3 Attention

Our sensory system is exposed to plenty of impulses every moment. It is overwhelmed by scenes, sounds, smells,… But not all of them can be perceived equally. Some of them are important, others are not. Since it is impossible for our brain to process such amount of information, there certainly is a mechanism which select among them and this mechanism is an attention (Tomoriová, 2007).

Despite the fact that concept of attention is intuitively clear, there is not a uniform definition of term 'attention' since it involves many aspects. For example, Dana Murphy defines it as mental concentration on sensory or mental events. Another definition, according to David Sommers, describes attention as activation of mechanisms which provide continual cognitive activity focused on the object of attention (Kopčo, 2007).

From many studies about character of attention, some its features were clarified:

Attention is limited, which means it is impossible to focus it on all stimuli at once. It is selective, which means that among all the inputs it chooses only few that will be processed, by inputs meaning not only sensory stimuli but also information from memory and motor responses (Kopčo, 2007). We could liken attention to a podium reflector; we only see in detail things that are lit and others we ignore.

If we say we pay attention to something we usually mean that we consciously target it to that thing. Therefore it can be assumed that this type of attention is a targeted process, and accordingly it is called strategic or endogenous or goal-driven attention. On the other hand, sometimes a certain stimulus disturbs us (e.g. noisy sound, intensive light) and draws our attention towards itself without any conscious control. This another type of attention is called automatic or exogenous or stimulus-driven.

3.1 Attentional cueing

In order to localize sound and develop spatial orientation the attention has to be drawn to the specific location. Drawing of auditory attention is dependent on appearance of the stimulus sound.

3.1.1 Posner´s cueing task

Michael Posner developed a neuropsychological task which is nowadays used as an underlying test to assess attention. In this task, attention is drawn by presented stimulus and it is called cueing. The major phenomenon resulting from the task and forming the Posner paradigm is that subject have better performance in detecting objects in space if the locations have been cued before, using a salient stimulus as the cue.

In his experiment, the measured responses which reveal the effect of attention were reaction times (RT) to target stimuli. The subjects were seated in front of a computer screen with fixed gaze at the central point on the screen. For a short period, a visual cue (of an arrow shape) was presented, appearing at the left or the right side of the screen and pointing to one or another side. Subsequently, the target visual stimulus appeared at either left or right side. When subjects detected the target they were instructed to respond manually by pressing buttons on a keyboard depending on which side the target stimulus was presented. Multiple trials were recorded and the results then analyzed.

In order to analyze the attention based on type of visual input, two different cue positions were used. An endogenous cue was presented at the central point of the screen, evaluating the input in the central visual field. An exogenous cue, on the contrary, was presented at a lateral position, visible and evaluating the input in the peripheral visual field. This experimental design leveraged a ratio between valid and invalid trials equal 80/20. It means, 80% of trials employed a valid cue (matching the cue information - suggested direction – with the actual position of the target) and 20% invalid one (the cue information was incorrect). This gave the listeners the belief to rely on the information provided by the cue, which reinforced the tendency to direct attention to the cued side. Another cue type was so called neutral cue, which information value was ambiguous (double-sided arrow) and this condition was used to analyze whether the attention is directed by cues to a specific area and whether it is beneficial or hindering.

A way how to assess the type of attention involved in performing the task may be to monitor eye movements. For this, a video-based eye tracking system or an EOG recording can be used. Based on these methods one can differentiate whether an overt or covert attention was implied. Overt attention includes saccadic eye movements consciously directed towards a visual stimulus. Covert attention involves only mental focus without directing eye gaze to the visual stimulus. Since the subjects were not allowed to move their gaze to any direction, but keep it fixated in the center of the screen, the measured RTs correspond to employment of a covert orienting of attention (Posner et al., 1978).

Results of Posner´s paradigm show that when attending a location with preceding valid cue, even without looking at it directly, the time needed to determine the stimulus position (RT) is decreased. "Detection latencies are reduced when subjects receive a cue that indicates where in the visual field the signal will occur" (Posner et al., 1980). Moreover, the processing of the stimuli seems to be more intense (Prinzmetal, 2005) and the probability of detecting a near-threshold stimuli appearing in the periphery seems to increase (Bashinski, 1980). Furthermore, the strength of effects of cueing is directly proportional to percentage of valid cues (Chun, 2000).

Attention can be categorized according to its components: selection, following, and control.

The question when assessing selection is what attention chooses based on. Whether it focuses only on a certain area in space or on particular objects independently on their spatial distribution. Studies show that selection can be based on both, spatial location (location-based attention) or on the individual object (object-based attention) or on specific tokens of objects (object-token-based attention).

Following as the second component means keeping attention on a target object despite presence of attentional distractors.

The third component, control, is a process through which the nervous system is able to change settled attention from one object to another which becomes a new target of the interaction. Attention can be controlled by different sensory modalities, e.g. hearing, vision, touch,… Depending on how many modalities are involved, the control process of attention can be unimodal (auditory attention controlled by hearing, visual attention controlled by vision,…) or crossmodal (auditory attention controlled by vision, visual attention controlled by hearing,…) (Tomoriová, 2007).

The attention employed in visual and auditory complex scenes is an object-based attention. (Shinn-Cunningham, 2008) Therefore this type of attention will be described below.

3.2 *Object-based attention*

"Theories of visual attention explain many striking perceptual phenomena that arise when viewing complex scenes." (Shinn-Cunningham, 2008)

In contrast to visual objects, auditory objects form their identity as their content develops over time. But many auditory phenomena can be understood and explained by properly extended theories of visual attention. This similarity is supported by the fact that visual and auditory attention are controlled by the same neural processes. (Serences et al., 2004)

As an object is considered "a perceptual estimate of the content of a discrete physical source" (Shinn-Cunningham, 2008). Unlike visual objects which are quite precisely understood, auditory objects lack accurate definition. This occurs because an auditory stimulus (sound) in a mixture of other sounds cannot always be allocated between the objects perceived in the environment, it can contribute to either no object (Shinn-Cunningham et al., 2007) or multiple objects (Whalen and Liberman, 1996). Despite this, we intuitively understand the concept of an auditory object. Each of auditory objects can be seen as an estimation of the sound radiating from a certain sound source. In other words, "an auditory object is a perceptual entity that, correctly or not, is perceived as coming from one physical source" (Shinn-Cunningham, 2008). For example, in a party, we perceive a friend´s voice talking with us, glasses cling or a music playing in the background, each as a particular object.

3.2.1 Object formation

Formation of objects in visual scene occurs based on locally continuing geometric structure (contours, boundaries, edges) or similar pattern (color, texture) (Feldman, 2003). Forming of auditory objects depends on their contiguous spectro-temporal structure (continuity of frequency over time, harmonic structure, common onsets and offsets).

How a particular feature or cue influences the object formation depends on "the scale of the analysis" (Shinn-Cunningham, 2008)." For example, spatial auditory cues (e.g. ILD) have a relatively weak influence over local time scales (Darwin and Carlyon, 1995; Carlyon, 2004). On the other hand, "perceived location strongly influences how we link short-term auditory objects into a coherent stream" (Darwin and Hukin, 2000).

Even though the description above may suggest that objects are formed through a hierarchical processing, meaning initially grouped based on local structure, and only then organized across longer temporal or spatial scales, in fact the process is more complex. Top-down attention and higher-order features can modulate how objects form locally. “Rather than a hierarchical processing structure, objects are formed through heterarchical interactions across different scales. The ultimate perceptual organization of the scene, at all scales, depends on the preponderance of all evidence” (Shinn-Cunningham, 2008).

Object formation directly influences our perception and processing of complex environments. In all sensory modalities, generally speaking, the processing of a complex environment happens by focusing on one object only while other objects mold the perceptual background. (Duncan, 2004) In vision scene, there usually are multiple objects which compete for attention. Which object is attended in the end depends on salience (“the perceptual strength of an input based purely on stimulus attributes“ (Shinn-Cunningham, 2008)) of the objects and the influence of the top-down (voluntary) attention. These factors bias the competition towards the favored object with required perceptual features (Desimone and Duncan, 1995). That is why it is called a biased competition. (Desimone and Duncan, 1995)

Even though observers select what they attend based on low-level features, attention tend to operate on objects (Serences et al., 2005). In other words, even when the attention of an observer is spatially focused, his sensitivity to other features which are part of the object placed in the attended location is enhanced too. Therefore it can be claimed that object formation is linked with selective attention and the object is the perceptual unit of attention.

3.2.2 Shifting attention

Evidence attests that we are able to listen to one object at a time only. Listeners might have difficulties when making judgments about the relative timing of events across streams, but not within (Bregman, 1990). In an experiment when listeners need to divide their attention between two speech streams which are close to each other in space and asked to report words from the two streams, they are able to recall many words from both streams, but intermingle words of the two messages (Best et al., 2006). On the contrary, when the two streams are distant in space, the listeners confuse words across streams less, but also recall fewer words in general (Best et al., 2006). These results suggest that the more distant the competing streams are from each other, the more extensive the suppression of the competing stream which is in the perceptual background.

One could ask a question the, how is it possible, that in everyday situations we are able to listen to and understand multiple sound sources, especially when attending social meetings where the conversation flow is chaotic and unpredictable?

The answer could lay within a concept of switching attention between objects in complex environment, a time-sharing attention switching between competing sources. Even though we do not perceive everything of the signal content, we are still able to fill in missing pieces of the information (Shinn-Cunningham, 2008). Moreover, the short-term sensory memory can help us in this filling-in process, as we replay mentally the fragments of the input signal which we did not pay attention to initially (Shinn-Cunningham, 2008).
Because of the facts that switching attention takes ordinarily 100 - 200 ms and the sensory memory declines with time, some pieces of information from a newly attended stream may be missed even if the listener switches attention afterwards (Shinn-Cunningham, 2008). In addition, auditory streams build up over time as stated above, which might enhance the listener´s ability to concentrate on the stream which is in perceptual foreground and to understand its content (Shinn-Cunningham, 2008).

Moreover, when the attention is sustained on one object in a complex environment, it might yield a bigger refinement of selectivity for the attended object over time. “In turn, switching attention to a new object may reset object formation and therefore reset attentional selectivity” (Best, Ozmeral and Shinn-Cunningham, 2007).

To conclude, switching attention between streams degrades the listener´s performance as a direct cost of switching attention. Additionally, because switching attention cancels streaming, it results in canceling out also benefits of object build up (Shinn-Cunningham, 2008). “The cost of switching attention between objects may not only be related to the time required to dis- and reengage attention but also to the time it takes to build up an estimate of the identity of an object in a scene.“ (Best, Ozmeral and Shinn-Cunningham, 2007).

3.3 Crossmodal perception and attention

Most of attentional research studied unimodal influence on attention so far. They were mostly focused on visual perception or selective hearing, only few were dedicated to interactions among different modalities. Attention was also comprehended as a specific phenomenon within a certain modality, independent from other modalities. (Andoga and Kopčo, 2005)

Since the stimuli from environment are perceived multimodally (e.g. by more than one sense) it is possible that there exist interactions which should be taken into account when considering attention. For instance, when listening to a person in a noisy surrounding, the speech comprehensibility becomes more difficult. The hearing perception itself is not sufficient so a person helps himself by using visual information, in other words, following the speaker´s lips´ movements, gestures, or mimics.

