Buenos Aires – 5 to 9 September 2016
Acoustics for the 21st Century…

PROCEEDINGS of the 22nd International Congress on Acoustics

The Technology of Binaural Listening & Understanding: Paper ICA2016-161

Examining auditory selective attention in realistic, natural environments with a newly designed paradigm

Janina Fels[a], **Josefa Oberem**[a], **Iring Koch**[b]

[a]RWTH Aachen University, Institute of Technical Acoustics, Medical Acoustics Group, jfe@akustik.rwth-aachen.de / job@akustik.rwth-aachen.de
[b]RWTH Aachen University, Chair of Cognitive and Experimental Psychology, Institute of Psychology, Aachen, Germany, koch@psych.rwth-aachen.de

Abstract

The topic of the present collaborative project (Medical Acoustics and Cognitive Psychology) is the exploration of cognitive control mechanisms underlying auditory selective attention. The aim is to examine the influence of variables that increase the complexity of the auditory scene with respect to technical aspects (dynamic binaural hearing with consideration of room acoustics and head movements) and that influence the efficiency of cognitive processing. Using a binaural-listening paradigm, the ability to intentionally shift auditory attention in various anechoic setups was tested. An anechoic reproduction fails to represent realistic listening experiences. Room acoustics and distracting sources are essential parts of a natural acoustic scene. The original paradigm is limited to present relatively short stimuli (i.e. digits). Therefore, the paradigm is extended to use longer stimuli to offer more opportunities. Spoken phrases by two speakers were presented simultaneously to subjects from two of eight azimuth positions. The new stimuli were phrases that consist of a single number word (i.e., 1 to 9) followed by either the German direction "UP" or "DOWN". Guided by a visual cue prior to auditory stimulus onset, subjects were asked to identify whether the target number was arithmetically smaller or greater than five and to categorize the direction. Results showed generally greater reaction times and higher error rates using phrase stimuli than single word stimuli. The influence of spatial transition of the target speaker (shift or repetition of speaker's direction in space) was similar across both paradigms. The extended paradigm is therefore deemed suitable for studying auditory selective attention in more complex environments which can include room acoustics.

Keywords: auditory selective attention, binaural hearing, task shifting, realistic environments

Examining auditory selective attention in realistic, natural environments with a newly designed paradigm

1 Introduction

Communication in environments with reverberant energy and multiple distracting source, often described as “Cocktail-party-situations”, have been in the focus of research since Cherry [1] reported his study using dichotic-listening to investigate auditory attention. Traditionally, the task requirement is to maintain selective attention to a target source [2]. A shift of attention is therefore unfavorable and can be interpreted as a lack of a perfect inhibition of the distractor. Several studies focusing on attention shifts reported about attention shifts that were not instructed but occurred spontaneously and involuntary. (e.g., [3] [4]) These attention shifts can be described as exogen.
In contrast, [5] explicitly examined the endogenous, voluntary attention shits. Therefore, a new dichotic paradigm to explore the intentional shifting of auditory attention was introduced. A cued shift of the relevant target (i.e. the target’s position shifted between trials; e.g. in the preceding trial the target was on the left side and in the following trial the target was on the right side) resulted in a worse performance than in cued repetitions of the relevant target ([5] [**?**] [6] [7]). Furthermore, the role of attentional control in processing of task-irrelevant information in auditory task shifting was explored. The subjects’ task was always to categorize the relevant digit presented by the target speaker as smaller or larger than five and press the corresponding response button. The two presented stimuli of one trial could be congruent (both digits smaller than 5 or both greater than 5) and therefore call for the same response, or they could be incongruent (one digit was smaller and one was greater than 5) and therefore call for different responses. The “congruency effect” (see [8] for a review) implying that participants respond faster in congruent trials than in incongruent trials was confirmed [5].
To fully understand the cognitive control mechanisms underlying auditory selective attention examinations with a dichotic paradigm are not sufficient. A dichotic presentation is a highly artificial situation compared to natural listening. In reality, an acoustic scene differs enormously from a dichotic presentation of stimuli. Therefore, the authors extended the paradigm to a binaural paradigm [9] as a first step to examine the influence of variables that increase the complexity of the auditory scene and that influence the efficiency of cognitive processing. The effects of the intentional attention shift and congruency did not significantly differ between the dichotic and the binaural reproduction [10]. In a second step reverberant energy was added to the virtual scene of target and distractor [11]. Main effects and their interactions were not effected by the reverberant energy. However, stimuli consisting of monosyllabic digit words (730 ms) were too short to be negatively influenced in speech intelligibility by reverberation.
The present investigation extended the formerly dichotic paradigm to intentionally shift attention based on recorded digit words as stimuli into a binaural paradigm with more complex and longer stimuli. The stimulus presented by target and distractor now consists of a German digit/single number word and a German dissyllabic direction word (“UP”, in German “OBEN” and “DOWN”, in German “UNTEN”). Stimuli were therefore extended to (1200 ms).
The congruency effect had to be redefined. The subjects’ task was still to categorize the

