COVER PAGE

Bibliographical reference

Intentional Switching in Auditory Selective Attention: Exploring Different Binaural Reproduction Methods in an Anechoic Chamber, Josefa Oberem, Vera Lawo, Iring Koch and Janina Fels, Acta Acustica **vol. 100** (Number 6), 2014, pp. 1139-1148

DOI

https://doi.org/10.3813/AAA.918793

Acta Acustica united with Acustica
Vol. 100 (2014) 1139 – 1148

DOI 10.3813/AAA.918793

Intentional Switching in Auditory Selective Attention: Exploring Different Binaural Reproduction Methods in an Anechoic Chamber

Josefa Oberem[1)], Vera Lawo[2)], Iring Koch[2)], Janina Fels[1)]

[1)] Institute of Technical Acoustics, Medical Acoustics Group, RWTH Aachen University, Kopernikusstraße 5, 52074 Aachen, Germany. josefa.oberem@akustik.rwth-aachen.de

[2)] Institute of Psychology, RWTH Aachen University, Jägerstraße 17, 52066 Aachen, Germany

Summary

In a previous study the authors examined intentional switching in auditory selective attention using a dichotic-listening paradigm. In the present study this paradigm was extended to more natural and realistic environments by changing it to a binaural-listening paradigm in which human performance with different methods of spatial reproduction were compared. Four reproduction methods were used: real sources in an anechoic environment, individual binaural synthesis reproduced with headphones, non-individual binaural synthesis reproduced with headphones, and non-individual binaural synthesis reproduced with two loudspeakers and Cross-Talk-Cancellation-Filters. Speech of two speakers was presented simultaneously to subjects from two out of eight different directions. Guided by a visual cue, subjects were asked to categorize the target's speech while ignoring the distractor's speech. Results showed greater reaction times and error rates for non-individual repro duction methods. The influences of the spatial transition of the target-speaker (switch or repetition of speaker's direction in space) and of the spatial arrangement of the two speakers were largely identical across reproduction methods, even though it was generally easier to filter out distractor's speech when using real sources. The findings suggest that the reproduction methods can be usefully applied to study auditory attention with only very little loss in accuracy.

PACS no. 43.66.Pn, 43.66.Qp, 43.66.Dc, 43.66.Lj

1. Introduction

Communicative noisy situations such as cocktail-parties have been in the focus of research since Cherry [1] reported his study using dichotic-listening to investigate auditory attention. In the present investigation "dichotic" is used as in cognitive sciences referring to two different stimuli presented separately to the two ears. Subjects were asked to selectively listen to continuous speech presented to one ear and repeat the words of this speaker while ignoring the speech of a distracting speaker on the other ear. A strong attentional selection is needed to fulfill the required task of shadowing (i.e. repeat aloud) the relevant acoustic information [2]. Several questions regarding auditory attention have been asked in the last decades [3] and there has been a long tradition of using dichotic-listening paradigms in the study of auditory selective attention [1, 4, 5, 6, 7].

Recently Koch *et al.* [8] used dichotic listening to examine intentional attention switching using spoken digits/number words as auditory stimuli. For that investigation, dichotic listening was combined with the methodology of task cueing. In task cueing, the sequence of tasks is unpredictable and an instructional cue is needed to indicate the next task. More specifically, Koch and colleagues used a visual selection cue, preceding the auditory stimuli, to indicate either the gender of the relevant speaker in the upcoming trial (i.e. when one stimulus was spoken by a female speaker and the other stimulus by a male speaker) [8, 9, 10] or the relevant ear (see Lawo *et al.* [11] for a direct comparison of gender-based and ear-based selection criteria). The participants' task was always to categorize the relevant digit as smaller or larger than five and press the corresponding response button. The main finding of these investigations was that a cued switch of the relevant target (i.e. the target's position switched 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. The corresponding differences in reaction times and error rates were also called switch costs [12].

In simple experimental setups as presented by Koch *et al.* [8] where two sources are presented to the right and left ear, the dichotic reproduction of stimuli is convenient [5, 7]. In general, however, dichotic listening is a highly artificial situation compared to natural listening. A binaural reproduction of stimuli is more natural and offers several

Received 29 November 2013,
accepted 18 August 2014.

advantages. In the present investigation "binaural" does not only refer to the situation where sound reaches both ears, but it also includes spatial information.