Visual information has a great disposition to influence heard information. To demonstrate this, the McGurk effect can be described. Participants are presented by a video showing a person pronouncing a syllable 'ga' and at the same time another syllable 'ba' is played through loudspeakers. To the question what the participants heard they surprisingly did not answer either of the presented syllables, but another one ('da') which is somehow a combination of the two interacted ones.

Crossmodal interactions can influence perception in a positive or negative manner. The negative effect, of which manifestation is also the McGurk effect, occurs mostly when the information perceived by different modalities contradicts. This leads, firstly, to a combination of the stimuli, which is of some kind of weighted average. Secondly, to retaining one of the stimuli and suppressing the other one. And, thirdly, to a mutual suppression of both of the stimuli and confusion. (Kopčo, 2007)

A positive effect on neuronal level is that neuronal inputs from specific modalities sum up and transform in a way that if it comes to boosting of the signal only when the two stimuli of different modalities are weak enough. If one of the unimodal stimuli is strong it may attenuate the other stimulus. The intensity of the stimuli is therefore cardinal. (Tomoriová, 2007)

Studying of crossmodal interactions helps to investigate attention in the sense whether it is specific for particular modalities or shared. Since the attention is limited, it is considered regarding its sources. If the attentional sources are not shared and for example, tactile information (i.e. texture) is presented visually, the sources which are dedicated purely to visual stimuli will downsize. Contrariwise, visual information presentation through touch would open up visual sources. (Kopčo, 2007) Yet if the attentional sources are shared, presenting a stimulus of a certain modality by other modality would downsize the communal sources, never open up. Existence of communal sources can be supported by the phenomenon that when we expect a stimulus of a certain modality it will wreak expecting stimuli of other modalities coming from the same location. If the stimulus is of the expected modality it will load the attentional sources less than if of another modality (Tomoriová, 2007).

If to an observer stimuli of different modalities are presented he can either integrate into one event or perceive them as two separate events. It depends on many factors, for example, temporal or spatial proximity of stimuli, or their semantical congruency. It was shown that if a sound was presented together with a visual stimulus, localization of the sound would shift in direction towards the visual stimulus. Yet when the distance or time delay of the two stimuli increased, the effect would diminish (Spence, 2007).

3.4 Auditory spatial attention

Depending on whether an auditory task demands attending to a location in space or to an auditory object or feature, there are different activations of the 'where' and 'what' auditory pathways. Attentional mechanisms modulate neuronal activity which encodes the spatial location information and the acoustic attributes of the selected target stimuli and the early sensory representations of the attended stimuli. (Ahveninen et al., 2006)

Spatial attention is supramodal, which means, crossmodal spatial cues (visual, tactile) can enhance the auditory ERP when the acoustic stimuli are presented at the same location (Fritz et al., 2007). Some recent studies referred to the frontoparietal neural network which is shared for both auditory and visual spatial attention. Deterioration occurred in this network would lead to combined auditory and visual neglect (Clarke and Thiran, 2004; Spierer et al., 2007). The network includes prefrontal cortex which is involved in tracking goals of task and biasing sensory cortex towards task-relevant stimuli. Other part of the network, the anterior cingulate cortex, plays critical role in executive attentional control. Frontal eye fields is involved in attentional orienting. Another part which also contributes is posterior parietal cortex (Gottlieb, 2007). Other studies showed that the pathways of visual and auditory attentional top-down modulation of owl and visual descending pathways of primate are very closely parallel (Moore and Armstrong, 2003). From this it can be suggested that the brain´s strategy to direct the spatial attention is common for both sensory modalities and also that the top-down pattern of the attentional modulation is conserved across species (Knudsen, 2007).

Regarding crossmodal interaction of auditory and visual attention, there are many similarities between them. Two mechanisms are thought to function in both modalities:

bottom-up (automatic, image-based) cues, and top-down (attentional, task-dependent) cues (Fritz et al., 2007). Another similarity is that in both modalities, the attention modulates spatial and non-spatial features processing. There is also a neuroimaging evidence for crossmodal audio-visual interaction, which is evident visual modulation of activity in auditory cortical areas (Kayser et al., 2007). These similarities in perception between audition and vision in complex scenes also give the suggestion that there are common neural mechanisms of attention control across modalities (Shinn-Cunningham, 2008).

Several studies investigated how auditory cortical responses to an auditory stimulus are influenced by other ongoing sensory activity, attended or unattended. In conjunction with model of attention with limited resources, the usual result of these studies is that when a visual stimulus is presented and attended in a visual task, attention is drawn away from auditory stimuli which causes a decreased activity in auditory cortex (Woodruff et al., 2007), but not always. On the other hand, many studies also showed that paying attention to the auditory stimuli enhanced activity in auditory cortical areas and this result was confirmed by extended studies examining unimodal and bimodal conditions (Johnson and Zatore, 2005). In the unimodal condition tasks, when the subjects were actively listening to auditory stimuli, generally greater responses were seen in secondary auditory cortical areas. In the bimodal condition tasks, the auditory cortical responses were found to be enhanced during auditory attention task and suppresses during visual attention task. Another matching finding to these was brought by analysis of the functional connectivity between visual and auditory cortical areas during visual and auditory task. It showed a reciprocal inverse relationship, in other words a decreasing visual response when increased auditory response and vice versa. More experiments and studies were performed with similar supporting results, including neuroimaging or TMS. Results of all these studies attest that the flow of attended information is regulated by a top-down sensitivity control by modulating relative strengths of particular sensory information channels. (Fritz et al., 2007)

Although the previous stated studies dealt with “competition between sensory channels in the limited resources model, in relatively simple low-level task contexts (such as pitch discrimination or contrast discrimination) there may be no conflict over limited attentional resources since there are apparently sufficient separate attentional resources for both vision and audition” (Fritz et al., 2007).

When we look at experiments focused on source localization with bimodal cues we can find some of the following results. A study examining the effect of focused attention along the spatial dimension in a highly uncertain multitalker (more speakers, only one was the target) listening situation showed: In the case that listeners did not get a priori information about target location, or about correctness of the target choice, their performance was relatively poor. When there was a cue divulging the correct target to choose, preceding the target, but location was uncertain, performance increased significantly relative to the uncued condition. Also the performance increased significantly when the information about spatial location was provided in advance, in both cued and uncued conditions. When the target location was certain, correct identification performance proportion was higher than 0.9, independent on cued/uncued condition. (Kidd et al., 2005)

Another study investigated the effect of providing simple visual cues indicating either when or where the target would appear (but no information about the target content) in a complex acoustic mixture. A visual cue indicating ‘where’ (which loudspeaker out of five contained the target) improved performance for both kinds of target content. A cue indicating ‘when’ (which time segment out of five contained the target) also increased accuracy, but differently for particular contents of the target. “These results suggest that in real world situations, information about *where* a target of interest is located can enhance its identification, while information about *when* to listen can also be helpful when targets are unfamiliar or extremely similar to their competitors” (Best, Ozmeral and Shinn-Cunningham, 2007).

4 Event-Related Potentials

Electroencephalography (EEG) provides a very good medium to understand neurobiological regulation, and has also the potential to evaluate neurotransmission. Event-related potential (ERP), sometimes referred as time-locked EEG activity, helps to capture neural activity which is related to both, cognitive and sensory processes (Sur and Sinha, 2009).

ERPs are tiny electrical voltages generated by the brain structures as a response to specific stimuli, called events (Blackwood and Muir, 1990). These EEG changes are time locked to motor, sensory, or cognitive events and since the method is noninvasive it provides safe approach to studying of psychophysiological correlates of mental processes. Generally speaking, ERPs reflect a summed activity of postsynaptic potentials in neurons. Only when a large group of cortical pyramidal neurons (thousands or millions) which are similarly oriented fire in synchrony in order to process an information, an ERP can be recorded (Peterson *et al*., 1995). Human ERPs can be of 2 categories: The early waves are components with a peak approximately within the first 100 ms after a stimulus and they are often termed 'exogenous' or 'sensory' because they depend mainly on physical parameters of the stimulus. On the contrary, late waves are ERPs generated later on and reflect the manner how the subject evaluates the presented stimulus and are usually termed 'endogenous' or 'cognitive' because they examine information processing. ERP waveforms are described in manner of latency and amplitude. (Sur and Sinha, 2009)

4.1 ERPs as a tool for assessing attention

It is predicted that sensory ERP components are enhanced for those stimuli which are presented in an attended channel in comparison with stimuli which are presented in an ignored channel. Moreover, this effect is expected to be the same for both standard and deviant stimuli since it is assumed that attention operates before completed perception, in other words, before the brain determines whether the given stimulus is also the target. A few studies examined whether the attention to stimuli could enhance the brainstem auditory evoked responses which are evoked within 10 ms after stimulus onset. They provided a convincing evidence that attention does not influence these early components even when highly focused (Woldorff et al., 1991). Attention seems to have no impact on the very early transmission of auditory information (originating in the cochlea and spreading through the brainstem to auditory processing nuclei). Nevertheless, highly focused attention influenced ERPs in the 20–50 ms latency range, corresponding to the midlatency responses. The auditory midlatency responses arise primarily from the auditory cortex, thorough possibly with a contribution from upper brainstem and some nuclei of thalamus. Concluding this, auditory attention does occur to influence auditory sensory activity even before the signal reaches cortex. Measuring ERPs of latency lower than 50 ms is technically very demanding because these early ERPs are very small and can only very highly focused attention can influence them. (Luck and Vecera, 2002)

The most commonly observed effect of auditory attention is a later effect with the latency of the N1 wave (cca 100 ms). It was discovered that the N1 amplitude was greater when the stimulus was actively attended comparing to unattended. Since the N1 wave is a sensory response, attention was concluded to operate during perception. Moreover, the effect of N1 wave was the same no matter whether using standard or deviant stimuli, which indicates that N1 reflects selective processing of the attended channel even before the brain determined whether the stimulus is standard or deviant. (Luck, 2005)

Another later component, the P3 wave, was also observed for deviant stimuli in the attended channel, but not for the standard ones, which indicates that this component (like all the later components) reflects processes of classification of the stimulus as standard or deviant. In addition, P3 wave is not elicited by deviant stimuli in the unattended channel, what provides an evidence that the perceptual processing within the unattended channel was suppressed. Although it was clear that attention has an impact in the N1 latency range, it had to be questioned whether the effect was a result of modulation of the exogenous (stimulus-evoked) attention or rather endogenous (internally triggered) attention. It was noticed that the effect lasted significantly longer than the normal N1 wave, and from this observation it was finally concluded that this effect of attention occurred on the basis of the endogenous component, which was named the processing negativity. This processing negativity component overlapped with actual N1 component which created the appearance of larger N1 response to the attended stimuli. A newer study showed that attentional processing does not necessarily elicit a new peak independently from N1 but can simply modify the amplitude of the exogenous N1 wave, and named it Nd wave (for negative difference wave). However, it was shown that the late part of attentional effect reflects different neural processes than described by processing negativity. (Luck, 2005)

To assess the auditory attentional processes, the late-latency auditory ERPs are usually used. They are typically obtained from Fz electrode. The electrode site does not reflect the origin of ERP deflections accurately because auditory stimulation activates more pathways than just the one from ear to brain. The peaks of late-latency auditory ERP mostly represent summed neuronal activity from several different generators. For instance, even the main ERP used as marker of auditory attentional processes, N1, recorded frontally probably reflects a combined activity of more than 5 neural generators (Bishop et al., 2007).