relevant digit presented by the target speaker as smaller or larger than five and press the corresponding response button. These categories were mapped to the left hand buttons and the right hand buttons at the front side of a controler. Furthermore, the direction word presented by the target gave information whether the index finger (in case the direction word was “UP”) or the middle finger (in case the direction word was “DOWN”). Therefore, four response possibilities were given in a quadratic arrangement to be pressed by index fingers and middle fingers of both hands.
The new paradigm was tested in a listening test under the same conditions as the lsitening tests with the former paradigm.

2 Methods

2.1 Subjects

A number of 24 paid (8 euros) students aged between 18 and 35 (mean age: 24.5 years) participated in the experiment. Subjects were equally divided into male and female listeners. Listeners were screened to ensure that they had normal hearing (within 20 dB) for frequencies between 250 Hz and 8 kHz. All listeners could be considered as non-expert listeners since they had never participated in a listening test on auditory selective attention.

2.2 Room setup and source positions

The listening tests took place in a hearing booth (l$\times$ w$\times$ h = 9.2$\times$ 6.2$\times$ 5.0 m^3) to ensure a quiet environment during the test. Lights were turned off during the listening test to direct the focus to the aural sense other than the visual sense.

2.3 Binaural reproduction

For the binaural reproduction, head-related transfer functions (HRTFs) of an artificial head were measured in an anechoic chamber. The dummy head is a custom-made mannequin produced at the Institute of Technical Acoustics, RWTH Aachen University, with a simple torso and a detailed ear geometry ([12] [13]). Measurements ran automatically with the ITA-Toolbox ([14]) in Matlab. Interleaved exponential sweeps (frequency range: 70 Hz-20 kHz, bit rate: 24 bit, sampling rate: 44.1 kHz, total excitation length: 7.5 s, no averaging) ([15]) were first sent to the sound card, a Hammerfall DSP Multiface by RME, were then converted by a D/A-converter of type Behringer ADA8000 Ultragain Pro-8 and were subsequently amplified. Finally, they were played back by Genelec two-way active loudspeakers, model 6010A (frequency range: 73 Hz - 21 kHz (-3 dB)) in the anechoic chamber.
The generated stimuli were presented binaurally via headphones. Open headphones (Sennheiser HD 600) were used for the binaural reproduction. The convolution of stimuli and HRTFs was done off-line with Matlab, and each binaural stimulus was stored as a separate sound file in wave format.

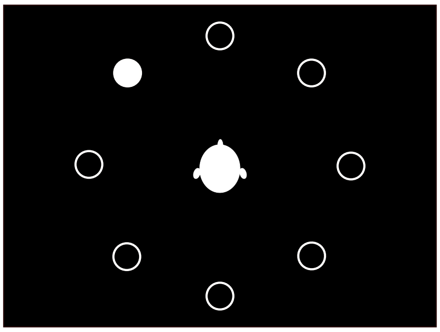

Figure 1: Visual cue with target cued in the direction front-left.