With a binaural reproduction, there are more degrees of freedom for the location of sources (in contrast to the limitation to left and right in dichotic listening, in binaural listening, sources can be positioned at any location on a sphere around the listener) and therefore the distance between sources as well as the distance of sources to the listener are more variable [13, 14, 15, 16, 17]. The greater range of source positions in a binaural listening scene also offers more possibilities for the number of maskers. Questions of whether a spatial separation of target and distractor improves attention performance [16, 18] or, whether attention acts like a "spotlight" [19], can only be analyzed and answered with a binaural experimental setup. For example, Bregman [20] and Deutsch [21] reported a benefit of binaural listening emphasizing the ability to switch voluntarily between multiple channels or streams of information. Regarding the environment of target and masking speakers, binaural technologies also allow the inclusion of room acoustics such as variable reverberation times [22].

With respect to the listed advantages, the aim of this investigation was to extend and transform the dichotic-listening paradigm, designed to analyze intentional switching in auditory selective attention by Koch *et al.* [8], into a binaural-listening paradigm. With this extension, the authors took a step towards natural listening in realistic scenes.

A binaural scene can be reproduced by different binaural reproduction methods. For example, the methods can be based on individual or non-individual head-related transfer functions (HRTFs). Furthermore, the stimuli can be presented over headphones or loudspeakers. For the listener. These binaural reproduction methods can differ in accuracy of the binaural synthesis.

Evaluations of binaural reproduction methods were usually performed with localization experiments. Comparisons of localization performance between real sources and individual binaural synthesis presented with headphones were analyzed and rated as similar by Bronkhorst [23]. Wightman and Kistler [24] found similar results, but they also reported about challenges in elevated positions for the individual binaural synthesis which became apparent through an increased angle of error. The results of comparisons between individual and non-individual binaural recordings were analyzed by several authors [25, 26, 27, 28]. All of them showed that individual recordings yielded better results than non-individual recordings for localizing sources in space. Detailed results also showed that in localization tasks non-individual binaural stimuli especially caused difficulties for sources located in the median plane, on cones of confusion, as well as elevated directions.

In real-life scenes, subjects are usually asked to process much more complex information than in simple localization tasks. Hence, this investigation tried to find a new measure to define the required accuracy of binaural syntheses in an "everyday-task", including localization but, with a main focus on a non-localizing task. Therefore, the main task was to analyze intentional switching in auditory selective attention. A listening test with the binaural-listening paradigm [29] was carried out and compared with different binaural reproduction methods (loudspeakers in an anechoic environment, binaural synthesis with individual HRTFs via headphones, binaural synthesis with non-individual HRTFs via headphones and binaural synthesis with non-individual HRTFs via two loudspeakers and a Cross-Talk-Cancellation-Filter (CTC)). It seems reasonable to assume that similar results to those of the investigations focusing only on localization such as worse performance with non-individual binaural stimuli will be observed with the present paradigm. However, it is not clear whether the task of the present paradigm regarding auditory selective attention will be as much effected as a simple localization task by the individuality of the binaural presentation.

2. Methods

The performance of intentional switching in auditory selective attention was evaluated in four binaural listening conditions:

- I. Real sources
- II. Individual binaural stimuli via headphones
- III. Non-individual binaural stimuli via headphones
- IV. Non-individual binaural stimuli via CTC.

The experimental procedure (c.f. section 2.5) was the same for all reproduction methods.

2.1. Subjects

A number of 96 ($4 \cdot 24$, between-subject-design) paid (8 euros) students aged between 18 and 35 (mean age: 24.5 years) participated in the experiment and were randomly assigned to the four reproduction methods. 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 10 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 and the required measurements took place in a fully anechoic chamber ($l \times w \times h = 9.2 \times 6.2 \times 5.0\,\mathrm{m}^3$) with a lower boundary frequency limit of 200 Hz. The subjects were asked to sit inside a frame of eight loudspeakers (cf. Figure 1, 2), which were equally distributed over azimuth (every 45°), whereas the distance between the subject and the loudspeakers was kept constant at 1.8 m. The chair was provided with a backrest, armrests, and an adjustable head rest. An electromagnetic tracker (Polhemus Patriot) was used during the HRTF measurements and the listening test to control and supervise the movements of the subject's head. Therefore, data associated with head movements was eliminated a posteriori. Based on a series of pretests, limits for the allowed

head movements were set to ±1 cm in translation and ±2° in rotation. The head tracker was not used to adjust binaural stimuli; the binaural presentation via headphones and CTC was static. To take the focus from vision to audition, lights were turned off during the listening test [30, 31].