The late-latency auditory ERPs are a good choice as an index of the real-time brain response to auditory stimuli while minimizing demands on listeners´ performance (Bishop et al., 2007). Fig. 2 depictures ERPs which are usually found when attention is involved it performing an auditory task.

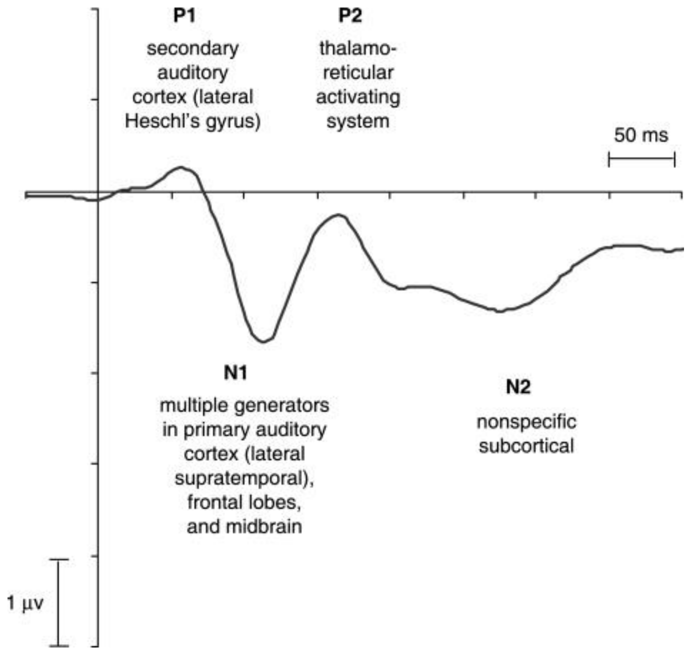

Figure 2. Grand average auditory ERP with a description which neural structures are involved in their production (Bishop et al., 2007)

5 Experimental part

5.1 *Characteristics of the object of research*

The influence of attentional cueing on sound localization was examined and analyzed in this experiment. The experimental design and analyzes were based on the ones used in few previous studies (mentioned in the Introduction).

We collected two types of data, behavioral and electrophysiological. Therefore behavioral and electrophysiological part are described and analyzed separately, after describing the task in general.

5.2 *The mode the data have been collected and their resources*

14 subjects (10 male, aged 20 - 38 years) participated in the two-session experiment. All participants were without any known hearing deficiencies. Some initial practice trials on each of the different experimental conditions were given prior to data collection. All provided written informed consent as approved by the P. J. Šafárik University in Košice.

5.3 *Working procedures*

Auditory and visual stimuli were generated using Matlab (Mathworks, Natick, MA). The experiment was controlled using Matlab with the Psychtoolbox 3 extension (Brainard, 1997). Sound stimuli were presented using Etymotic Research (Elk Grove Village, IL) ER-1 insert headphones connected to a Datapixx system (VPixx Technologies, Saint-Bruno, QC). During the experiment subjects were sitting in a sound-treated booth (Eckel Laboratories).

5.3.1 Types of trials

According to the modality of cues used, two different trial types molded essential conditions of this experimental design.

Visual cue

Visual cue trials included a 100 ms white dot presented at either 0 degrees or +/- 25 degrees on the computer screen at the beginning of each trial (see Figure 3).

Auditory cue

Auditory cue trials were similar to the visual primer trials, including an auditory 100 ms click train at 170 Hz served as an auditory cue at the beginning of each trial (see Figure 3).

5.3.2 Design of the experiment (Subjects´ task)

Subjects were directed to fix their gaze at +12.5° or -12.5° position (i.e. 12.5° to the left or to the right), marked with a small white dot on the computer screen, during the whole trial (balanced across trials). To begin the trial, a cue lasting for 100 ms was presented (see descriptions of cues above). Subjects were instructed to pay attention to the cue and to expect target stimulus from the same position as the presented cue. At 800 ms time point an auditory target was presented either at 0 central position or +/-25 degrees lateral position (ipsilateral with the fixation point) through inserted earphones. The auditory target (a click train lasting for 200 ms) consisted of two, 100 ms clicks played successively. The first click of the train was presented at 0 or +/-25 degrees, and the second click of the train at a location slightly shifted (4.2° for central and 8.4° for lateral position) relative to the location of the first one. After target presenting, a subject´s response was expected, conveyed by changing the color of the fixation point to grey. The subject's task was to respond whether the target moved to the left or to the right (using 1 or 2 keyboard buttons, respectively).

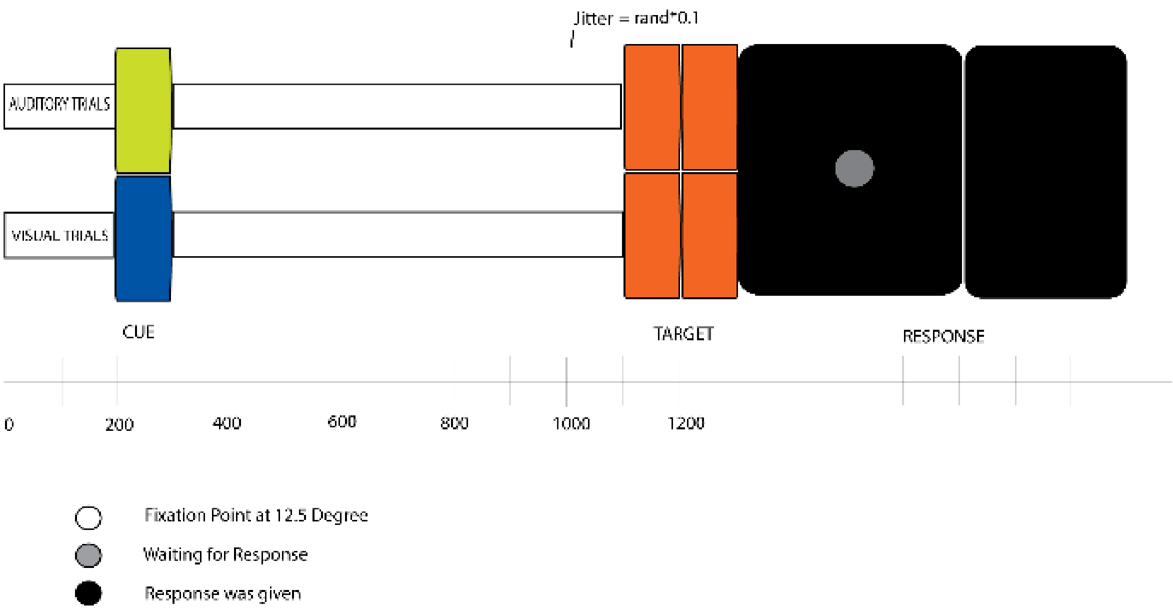

Figure 3. Schematic design of the experiment, depicting the trial sequences and the two experimental conditions.

5.3.3 Experimental conditions

5 condition classes were assessed and resulting in 32 (2^5) combinations of conditions. The condition classes are described below.

Cue modality

In ‘visual’ trials, visual cues (see above) were used.

In ‘auditory’ trials, auditory cues (see above) were used.

Cue validity

In ‘valid (matched)’ trials, the target stimuli were presented at the same location as the cue (both cue and target at 0° or 25° location).

In ‘invalid (mismatched)’ trials, the target stimuli were played at the opposite location from the cue (if cue at 0° then target at 25°; if cue at 25° then target at 0°).

Fixation point

In trials with ‘left fixation’, fixation point at -12.5° was used.

In trials with ‘right fixation’, fixation point at +12.5° was used.

Target position

In trials labeled ‘central’, target was presented at the 0° location.

In trials labeled ‘lateral’, target was presented at the 25° location, ipsilateral to the fixation point side.

Target shift

Trials labeled ‘towards’ or ‘away’ depending on the shift of the target sound relatively to the fixation point (towards the fixation point or away from fixation point, respectively).

**Nocue condition*

In order to record baseline performance, nocue condition trials were presented to participants before starting actual experimental sessions. In these trials no cue was presented before the target.

Trials were organized into block of trials (each block contained 40 trials) and they were organized into sessions (each session contained 20 blocks of trials). Each of the subjects absolved two sessions, which resulted in obtaining 1600 trials per subject. Particular conditions were randomly distributed within these trials, but balanced across the whole experiment, therefore number of trials for each particular condition was 50.

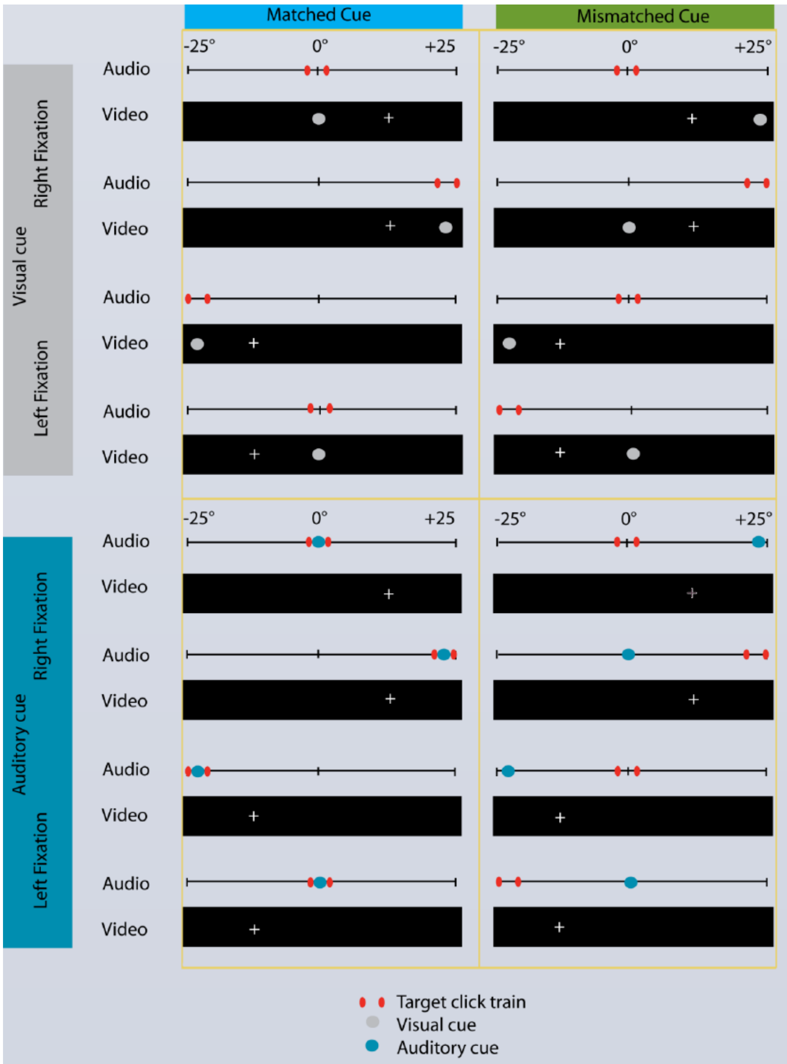

Figure 4. Scheme of the experimental design considering the particular conditions

5.3.4 Data analysis

Data were analyzed for each condition alluded above. For most of the conditions, there was a left-right symmetry in results. Therefore, data collected with fixation point on the left were mirror-flipped and combined with the data collected with fixation point on the right.