2.4 Stimulus material

Speech material was recorded under anechoic conditions with two male and two female native German speakers. The used hardware, a studio microphone (TLM170 by Neumann) and a sound card (Hammerfall DSP Multiface by RME), allowed recordings with a frequency range from 70 Hz to 20 kHz. The stimuli consisted of single spoken digits (1-9, excluding 5). With a time stretching algorithm that maintains the original frequencies of the recording [14], stimuli were shortened or extended to 730 ms (max. modification of length: 20 %). Therefore, stimuli started and ended synchronously when presented at the same time. The loudness of the recorded stimuli was adjusted according to DIN 45631 [16].

2.5 Experimental Design

The independent within-subject variables were spatial position of target (median vs. diagonal vs. frontal), auditory attention shift of target (repetition vs. shift) and congruency of stimuli (congruent vs. incongruent). Dependent variables were reaction time and error rates.

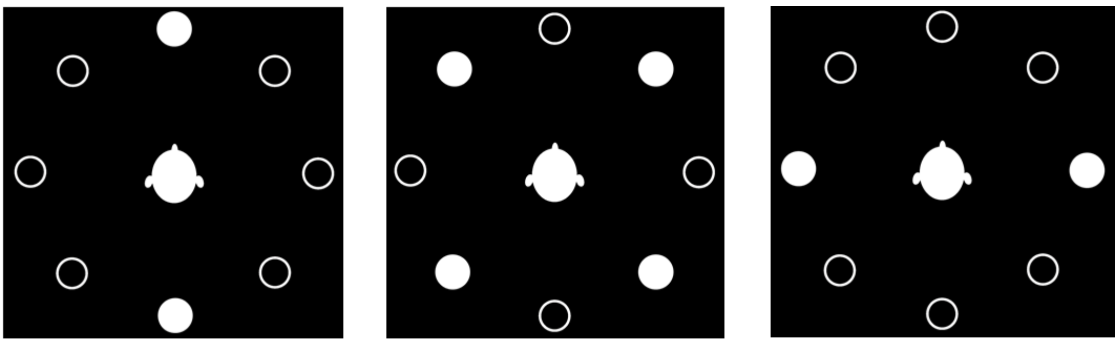

Figure 2: Classes of target's spatial position: Median plane (Front; Back), Diagonal planes (Front-Left; Front-Right; Back-Left; Back-Right), Frontal plane (Left; Right)

2.5.1 Spatial position of the target (CP)

The effect of the target's position was studied. There were eight possible positions for the target (cf. Figure 1). The positions of the target's direction were categorized into three different classes (cf. Figure 2). The classes were designed with respect to the planes within the head-related coordinate system ([17]). The first class included all positions on the median plane (Front; Back) and was therefore called “Median”. The second class described all positions placed on the inter-aural axis and therefore on the frontal plane (Left; Right); it is later called “Frontal”. The third class, called “Diagonal”, includes all other possible spatial positions which are positioned in 45° from the defined planes of the head-related coordinate system.

2.5.2 Auditory attention shift (AS)

Auditory attention shift referred to the target's spatial position in two consecutive trials. The target's spatial position could either be repeated from one trial to another (e.g. front - front) or shifted between trials (e.g. left - back). The distractor's position was shifted between all trials.

2.5.3 Congruency (C)

Congruency referred to the stimuli of target and distractor within one trial. The variable had two different levels (congruent vs. incongruent). A trial was considered congruent when target's digit and distractor's digit belonged to the same category (both digits were smaller than 5 or both greater than 5 (e.g. 2 and 4, 6 and 9)) and the direction word was identical. In case, the digits belonged to different categories (one digit was smaller and one was greater than 5 (e.g. 1 and 7, 8 and 3)) and/or the direction word was not identical the trial was considered as incongruent.