2.3. Reproduction Methods

2.3.1. Reproduction method I: Real sources

In reproduction method I, the scene of competing speakers was reproduced by loudspeakers placed in the anechoic chamber (cf. Figure 1, Figure 2). The used loudspeakers were Genelec two-way active loudspeakers, model 6010A (frequency range: 73 Hz - 21 kHz (-3 dB)) fed by the sound card, a Hammerfall DSP Multiface by RME.

2.3.2. Reproduction method II: Individual binaural stimuli via headphones

In reproduction method II, individually generated stimuli were presented binaurally via headphones. Therefore, HRTFs were measured individually.

Measurements ran automatically with the ITA-Toolbox [32] in Matlab. Interleaved exponential sweeps [33, 34] (frequency range: 70 Hz-20 kHz, bit rate: 24 bit, sampling rate: 44.1 kHz, total excitation length: 7.5 s, no averaging) were first sent to the sound card, then converted by an D/A-converter of type Behringer ADA8000 Ultragain Pro-8 and amplified, and finally played by the loudspeakers in the anechoic chamber. For recording, microphones (KE3 by Sennheiser) were placed at the entrance of the ear canal with an open-dome, a little silicon carrier, so that the ear canal stayed partly open. The in-ear recorded signal went through the above-mentioned A/D-converter and the sound card before being post-processed (including time windowing).

Open headphones (Sennheiser HD 600) were used for the binaural reproduction. As shown by Masiero and Fels [35], headphone equalizations are also necessary when the headphones are open. A robust headphone equalization is especially important when headphones are taken off or repositioned on the head during the experiment. Hence, a number of eight headphone-transfer-functions (HpTFs) were measured with the described hardware, microphones, and exponential sweeps. After every measurement, the subject was asked to reposition the headphones in a comfortable position. Measurements were averaged and a minimum-phase-filter was applied [35].

The convolution of stimuli, HRTF (filter length: 40 ms), and equalization (filter length: 23 ms) was done off-line with Matlab, and each binaural stimulus was stored as a separate sound file in wave format.

Since only HRTFs from defined directions were measured, the presentation of the binaural stimuli was static.

2.3.3. Reproduction method III: Non-individual binaural stimuli via headphones

Reproduction method III was a generalization of reproduction method II. Instead of individually measured HRTFs, the HRTFs of an artificial head were used to create binaural stimuli. The dummy head is a mannequin produced at the Institute of Technical Acoustics, RWTH Aachen University, with a simple torso and a detailed ear geometry [36, 37].

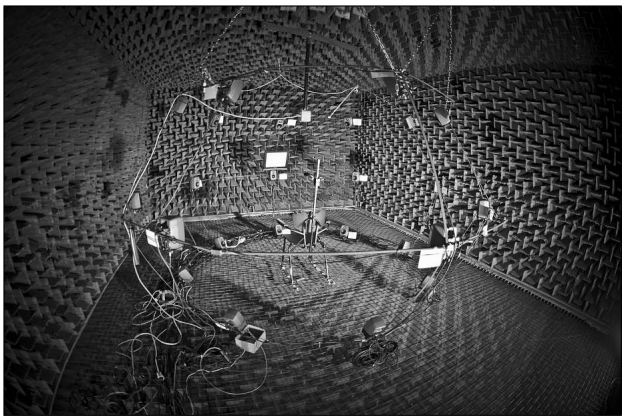

Figure 1. Anechoic room with loudspeaker setup and monitor in front.

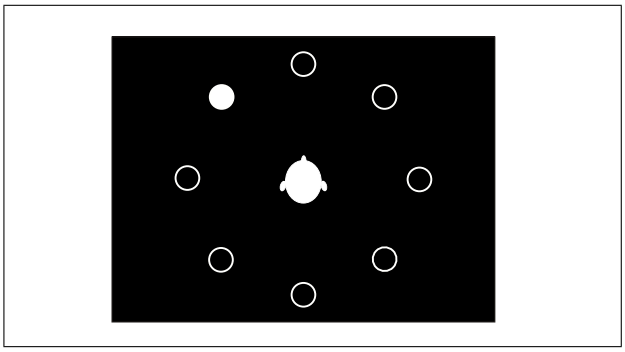

Figure 2. Visual Cue with target cued in the direction front-left.