5.4 Results of the thesis

5.4.1 Behavioral part

5.4.1.1 Hypothesis

We hypothesized following:

1. Automatic attention would improve performance for valid trials and have only a little effect or decrement in accuracy for invalid trials.
2. Effect of automatic attention would be modality dependent.
3. Specifically, based on Maddox study (Maddox et al., 2014), we assumed that there would be better performance in valid visual cue trials compared to valid auditory cue trials.

5.4.1.2 Methods of evaluation employed and interpretation of results

For all participants, the percentage of correct responses and their means were computed to see how accurately subjects responded. Data of all subjects were averaged across various conditions.

All graphs were generated by using Matlab (Mathworks, Natick, MA). Graphs of subjects´ performance display the % correct.

To identify significant differences between experimental conditions, a repeated measures ANOVA was performed on RAU-transformed % correct data.

5.4.1.3 Results

Figures 5 and 6 show performance (percentage of correct answers) of all individual subjects (marked by color bars) for particular conditions (labeled by 4-5-letter names, the legend is below), for auditory and visual conditions, respectively. Means across all subjects for each particular condition are depicted by black bars.

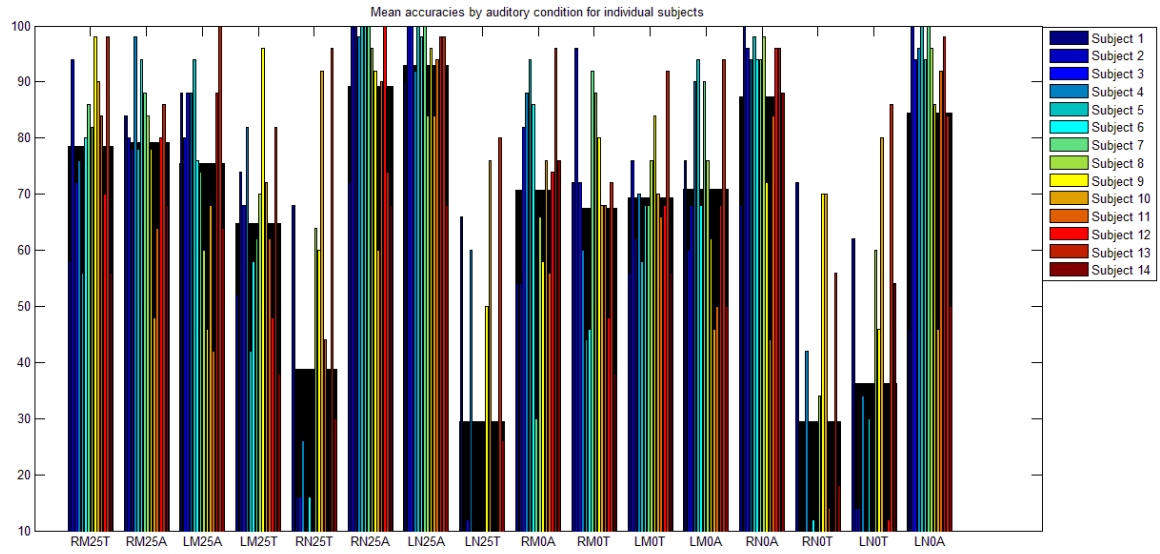

Figure 5. Performance of individual subjects in different conditions in auditory trials

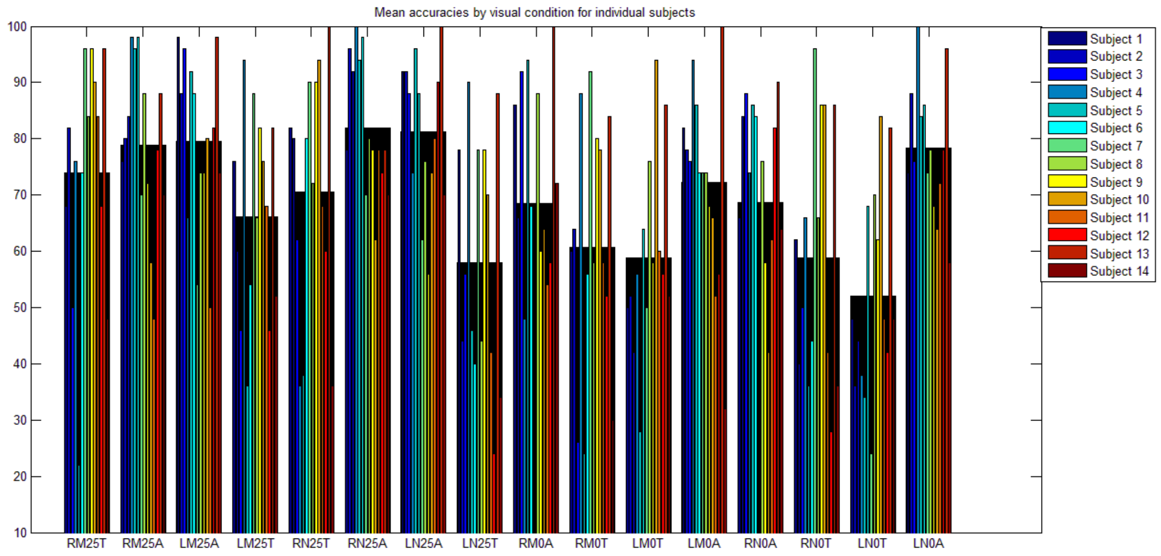

Figure 6. Performance of individual subjects in different conditions in visual trials

Legend of graph labels:

FP – fixation point
R – right position of FP
L – left position of FP
M – matching position of cue and target
N – nonmatching position of cue and target
25 – 25° azimuth of cue
0 – 0° azimuth of cue
T – target shift towards FP
A – target shift away from FP

For identifying significant effects in particular conditions and significant differences in performance in particular interactions between conditions rANOVA analysis was done. Table 1 shows the results.

Effect/interaction	df	F	pValue
Cue	1,13	7.32	0.018*
Position	1,13	35.97	< 0.001*
Fixation	1,13	3.93	0.069
Matching	1,13	47.13	< 0.001*
Cue x Matching	1,13	12.62	0.004*
Cue x Position	1,13	6.62	0.023*
Fixation x Position	1,13	5.87	0.031*
Cue x Fixation x Matching	1,13	3.84	0.072

Cue x Matching x Shift	1,13	23.318	< 0.001*

Table 1. rANOVA table for testing within-subjects effects

The ANOVA showed a significant effect of the Cue type (F(1,13) = 7.3; p < 0.05), indicating that in auditory cue trials, participants performed worse than in visual cue trials.

There was also a main effect of Position ((F1,13) = 35.9; p < 0.001) resulting in a less pronounced decrease in performance in lateral than central position.

ANOVA also indicated a significant Cue by Position interaction (F(1,13) = 12.6; p < 0.01) (Figure 7). In lateral position performance was better for visual cue compared to auditory cue.

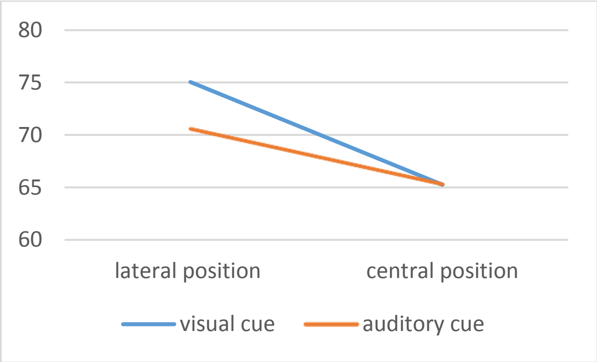

Figure 7. Cue modality by Position of cue interaction

As hypothesized we also found a main effect of Matching ((F1,13) = 47.13; p < 0.001) resulting in a significantly better performance for matched than mismatched trials. There was also a significant Cue x Matching interaction. The difference between 'Match' and 'Nonmatch' condition for visual cue trials was not significant (F(1,13) = 1.5; p = 0.24), but for auditory cue trials this difference was significant (F(1,13) = 28.9; p < 0.001), the performance was better for 'Match' experimental conditions (Figure 8).

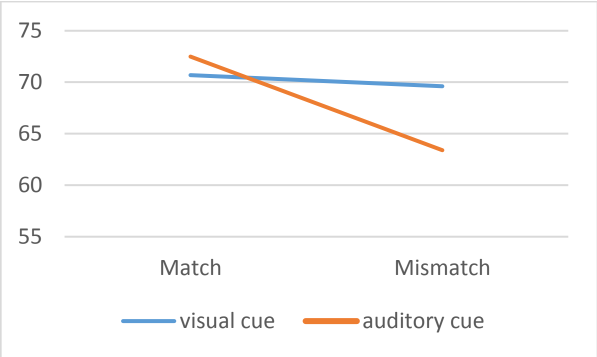

Figure 8. The difference between 'Match' and 'Nonmatch' condition for visual and auditory cue trials

Finally, we found no significant main effect of Fixation point location (F(1,13) = 3.93; p = 0.07). The performance was relatively the same for left fixation and right fixation. On the other hand, there was a significant Fixation x Position interaction. With right fixation, performance was better for laterally presented target stimuli (Figure 9). Cue by Fixation by Matching interaction was not significant.

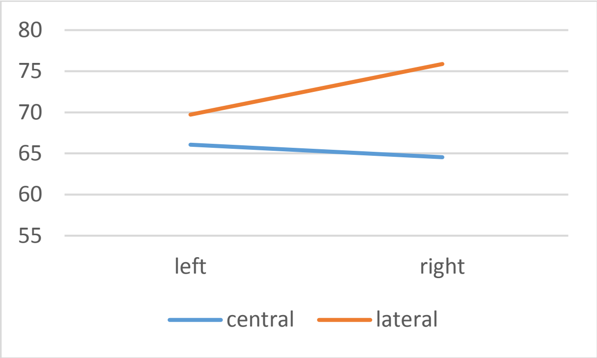

Figure 9. Eyes fixation by Cue position interaction

5.4.1.4 Discussion

Performance was better when visual cue was used

We aimed to compare subjects´ performance when visual cue versus auditory cue was used. Based on our findings it can be concluded that subjects performed better when visual cue was used. This is in line with some other studies showing that visual cues help auditory perception by guiding attention to discriminate target either by enhancing sounds near the threshold of audibility when the target is energetically masked or by enhancing segregation when it is difficult to direct selective attention to the target (Varghese et al, 2012). It seems that visual cues can provide perceptual benefits helping listeners focus selective attention on the target.

Auditory cue presented from incongruent location resulted in deteriorating performance

In our experiment all trials had either ´Matched´ or ´Nonmatched´ cues. We found that only for auditory cue trials performance was better when cue was matching the target position. This result is surprising, partly in contrast to the previous experiments which showed intelligibility and discrimination benefits of knowing where to listen (Best et al, 2007; Maddox et al., 2014). Those gains may come from facilitated selective attention (Mesgarani, Chang, 2012).

This finding is in opposite with Maddox study (Maddox et al., 2014), who found no benefit in performance for directional (informative) auditory primers compared to uninformative and better performance in spatial discrimination with visual directional (informative) primers than with visual uninformative primers, for ILD primers at both, the central and side positions, and for ITD primers only when stimulus was located on the side.