3 Results and discussion

For reaction times, the ANOVA yielded a significant main effect of of spatial position of target [*CP*: $F(2,46)=34.45$, $P<0.0001$] was significant (cf. Figure 3). Reaction times were highest for trials where the target was positioned in the median plane and smallest for trials where the target was positioned in the frontal plane (Median: 1316 ms vs. Diagonal: 1255 ms vs. Frontal: 1145 ms).
The main effect of attention shift was significant and indicated a higher reaction time for shifts than for repetitions (1288 ms vs. 1190 ms) [*AS*: $F(1,23)=20.54$, $P<0.0001$]. The shift costs (difference between reaction times of shift trials and those of repetition trials) amounted on average to 100 ms.
The congruency effect was not significant [*C*: $F(1,23)=2.91$, $P=0.1014$]. A non-significant trend towards higher reaction times for incongruent stimuli than for congruent stimuli (1250 ms vs. 1228 ms) could however been observed.
The attention shift effect interacted with the target's spatial position [$CP\times AS$: $F(2,46)=3.26$,

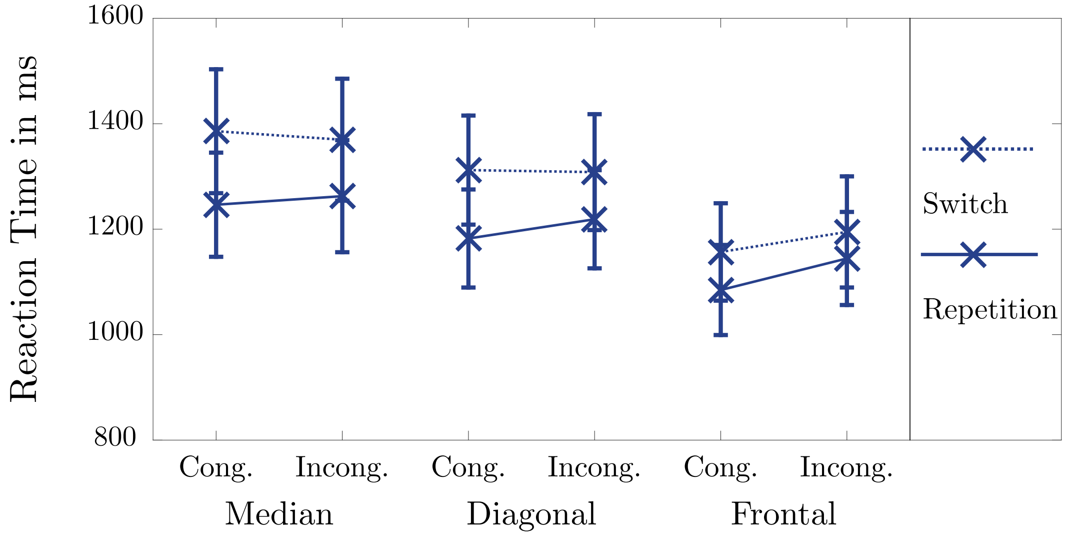

Figure 3: Results of a direct comparison of real sources and binaural synthesis in 3-AFC design with three different stimuli.

P<0.0475], indicating greatest shift costs in Median and smallest reaction time differences in Frontal (Median: 124 ms vs. Diagonal: 109 ms vs. Frontal: 61 ms). The two-way interaction of target's spatial position and congruency was not significant [*CP*×*C*: F(2,46)=2.86, P=0.0673]. The two-way interaction of attention shift and congruency, as well as the three-way interaction of target's spatial position, attention shift and congruency was not significant [*AS*×*C*: F(1,23)=1.75, P=0.1972], [*CP*×*AS*×*C*: F<1].