An important difference between this study and the previous studies is that our study examined only automatic spatial auditory attention since the cue was only informative at 50% of trials, thus making it unlikely that the subjects would use the cue informativeness to direct their strategic attention. However, it is possible that some strategical attention was engaged. Additional experiments need to be performed to distinguish between these two options.

Better performance in lateral than central position

We also found significant main effect of position with better performance in lateral than central position. This finding is in contrast with Maddox who found central performance better than side performance.

In our case lateral position was much easier to discriminate due to more spatial (azimuthal) difference between the two target clicks (8.4°), compared to 4.2° azimuthal difference for central click trains as is obvious from initial practice no cue trials on each experimental conditions which were given prior to data collection (Figure 9).

We also observed an asymmetry between central and lateral performance for the left vs. right fixation point (Figure 9). This asymmetry is likely due to the use of non-individualized HRTFs which might have been better matched to the individual subjects' HRTFs on the righthand side compared to the lefthand side. However, as shown in Figure 10, which shows the nocue baseline performance measured prior to the experiment, this performance was well matched across the locations in the experiment.

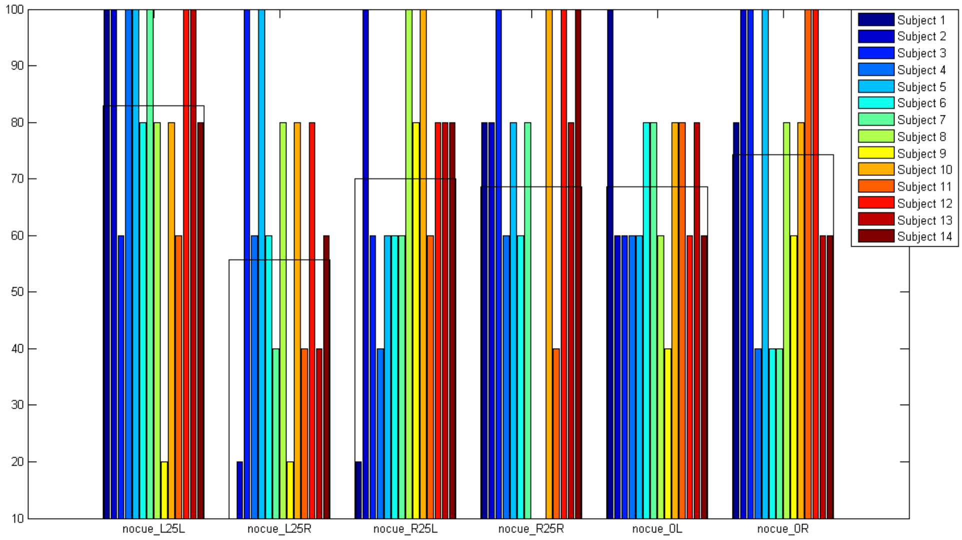

Fig.10 No cue practice trials on 6 experimental conditions. Each group of bars corresponds to performance of individual subjects (color bars) and across-subject mean (open bar) for one combination of target position (L25, R25 or 0) and direction of shift.

In lateral position performance was better for visual cue compared to auditory cue

Maddox study confirmed that gazing leftwards would shift the receptive field to the left, resulting in better discrimination of the left-lateralized sounds (Maddox et al., 2014). It is not clear how to relate this result to the current results, given that our subjects were specifically instructed not to move their eyes (and we monitored eye position using electro-oculography). It is possible that the presentation of visual cue or auditory cue induced an automatic orienting response or response planning, which then affected performance, in particular for the non-matching cues in the auditory condition.

5.4.1.5 Future studies

With regards to these studied experimental conditions and data, we are working on d-prime (subjects´ sensitivity) analysis and criterion (response bias) analysis.

5.4.2 Electrophysiological part

5.4.2.1 Aim

The aim of this part was to identify electrophysiological correlates of auditory target stimuli used in the experiment. As these correlates and as brain markers of "low-level" auditory attention (i.e., acoustic representations in sensory memory), event-related potentials (ERPs) were set. Among the auditory response elements (P1, N1, P2, N2) and the auditory attentional elements (N1, P2, P3a), the main focus was allocated to the N1, and the P2 elements.

The current goal should serve as a base for further analyses of the recorded EEG signal and extracted ERPs. Since the analyses are still in process we do not dedicate to this issue in bigger extent in this thesis.

5.4.2.2 Methods of evaluation employed and interpretation of results

EEG data acquisition took place at the Perception and Cognition Laboratory at P. J. Šafárik University in Košice. Normal-hearing subjects (as determined by audiometric screening) sat in a sound-treated booth (Eckel laboratories). A Biosemi ActiveTwo system was used to record EEG data from 32 scalp electrode positions in the standard 10/20configuration. Two extra electrodes were placed on the earlobes for reference, as well as two electrodes above and below the left eye (recording vertical EOG), and two additional ones at the outside corners of each eye (recording horizontal EOG). TDT hardware sent timing signals for all events, which were recorded in an additional channel. Recordings were re-referenced to the average of the two reference electrodes.

The preprocessing consisted of first downsampling both the EOG and EEG data from 4096 to 256. For EOG data, no filter was applied due to the fact that this action would eliminate sudden voltage changes generated from eye saccades, making it difficult to differentiate between the natural fluctuation of EOG voltage and actual eye saccades. For the EEG data, a bandpass filter from 1-40 Hz was subsequently applied to remove noise from the data. The processing for the EEG data involved epoching the trials between -0.5 to 1 second relative to the onset of the auditory primer in the auditory condition and -0.5 to 5 seconds relative to the onset of the first click of the target stimulus (occurring at 800 ms, see fig. 3 for the trigger scheme). The signals were referenced against the two external electrodes placed on the subject's earlobes. Horizontal EOG data was epoched between -0.2 second to

0.3 second relative to the onset of the first target click. This epoch interval allowed for an evaluation of the eye gaze direction during the time that the auditory stimulus was playing. It was subsequently matched to their EOG calibration data to determine if they were or were not fixating their gaze appropriately on the fixation point during the time of the sound presentation. Interpretation of saccade information: A linear regression was performed on the EOG calibration data for each subject. EEG and horizontal EOG trials contaminated with eye blink artifacts were rejected using vertical EOG data (using thresholding as the artifact rejection procedure). Trials where the subject did not fixate their gaze correctly (as determined from the horizontal EOG data during real time analysis) were also rejected.

The indices of trials in which the subject answered correctly were cross-referenced and ERPs were extracted and plotted. In order to visualize the response to the second click only, the grand average ERPs for each condition during the single click train response (obtained from the auditory session trials) were subtracted from the grand average condition ERPs of both the visual and auditory sessions. The ERP waveforms were analyzed from Cz electrode.

5.4.2.3 Results

The figures 11 and 12 display the grand average ERPs (GERPs) for 8 basic model conditions (4 for visual and 4 for auditory conditions) in trials when participants responded correctly. For each listener, the ERP was obtained by averaging responses to target stimuli (click trains) , 200 ms in duration.

Green vertical line represents the start of the trial, black vertical line represents onset of the first click of the target stimulus, and blue dashed line represents expected time of N100, e.g. 100 ms after the stimulus onset. Individual GERPs are in red, green and blue.

Since the analysis is still waging the figures show ERPs of only 3 participants, whose analysis has already been completed.

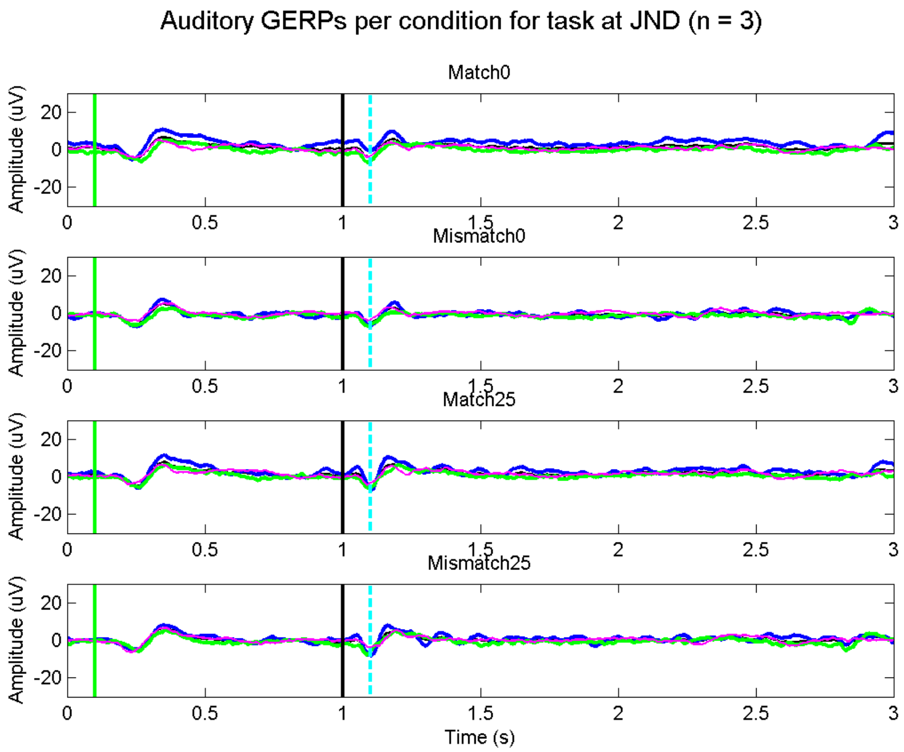

Figure 11. Auditory grand ERP per condition (n=3)

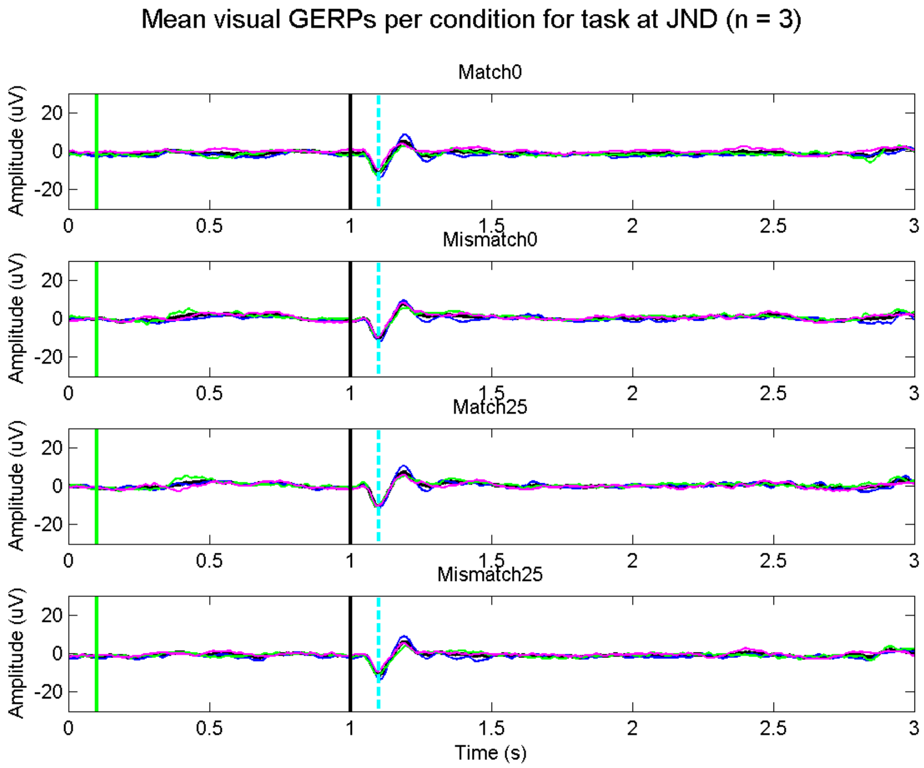

Figure 12. Visual grand ERP per condition (n=3)

As seen in the figures for both, auditory and visual trials, there is an evident negative deflection at 100 ms after the stimulus onset in every condition, which can be identified as N100 ERP element. However, in auditory trials this deflection seems to be of a larger amplitude. For evaluating this estimation, further analysis needs to be done.
Following N100, an obvious positive deflection can be found in every of the depicted waveforms, with typical shape and delay of the P2 ERP element.

5.4.2.4 Discussion

The main event-related potential elements (ERPs) of auditory spatial attention were investigated and identified in each of the experimental condition trials.

5.4.2.5 Future studies

To evaluate the differences between particular conditions further analyses are being done. Another goal would be to study how the crossmodal interaction would influence these early auditory ERP reflections of selective attention. Such effects could further substantiate the claim that selective attention operates at the level of early perceptual processing and could provide evidence regarding the role of different auditory routes in selective attention.

6 Conclusion

The aim of this thesis was to study brain mechanisms of sound localization and the effect of cued attention on it.

The theoretical base is dedicated to topics like sound perception in general, then sound localization and finally attention with the emphasis on auditory attention and attentional cueing in spatial localization tasks. The second theoretical part describes ERP as a very important tool of assessing and analyzing mechanisms of attention.

In our first experiment we studied how cueing of automatic attention influences performance in the auditory task. We compared effect of visual vs. auditory cue, effect of valid vs. invalid cue, influence of left-right situating, and influence of centrality vs. laterality of target position. We accounted also impact of direction of target shift but this condition requires more analyzing.

We confirmed our hypothesis that automatic attention would improve performance when valid cue was used. We also confirmed conclusion of Maddox paper, concretely an improvement in performance when visual cue was used comparing to auditory. When auditory cue was presented from incongruent position, this lead to deteriorating in performance. Another significant enhancement was observed when the cue was presented from lateral position vs. central, especially visual over auditory.

In the second experiment, in our experimental EEG data, we identified ERP components characteristically found in tasks which employ attention. This serves as a base for further analyses of specific components which are being processed.

Further analyses also include subjects´ sensitivity measurements and analysis of response bias.

These experiments were performed in virtual environment and this carries certain risks. HRTF, by which the locations of sounds were simulated, could possibly cause unnatural perception of the location because it depends on individual characteristics of human head. Even though a training procedure with standardized HRTF preceded the experimental one, using a real environment would be more efficient. Other enhancement of the experiment are the subject of current discussions and modifications of experimental conditions.

7 Resumé

Teoretický základ diplomovej práce

Objekty a javy v reálnom svete sú tvorené multimodálnymi zmyslovými atribútmi. Náš nervový systém spracúva informácie z rôznych zmyslových orgánov nezávisle a tieto informácie sú v určitom bode kombinované do vnemovej skúsenosti. Vnímanie je viaczmyslový proces, kde sú senzorické informácie integrované aj s aj naprieč rozličnými zmyslovými modalitami. Niektoré štúdie ukázali, že sluchové a zrakové stimuly môžu byť integrované bimodálnymi bunkami, ktoré vykazujú priestorové prekrývanie sa sluchových a zrakových vnemových polí. Tieto neuróny boli nájdené v skorých senzorických kôrových oblastiach ako napr. Colliculus superior (napr. Lakatos a kol., 2007, 2008; Kayser a kol., 2009), tiež sa však multisenzorické efekty dokázali aj v primárnych senzorických oblastiach (Lemus a kol., 2010).

Multimodálna aktivácia bola tiež nájdená v ľudskej parietálnej kôre (Bremmer et al. 2001; Bushara et al. 1999, 2003; Cusack et al. 2000; Warren et al. 2002) a Sulcus intraparietalis, ktorý sa v tejto oblasti bežne označuje ako LIP (lateral intraparietal sulcus) a MIP (medial intraparietal sulcus). U nervových buniek tejto oblasti bola zistené že sú citlivé na priestorové umiestnenie aj zrakových aj sluchových stimulov (O'Dhaniel et al., 2005; Ben Hamed et al. 2001, 2002; Cohen et al. 2004; Gifford and Cohen 2004; Cohen and Andersen 2000).

Informácie z jedného zmyslu majú potenciál ovplyvniť, ako vnímame informácie z iného zmyslu. Napríklad nerelevantný zrakový stimul môže ovplyvniť schopnosť detekcie sluchového stimulu (Lovelace a kol., 2003) ako tiež vnímanej hlučnosti (Odgaard a kol., 2004).

Pozornosť uľahčuje výber objektov, javov, alebo priestorových oblastí v komplexnom prostredí. Veľmi málo štúdií sa zaoberalo efektom pozornosti na sluchovú lokalizáciu. Ešte menej štúdií skúmalo, či je jej efekt závislý od vnemovej modality. Len zopár predošlých výskumov zisťovalo, či zameranie automatickej (exogénnej, mimovoľnej, stimulom-poháňanej) alebo strategickej (endogénnej, vôľovej, cieľom-riadenej) pozornosti pomocou sluchovej nápovedy môže zlepšiť lokalizáciu zvuku (Spence & Driver, 1994; Sach a kol., 2000; Kopco a kol., 2001). Výsledky ukázali, že napovedanie lokality spôsobilo zlepšenie reakčných časov (Spence & Driver, 1994), ale malé (Sach a kol., 2000) resp. žiadne zvýšenie lokalizačnej presnosti (Kopco a kol., 2001). Možné dôvody boli, že testované SOA boli priveľmi krátke na zameranie pozornosti a sluchová nápoveda nie je efektívna, pretože sluch nie je primárne priestorovou modalitou.

Ciele práce a hypotézy

Nedávna behaviorálna štúdia preukázala zvýšenie diskriminačnej schopnosti sluchovo napovedanej lokality, ak pohľad počúvajúceho bol nasmerovaný smerom ku zvukovému stimulu (Maddox et al., 2014). Tento efekt bol však preukázaný len pri použití binaurálnych vodítok priestorovej lokalizácie (ITD = „interaural time difference" a ILD = „interaural level difference"). Cieľom tejto diplomovej práce bolo nadviazať na túto štúdiu a rozšíriť poznatky tejto tematiky využívajúc monaurálne vodítko priestorovej lokalizácie, HRTF (head related transfer function) a otestovaním, či priestorovo sluchová pozornosť tiež vplýva na toto krosmodálne zlepšenie výkonu v danej úlohe.

Stanovili sme hypotézu, že automatická pozornosť zlepší výsledky v pokusoch s pravdivou nápovedou, ale v pokusoch s nepravdivou prinesie malý efekt alebo pokles v presnosti lokalizácie. Taktiež sme predpokladali, že tento efekt automatickej pozornosti bude závislý od použitej modality nápovedy (sluchová alebo zraková). Ďalší predpoklad, ktorý je osobitne založený na Maddoxovom článku je, že v pokusoch s pravdivou zrakovou nápovedou sa dosiahnu lepšie výsledky ako s použitím sluchovej nápovedy. Cieľom tejto práce bolo tiež získať behaviorálne údaje pre elektrofyziologickú analýzu zmien v sluchových ERP („event-related potentials") v kôrových častiach mozgu.

Metodológia experimentálnej časti

14 subjektov sa zúčastnilo experimentu, pozostávajúceho z dvoch stretnutí. Všetci účastníci boli bez známej poruchy sluchu. Pred samotným zberom dát subjekty podstúpili tréning v rôznych experimentálnych podmienkach.

Pokusy so zrakovou a sluchovou nápovedou

Vizuálny pokusný set obsahoval 100 ms zrakovú nápovedu, ktorou bola biela bodka objavujúca sa po dobu prvých 100 ms na pozícii buď 0° alebo +/- 25° (horizontálneho azimutu) na počítačovej obrazovke (pozri Figure 3). Subjekty boli inštruované k tomu, aby fixovali svoj pohľad na pozícii +12.5° alebo -12.5° na obrazovke počas celého pokusu a k tomu, aby dávali pozor na nápovedu a očakávali cieľový stimul na rovnakej pozícii ako objavujúca sa nápoveda. V čase 800 ms každého jednotlivého pokusu sa objavil sluchový cieľ buď na pozícii 0° alebo +/- 25° (ipsilaterálne s fixačným bodom) prehratý cez vložené slúchadlá. Sluchový cieľ pozostával z dvoch 100 ms zvukov („klikov") nasledujúcich za sebou. Prvý klik bolo prehratý z pozície 0° alebo +/- 25° horizontálneho azimutu a druhý klik z mierne posunutej pozície oproti prvému. Takto cieľový stimul vyvolával dojem posúvajúceho sa zvuku. Úlohou subjektu bolo po zmiznutí zvukového stimulu odpovedať, či sa cieľ pohyboval doľava alebo doprava (s použitím numerických klávesov 1 a 2 na počítači).

Sluchový pokusný set bol podobný vizuálnym pokusom, pozostávajúc zo sluchového cieľa, ktorý bol prehraný v čase 800 ms s trvaním 200 ms, avšak ako sluchová nápoveda slúžil iný, 170 Hz klik trvajúci tiež 100 ms a prehratý na začiatku každého z týchto pokusov. Subjekty mali v zmysle inštrukcií dávať pozor na nápovedu a očakávať cieľový stimul v rovnakej pozícii ako nápovedu a odpovedať rovnakým spôsobom. Schematické zobrazenie experimentálneho dizajnu je ukázané vo Figure 3.

V „zhodujúcich sa“ pokusoch boli cieľové stimuly prezentované v tej istej polohe ako nápoveda. V „nezhodujúcich sa“ pokusoch bol cieľ prehraný z inej polohy ako nápoveda (pre nápovedu z 0° zodpovedal cieľový stimul z 25° a naopak).

Behaviorálna časť

Dáta boli analyzované pre sluchové a vizuálne nápovedy ako „Zhodné“ versus „Nezhodné“ podmienky a pre stredovú (0°) a laterálnu (25°) pozíciu cieľa. ˇˇĎalšími zohľadnenými experimentálnymi podmienkami bola fixácia pohľadu (vľavo alebo vpravo) a tiež posun cieľového zvuku (ku alebo od fixačného bodu). Pre každý subjekt bola presnosť lokalizácie vyrátaná ako percento správnych odpovedí (vykreslenie podielu správnych odpovedí možno vidieť vo Figure 5 a 6). Tieto dáta podliehali ďalšiemu spracovaniu.

Výsledky a diskusia

1. *Úspešnosť lokalizácie je lepšia s použitím vizuálnej nápovedy*

V práci sme sa usilovali porovnať úspešnosť subjektov lokalizovať cieľový zvuk pri použití zrakovej alebo sluchovej nápovedy. Vychádzajúc z našich zistení možno konštatovať, že subjekty dosiahli vyššiu úspešnosť pri použití zrakovej nápovedy. Tento fakt je v súlade s inými štúdiami ukazujúcimi, že vizuálne nápovedy pomáhajú sluchovému vnímaniu vedúc pozornosť k diskriminácii cieľa buď vylepšeným vnímaním zvukov blízko prahu počutia, keď je cieľ energeticky maskovaný alebo zvýšením segregácie, keď je ťažké nasmerovať selektívnu pozornosť k cieľu (Varghese a kol., 2012). Zdá sa, že vizuálna nápoveda môže poskytnúť poslucháčovi výhodu v zameraní selektívnej pozornosti na cieľ.