For error rates, the ANOVA yielded a significant main effect of of target's spatial position [*CP*: F(2,46)=27.13, P<0.0001] was significant (cf. Figure 4). Error rates were significantly smaller (Post-hoc: Bonferoni) for trials where the target was positioned in the frontal plane (Median: 13.9 % vs. Diagonal: 14.7 % vs. Frontal: 7.7 %).
The main effect of attention shift was significant and indicated a higher error rates for shifts than for repetitions (13.1 % vs. 11.1 %) [*AS*: F(1,23)=9.48, P=0.0053]. The shift costs (difference between error rates of shift trials and those of repetition trials) amounted on average to 2 %.
The congruency effect was significant [*C*: F(1,23)=159.64, P<0.0001] indicating that error rates for incongruent stimuli were higher than for congruent stimuli (19.5 % vs. 4.7 %).
The two-way interaction of target's spatial position and attention shift was not significant [*CP*×*AS*: F<1]. The congruency effect interacted with the target's spatial position [*CP*×*AS*: F(2,46)=24.82, P<0.0001], indicating smallest error rate differences in Frontal (Median: 17.9 % vs. Diagonal: 19.2 % vs. Frontal: 7.2 %). The two-way interaction of attention shift and congruency, as well as the three-way interaction of target's spatial position, attention shift and congruency was not significant [*AS*×*C*: F(1,23)=1.83, P=0.1888], [*CP*×*AS*×*C*: F<1].

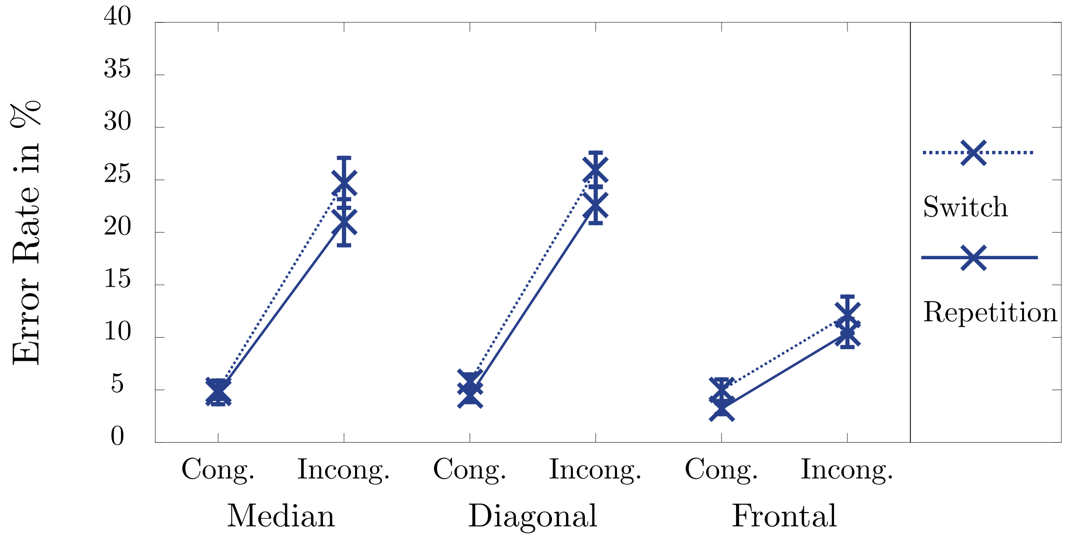

Figure 4: Results of a direct comparison of real sources and binaural synthesis in 3-AFC design with three different stimuli.

4 Discussion

The main aim of this investigation was to test the newly developed binaural paradigm to examine intentional shifting in auditory selective attention in realistic, complex environments. To be able to compare results with results from the former paradigm the scene was presented under anechoic conditions and with static sound sources. In general, reaction times and error rates are higher than those collected with the old paradigm (1238 ms vs 1146 ms; 12.1 ms vs 9.1 %)[9]. The increase in reaction time and error rates is reasonable since the answering task is more demanding (four answering possibilities vs. two answering possibilities).
Significant effects of the target' s position were also found with the former paradigm [18] indicating highest error rates and reaction times in median plane and lowest in frontal plane. Therefore, it can be assumed that the target's position is not influenced by the new paradigm. However, slight differences in main effects and interactions of attention shift and congruency can be found between the two paradigms. The congruency effect turned out to be significant in reaction times with the original paradigm. With the new paradigm a non-significant trend towards greater reaction times in incongruent trials was only found. The authors suggest that the new more demanding task of categorizing the number word and rating the direction word masks the effect of congruency in reaction times. The main effect of congruency was, however, always more pronounced in error rates [9] [6] and is also significant in error rates with the new paradigm. Furthermore, the new paradigm yielded a significant main effect of attention shift in error rates, which was only a non-significant trend with the old paradigm.

5 Conclusion

First results show that the paradigm is robust and provides comparable findings to the original paradigm. Further steps to examine intentional shifting in auditory selective attention are planned in terms of adding reverberant energy and adding moving distracting sources to a dynamic binaural scene.

Acknowledgements The authors are grateful for the provided financing by DFG (Deutsche Forschungsgemeinschaft, FE1168/1-1 and KO2045/11-1).

References

[1] E. Colin Cherry. Some experiments on the recognition of speech, with one and two ears. *Journal of the Acoustical Society of America*, 25:975–979, 1953.

[2] Daniel Holender. Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal. *Behavioral and Brain Sciences*, 9(01):1–23, 1986.

[3] Joel Lachter, Kenneth I. Forster, and Eric Ruthruff. Forty-five years after broadbent (1958): still no identification without attention. *Psychological review*, 111(4):880, 2004.

[4] Barbara G. Shinn-Cunningham. Object-based auditory and visual attention. *Trends in Cognitive Sciences*, 12(5):182–186, 2008.

[5] Iring Koch, Vera Lawo, Janina Fels, and Michael Vorländer. Switching in the cocktail party: Exploring intentional control of auditory selective attention. *Journal of Experimental Psychology / Human Perception and Performance*, 37(4):1140–1147, 2011.

[6] Vera Lawo and Iring Koch. Examining age-related differences in auditory attention control using a task-switching procedure. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 69:237–244, 2014.

[7] Vera Lawo and Iring Koch. Attention and action: The role of response mappings in auditory attention switching. *Journal of Cognitive Psychology*, 27(2):194–206, 2015.

[8] Andrea Kiesel, Marco Steinhauser, Mike Wendt, Michael Falkenstein, Kerstin Jost, Andrea M. Philipp, and Iring Koch. Control and interference in task switching—a review. *Psychological Bulletin*, 136(5):849, 2010.

[9] Josefa Oberem, Vera Lawo, Iring Koch, and Janina Fels. Intentional switching in auditory selective attention: Exploring different binaural reproduction methods in an anechoic chamber. *Acta Acustica united with Acustica*, 100(6):1139–1148, 2014.

[10] Janina Fels, Josefa Oberem, Britta Karnbach, Vera Lawo, and Iring Koch. Comparison of dichotic and binaural reproduction in an experiment on selective auditive attention. In *Proceedings of the International Conference on Acoustics*, 2013.

[11] Josefa Oberem, Vera Lawo, Iring Koch, and Janina Fels. Evaluation of experiments on auditory selective attention in an anechoic environment and a reverberant room with non-individual binaural reproduction. In *Fortschritte der Akustik : DAGA - DGA 2014*, 2014.

[12] Alfred Schmitz. Ein neues digitales kunstkopfmesssystem. *Acta Acustica united with Acustica*, 81(4):416–420, 1995.

[13] Pauli Minnaar, S. Krarup Olesen, Flemming Christensen, and Henrik Møller. Localization with binaural recordings from artificial and human heads. *Journal of Audio Engineering Society*, 49:323–336, 2001.

[14] Institute of Technical Acoustics, RWTH Aachen. Ita toolbox, 2012.

[15] Pjotr Majdak, Peter Balazs, and Bernhard Laback. Multiple exponential sweep method for fast measurement of head-related transfer functions. *Journal of Audio Engineering Society*, 55:623–637, 2007.

[16] German Standard. Calculation of loudness level and loudness from the sound spectrum - zwicker method - amendment 1: Calculation of the loudness of time-variant sound; with cd-rom, 03.2010.

[17] Jens Blauert. *Spatial Hearing - The psychophysics of human sound localization*. MIT Press, Cambridge MA, 2nd edition, 1997.

[18] Josefa Oberem, Iring Koch, and Janina Fels. Intentional switching in auditory selective attention: Exploring age-related effects in a spatial setup. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, page (submitted), 2016.