2. *Sluchová nápoveda prezentovaná z nezhodnej polohy spôsobila zhoršenie presnosti lokalizácie*

V našom experimente boli nápovedy prezentované buď zo zhodnej alebo nezhodnej lokality oproti lokalite cieľa. Zistili sme, že iba pre pokusy so sluchovou nápovedou bola úspešnosť vyššia, keď nápoveda bola prezentovaná zo „zhodnej" lokality. Tento výsledok je prekvapujúci, čiastočne v kontraste s predošlými experimentmi, ktoré ukázali vylepšenie zrozumiteľnosti zvuku a jeho diskriminácie, ak subjekty vedeli, kde počúvať (Best a kol., 2007; Maddox a kol., 2014). Tieto výsledky mohla spôsobiť upriamená selektívna pozornosť (Mesgarani, Chang, 2012).

Toto zistenie je však opačné ako v Maddoxovom článku, ktorý nezistil žiadne zlepšenie výkonu pre informatívnu („zhodnú") sluchovú nápovedu v porovnaní s neinformatívnou a tiež našiel lepší lokalizačný výkon v pokusoch s vizuálnou informatívnou nápovedou ako s neinformatívnou, pre ILD nápovedy aj v stredových aj v okrajových pozíciách a pre ITD nápovedy len ak stimul bol umiestnený laterálne.

Dôležitý rozdiel medzi touto štúdiou a predošlými štúdiami je, že len automatická priestorová sluchová pozornosť bola skúmaná, pretože nápoveda bola informatívna (jej pozícia sa zhodovala s pozíciou cieľa) v 50% pokusov, teda spôsobujúc ako nepravdepodobné, že by ju subjekty využili na nasmerovanie svojej strategickej pozornosti. Napriek tomu je možné, že určitá strategická pozornosť bola zapojená. K rozlíšeniu medzi týmito dvoma možnosťami sú potrebné dodatočné experimenty.

3. *Lepšia úspešnosť lokalizácie v laterálnych ako stredových pozíciách*

Zároveň sme zistili dôležitý efekt pozície s lepšou úspešnosťou v laterálnej ako centrálnej pozícii. Toto zistenie je v protiklade s Maddoxom, ktorý našiel lepšiu úspešnosť centrálne. V našom prípade bolo oveľa jednoduchšie diskriminovať laterálnu pozíciu v dôsledku väčšieho priestorového (azimutálneho) rozdielu medzi jednotlivými klikmi cieľového zvuku, čo bolo 8.4° laterálne v porovnaní s 4.2° centrálne.

Objavili sme tiež asymetriu medzi úspešnosťou lokalizácie v centrálnej a laterálnej pozícii pre ľavý a pravý fixačný bod. Táto asymetria je pravdepodobne spôsobená použitím neindividualizovaných HRTF, ktorý mohli byť viac podobné individuálnym HRTF našich subjektov na pravej strane v porovnaní s ľavou.

4. *V laterálnej pozícii bola úspešnosť lepšia pre zrakovú nápovedu v porovnaní so sluchovou*

Maddoxova štúdia potvrdila, že pozeranie vľavo posunie vnemové pole doľava spôsobujúc výslednú lepšiu diskrimináciu ľavo lateralizovaných zvukov (Maddox et al., 2014). Nie je nám jasné, ako prepojiť tento výsledok so súčasnými výsledkami, keďže naše subjekty boli inštruované, aby nepohybovali očami (monitorovali sme pozíciu očí elektro-okulografiou). Je možné, že prezentácia zrakovej nápovedy alebo sluchovej nápovedy vzbudila automaticky sa orientujúcu odpoveď alebo proces plánovania odpovede, ktoré potom ovplyvnili schopnosť lokalizácie, konkrétne platné pre „nezhodné“ sluchové nápovedy.

Budúcnosť štúdie

V budúcnosti s ohľadom na tieto študované experimentálne podmienky a dáta plánujeme uskutočniť d-prime a bias analýzy.

Elektrofyziologická časť

Cieľom elektrofyziologickej časti bolo identifikovať elektrofyziologické koreláty cieľových sluchových stimulov prezentovaných v experimente. Ako tieto koreláty a zároveň ako mozgové markery nízko-úrovňovej sluchovej pozornosti (t.j. akustické reprezentácie v senzorickej pamäti) boli vybrané a študované ERP (event related potentials). Medzi týmito elementmi sluchovej odpovede (P1, N1, P2, N2) a elementmi sluchovej pozornosti (N1, P2, P3a), sme sa hlavne sústredili na elementy N1 a P2. Súčasný cieľ má tiež poslúžiť ako základ pre ďalšie analýzy zaznamenaného EEG signálu a vybraných ERP. Keďže sú tieto analýzy stále v procese realizácie, nevenujeme im v tejto práci veľký rozsah.

Získanie dát EEG sa uskutočnilo v Laboratóriu vnímania a kognície na Univerzite P. J. Šafárika v Košiciach. Normálne počujúce subjekty (dokázané audiometrickým screeningom) sedeli počas nahrávania EEG dát v zvukotesnej kabíne. Na zaznamenanie EEG signálu bol použitý systém Biosemi ActiveTwo, pričom bolo použitých 32 elektród v štandardnom 10/20 nastavení. Ďalších 6 extra elektród bolo umiestnených na ďalšie časti hlavy, snímajúc vertikálne a horizontálne pohyby očí (EOG).

Výsledky a diskusia

Jednotlivé ERP sluchovej priestorovej pozornosti boli skúmané v pokusoch s totožným experimentálnym dizajnom behaviorálneho experimentu. Identifikovali sme elementy N1 a P2 v každej z experimentálnych podmienok.

Budúcnosť štúdie

Na zhodnotenie rozdielov medzi konkrétnymi podmienkami sú v súčasnosti robené ďalšie analýzy. Ďalším cieľom bude študovať, ako krosmodálna interakcia ovplyvní tieto skoré sluchové ERP, ktoré sú zároveň odrazom selektívnej pozornosti. Takéto efekty by mohli neskôr byť základom tvrdenia, či selektívna pozornosť operuje na úrovni skorého vnemového spracovania a mohli by tiež poskytnúť dôkaz úlohy rôznych sluchových dráh v jave selektívnej pozornosti.

8 Bibliography

1. AHVENINEN, J., JÄÄSKELÄINEN, I.P., RAIJ, T., BONMASSAR, G., DEVORE, S., HÄMÄLÄINEN, M., LEVANEN, S., LIN, F.H., SAMS, M., SHINN-CUNNINGHAM, B.G., et al. Task-modulated ‘what’ and ‘where’ pathways in human auditory cortex. In *Proc Natl Acad Sci USA*. 2006, 103: 14608-14613.
2. ANDOGA, R., KOPČO, N. *Mozgové .truktúry a mechanizmy riadenia strategickej pozornosti*: Zborník z konferencie Kognícia a umelý .život V. Smolenice, 2005.
3. BASHINSKI, H.S., BACHARACH, V. R. Enhancement of perceptual sensitivity as the result of selectively attending to spatial locations. In *Perception & Psychophysics*. 1980, 28(3): 241-248.
4. BEN HAMED, S., DUHAMEL, J.R., BREMMER, F., GRAF W. Visual receptive field modulation in the lateral intraparietal area during attentive fixation and free gaze. In *Cerebral Cortex*. 2001, 12: 234–245.
5. BEN HAMED, S., DUHAMEL, J.R., BREMMER, F., GRAF, W. Representation of the visual field in the lateral intraparietal area of macaque monkeys: a quantitative receptive field analysis. In *Experimental Brain Research*. 2001, 140: 127–144.
6. BENSON, D.A., TEAS, D.C. Single unit study of binaural interaction in the auditory cortex of the chinchilla. In *Brain Res*. 1976, 103: 313-338.
7. BEST, V., et al. The influence of spatial separation on divided listening. In *J Acoust Soc Am*. 2006, 120: 1506–1516.
8. BEST, V., OZMERAL, E.J., AND SHINN-CUNNINGHAM, B.G. Visually guided attention enhances target identification in a complex auditory scene. In *J. Assoc. Res. Otolaryngol*. 2007, 8: 294–304.
9. BEST, V., OZMERAL, E.J., SHINN-CUNNINGHAM, B.G. Visually-guided Attention Enhances Target Identification in a Complex Auditory Scene. In *JARO: Journal of the Association for Research in Otolaryngology*. 2007, 8(2): 294-304.
10. BISHOP, D.V., HARDIMAN, M., UWER, R., VON SUCHODOLETZ, W. Maturation of the long-latency auditory ERP: step function changes at start and end of adolescence. In *Developmental Science*. 2007, 10(5): 565-575.
11. BLACKWOOD, D.H.R., MUIR W.J. Cognitive brain potentials and their application. In *The British Journal of Psychiatry*. 1990, 157: 96–101.
12. BLAUERT, J. *Spatial hearing: The psychophysics of human sound localization*. Cambridge, MA: MIT Press, 1997. ISBN 9780262024136.

13. BREGMAN, A.S. *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press, 1990. ISBN 978-0262521956.
14. BREMMER, F., SCHLACK, A., DUHAMEL, J.R., GRAF, W., FINK, G.R. Space coding in primate posterior parietal cortex. In *Neuroimage*. 2001, 14: 46–51.
15. BUSHARA, K.O, WEEKS, R.A., ISHII, K., CATALAN, M.J., TIAN, B., RAUSCHECKER, J.P., HALLETT, M. Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. In *Nature Neuroscience.* 1999, 2: 759–766.
16. BUSHARA, K.O., HANAKAWA, T., IMMISCH, I., TOMA, K., KANSAKU, K., HALLETT, M. Neural correlates of cross-modal binding. In *Nature Neuroscience.* 2003, 6: 190–195.
17. BUTLER, R.A.; HUMANSKI, R.A. Localization of sound in the vertical plane with and without high-frequency spectral cues. In *Perception & Psychophysics*. 1992, 51 (2): 182–186.
18. CAMPBELL, R. *Monaural and Binaural Level Cues: A Behavioural and Physiological Investigation*: A thesis submitted for the degree of Doctor of Philosophy. Oxford: University of Oxford, 2006.
19. CARLYON, R.P. How the brain separates sounds. In *Trends Cogn Sci*. 2004, 8: 465–471.
20. CHUN, M.M. Contextual cueing of visual attention. In *Trends in Cognitive Sciences.* 2000, 4, 170-178.
21. CLARKE, S., THIRAN, A.B. Auditory neglect: what and where in auditory space. In *Cortex*. 2004, 40: 291-300.
22. COHEN, Y., ANDERSEN, R. Reaches to sounds encoded in an eye-centered reference frame. In *Neuron.* 2000, 27: 647–652.
23. COHEN, Y.E., COHEN, I.S., GIFFORD, G.W. Modulation of LIP activity by predictive auditory and visual cues. In *Cerebral Cortex.* 2004, 14: 1287–1301.
24. *Comparison of Different HRTF-Individualization Methods*: Technical Report. Munich: Technische Universität, 2014. 205 p.
25. CROW, G., RUPERT, A.L. MOUSHEGIAN, G. Phase-locking in monaural and binaural medullary neurons: implications for binaural phenomena. In *J. Acoust. Soc. Am*. 1978, 64: 493-501.
26. CUSACK, R., CARLYON, R.P., ROBERTSON, I.H. Neglect between but not within auditory objects. In *Journal of Cognitive Neuroscience.* 2000, 12: 1056–1065.

27. DARWIN, C.J., CARLYON, R.P. Auditory grouping. In *Moore BCJ*. 1995, 387–424.
28. DARWIN, C.J., HUKIN, R.W. Effectiveness of spatial cues, prosody, and talker characteristics in selective attention. In *J Acoust Soc Am*. 2000, 107: 970–977.
29. DESIMONE, R., DUNCAN, J. Neural mechanisms of selective visual attention. In *Annu Rev Neurosci*. 1995, 18: 193–222.
30. DUNCAN J. EPS Mid-Career Award 2004: brain mechanisms of attention. In *Q J Exp Psychol*. 2006, 59: 2–27.
31. FELDMAN, J. What is a visual object? In *Trends Cogn Sci*. 2003, 7: 252–256.
32. FRITZ, J.B., MOUNYA, E., DAVID, S.V., SHAMMA, S.A. Auditory attention – focusing the searchlight on sound. In *Current Opinion in Neurobiology*. 2007, 17: 1–19.
33. GIFFORD, G.W., COHEN, Y.E. Effect of a central fixation light on auditory spatial responses in area LIP. In *Journal of Neurophysiology*. 2004, 91: 2929–2933.
34. GOTTLIEB, J. From thought to action: the parietal cortex as a bridge between perception, action, and cognition (review). In *Neuron*. 2007, 53: 9-16.
35. JENKINS, W.M., MASTERTON, R.B. Sound localization: effects of unilateral lesions in central auditory system. In *J. Neurophysiol*. 1982, 47: 987-1016.
36. JOHNSON, J.A., ZATORRE, R.J. Attention to simultaneous unrelated auditory and visual events: behavioral and neural correlates. In *Cereb Cortex*. 2005, 15: 1609-1620.
37. KAYSER, C., PETKOV, C.I., AUGATH, M., LOGOTHETIS, N.K. Functional imaging reveals visual modulation of specific fields in auditory cortex. In *J Neurosci*. 2007, 27: 1824-1835.
38. KAYSER, C., PETKOV, C.I., LOGOTHETIS, N.K. Multisensory interactions in primary auditory cortex: fMRI and electrophysiology. *Hearing Research*. 2009, doi:10.1016/j.heares.2009.02.011.
39. KIDD, G.Jr., ARBOGAST, T.L., MASON, C.R., GALLUN, F.J. The advantage of knowing where to listen. In *J Acoust Soc Am*. 2005, 118(6): 3804-15.
40. KNUDSEN, E.I. Fundamental components of attention. In *Annu Rev Neurosci*. 2007, 30: 57-78.
41. KOPCO, N., LER, A., SHINN-CUNNINGHAM, B. Effect of auditory cuing on azimuthal localization accuracy. In *Journal of the Acoustical Society of America*. 2001, 109: 2377.

42. KOPČO, N., *Výpočtová a kognitívna neuroveda,* lecture notes. 2007, accesible online: http://neuron-ai.tuke.sk/~kopco/kui440/
43. LAKATOS, P., CHEN, C.M., O'CONNEL, M.N., MILLS, A., SCHROEDER, C.E. Neuronal oscillations and multisensory interaction in primary auditory cortex. In *Neuron.* 2007, 53: 279–292.
44. LEMUS, L., HERNÁNDEZ, A., LUNA, R., ZAINOS, A., ROMO, R. Do sensory cortices process more than one sensory modality during perceptual judgments? In *Neuron.* 2010, 67: 335-348.
45. LOVELACE, C.T., STEIN, B.E., WALLACE, M.T. An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. In *Brain Research Cognitive Brain Research*. 2003, 17: 447–453.
46. LUCK, S.J. *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press, 2005. ISBN 9780262621960.
47. LUCK, S.J., VECERA, S.P. Attention. In *Stevens' handbook of experimental psychology*. 2002, Vol.1, 235–286.
48. MADDOX, R.K., POSPISIL, D.A., STECKER, G.C., LEE, A.K.C. Directing eye gaze enhances auditory spatial cue discrimination. In *Current Biology*. 2014, 24: 748–752.
49. MESGARANI, N., AND CHANG, E.F. Selective cortical representation of attended speaker in multi-talker speech perception. In *Nature.* 2012, 485: 233–236.
50. MIDDLEBROOKS, J.C., PETTIGREW, J.D. Functional classes of neurons in primary auditory cortex of the cat distinguished by sensitivity to sound location. In *J.Neurosci. I*. 1981, 107-120.
51. MOORE, C.N., CASSEDAY, J.H., NEFF, W.D. Sound localization: the role of the commissural pathways of the auditory system of the cat. In *Brain Res*. 1974, 82: 13-26.
52. MOORE, T., ARMSTRONG, K.M. Selective gating of visual signals by microstimulation of frontal cortex. In *Nature*. 2003, 421: 370-373.
53. MUSICANT, A.D, BUTLER, R.A. The influence of pinnae-based spectral cues on sound localization. In *The Journal of the Acoustical Society of America*, 1984, 75(4): 1195-1200.
54. O'DHANIEL, A., MULLETTE-GILLMAN, K., YALE, E. COHEN, Y.E., GROH, J.M. Eye-Centered, Head-Centered, and Complex Coding of Visual and Auditory

Targets in the Intraparietal Sulcus. In *Journal of Neurophysiology*. 2005, 94: 2331–2352.

55. ODGAARD, E.C., ARIEH, Y., MARKS, L.E. Brighter noise: sensory enhancement of perceived loudness by concurrent visual stimulation. In *Cognitive, Affective & Behavioral Neuroscience*. 2004, 4: 127–132.
56. PAUKNER, P., ROTHBUCHER, M., DIEPOLD, K. *Sound Localization Performance*
57. PETERSON, N.N., SCHROEDER, C.E., AREZZO, J.C., et al. Neural generators of early cortical somatosensory evoked potentials in the awake monkey. In *Electroencephalography and Clinical Neurophysiology*. 1995, 96: 248–260.
58. PICKLES, J.O. *An introduction to the physiology of hearing*. 2nd ed. San Diego, CA: Academic press limited, 1991. ISBN 0-12-554754-4.
59. POSNER, M.I., NISSEN, M.J., OGDEN, W.C. Attended and unattended processing modes: the role of set for spatial location. In *Modes of Perceiving and Processing Information.* 1978, 72(1): 137- 157.
60. POSNER, M.I., SNYDER, C.R., DAVIDSON, B.J. Attention and the detection of signals. In *Journal of Experimental Psychology*. 1980, 109(2): 160-174.
61. PRINZMETAL, W., MCCOOL, C., PARK, S. Attention: Reaction time and accuracy reveal different mechanisms. In *Journal of Experimental Psychology-General*. 2005, 134(1): 73-91.
62. ROADS, C. *The Computer Music Tutorial*. Cambridge, MA: MIT, 2007. ISBN 9780262680820.
63. ROFFLER, S.K.; BUTLER, R.A. Factors That Influence the Localization of Sound in the Vertical Plane. In *J. Acoust. Soc. Am*. 1968, 43 (6): 1255–1259.
64. SACH, A.J., HILL, N.I., AND BAILEY, P.J. Auditory spatial attention using interaural time differences. In *Journal of Experimental Psychology: Human Perception and Performance*. 2000, 26(2): 717-729.
65. SERENCES, J.T., et al. Parietal mechanisms of switching and maintaining attention to locations, objects, and features. In *Neurobiology of Attention*. 2005, 35–41.
66. SERENCES, J.T., et al. Preparatory activity in visual cortex indexes distractor suppression during covert spatial orienting. In *J Neurophysiol*. 2004, 92: 3538–3545.
67. SERENCES, J.T., et al. Preparatory activity in visual cortex indexes distractor suppression during covert spatial orienting. In *J Neurophysiol*. 2004, 92: 3538–3545.

68. SHINN-CUNNINGHAM, B.G. Object-based auditory and visual attention. In *Trends in Cognitive Sciences*. 2008, *12*(5): 182–186.
69. SHINN-CUNNINGHAM, B.G., et al. A sound element gets lost in perceptual competition. In *Proceedings of the National Academy of Science*. 2007, 104.
70. SPENCE, C.J. AND DRIVER. J. Covert spatial orienting in audition: Exogenous and endogenous mechanisms. In *Journal of Experimental Psychology: Human Perception and Performance.* 1994, 20(3): 555-574.
71. SPENCE, CH. Audiovisual Multisensory Integration. In *Acoustical Science and Technology*. 2007, Vol. 28, No. 2.
72. SPIERER, L., MEULI, R., CLARKE, S. Extinction of auditory stimuli in hemineglect: space versus ear. In *Neuropsychologia*. 2007, 45: 540-551.
73. SUR, S., SINHA, V.K. Event-related potential: An overview. In *Industrial Psychiatry Journal*. 2009, 18(1): 70-73.
74. THOMPSON, D.M. *Understanding Audio: Getting the Most out of Your Project or Professional Recording Studio*. Boston, MA: Berklee, 2005. ISBN 978-0634009594.
75. TOMORIOVÁ, B. *Krosmodálne riadenie pozornosti pri priestorovom počúvaní:* Diploma thesis. Košice: Technická univerzita v Košiciach, 2007.
76. VARGHESE, L.A., OZMERAL, E.J., BEST, V., SHINN-CUNNINGHAM, B.G. How visual cues for when to listen aid selective auditory attention. In *J. Assoc. Res. Otolaryngol.* 2012, 13: 359–368.
77. WALLACH, H; NEWMAN, E.B.; ROSENZWEIG, M.R. The precedence effect in sound localization. In *American Journal of Psychology*. July 1949, 62 (3): 315–336.
78. WARREN, J.D., ZIELINSKI, B.A., GREEN, G.G., RAUSCHECKER JP, GRIFFITHS, T.D. Perception of sound-source motion by the human brain. In *Neuron.* 2002, 34: 139-148.
79. WARREN, R.M. *Auditory perception: a new analysis and synthesis*. Cambridge: Cambridge University Press, 1999. ISBN 0-521-58783-2.
80. WHALEN, D.H., LIBERMAN, A.M. Limits on phonetic integration in duplex perception. In *Percept Psychophys*. 1996, 58: 857–870.
81. WOLDORFF, M.G., HACKLEY, S.A., HILLYARD, S.A. The effects of channel-selective attention on the mismatch negativity wave elicited by deviant tones. In *Psychophysiology.* 1991, 28: 30–42.
82. WOODRUFF, P.W., BENSON, R.R., BANDETTINI, P.A., KWONG, K.K., HOWARD, R.J., TALAVAGE, T., BELLIVEAU, J., ROSEN, B.R. Modulation of

auditory and visual cortex by selective attention is modality-dependent. In *Neuroreport.* 1996, 7: 1909-1913.

83. ZHOU X. Virtual reality technique. In *Telecommunications Science*, 1996, 12(7): 46-50.