2.3.4. Reproduction method IV: Non-individual binaural stimuli via CTC

Reproduction method IV was based on the procedure of Cross-Talk-Cancellation (CTC). First introduced by Atal and Schröder [38], CTC makes it possible to present binaural stimuli via loudspeakers. Detailed information about the theory and procedure can be found in Møller [39], Schmitz [40], and Lentz [41].

In this investigation a third order CTC filter was used and stimuli were presented with two loudspeakers in the horizontal plane at ±45° (front-left and front-right). Further information about the CTC-filter were presented by Majdak *et al.* [42]. The used HRTFs were those of the artificial head described in reproduction method III. Since the subject's movements were restricted to ±1 cm in translation and ±2° in rotation, the subject was always within the sweet spot [41].

2.4. Stimulus material

Speech material was recorded under anechoic conditions with two male and two female native German speakers. The used hardware, studio microphone TLM170 by

Neumann and 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 [32], 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 [43].

2.5. Experimental Procedure

The paradigm was firstly introduced by Koch *et al.* [8] and developed to analyze the intentional switching in auditory selective attention using dichotic listening. It consists of two simultaneously presented stimuli. These stimuli were delivered by two speakers of opposite sex. In the present binaural-listening paradigm, the speakers were located in two different directions (out of eight possible, cf. Figure 2).

One speaker acted as the target and the other acted as the distractor. The participant was asked to focus on the target-speaker and ignore the distracting speaker. To distinguish between target and distractor, the target-speaker's direction was cued in advance. Hence, a visual cue highlighting the target's direction was shown on a monitor (15 inch screen, 1.8 m distance). The visual cue consisted of a sketch of all directions indicating the target direction with a filled dot (cf. Figure 2).

The listener's task was to categorize the target's speech into smaller vs. greater than five. The speech material does not include the digit five. The two stimulus categories are mapped to two response buttons, held in hands, to be pressed by the left and right thumb.

Figure 3 shows the procedure of a trial. Each trial started with a visual cue presented on the monitor in front of the subject. After a cue-stimulus interval (CSI) of 500 ms, the two acoustic stimuli (target and distractor) were simultaneously presented. The visual cue remained on the screen until the subject responded to the acoustic target. The interval between response and next cue (RCI) was also set to 500 ms. In case of an error, visual feedback ("Fehler!", German for "error") was displayed for 500 ms, delaying the onset of the next cue.

In total 600 trials divided into four blocks of 150 trials each were separated by short breaks (5 min). The experimental blocks were preceded by one training block of 50 trials. The total duration of the experiment did not exceed 60 min including the audiometry.

In the blocks the location of the speakers was repeated or changed. Furthermore, trials were counterbalanced over combinations of digits. The speakers of the digit were assigned randomly.

2.6. Experimental Design

The experiment was designed to examine intentional switching in auditory selective attention and to compare

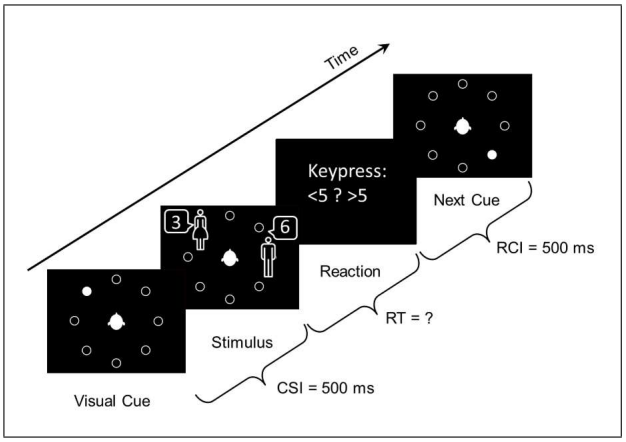

Figure 3. Procedure of a trial with a visual cue indicating the target direction, a cue-stimulus-interval (CSI) of 500 ms, the synchronous presentation of the stimuli, reaction time between onset of stimulus and the response of the subject, and the response-cue-interval (RCI) of 500 ms.

the performance across the four used reproduction methods (between-subject variable). Due to the extension to a binaural and therefore spatial reproduction, new acoustical aspects were also taken into account (cf. section 2.6.2). Reaction time (RT) and error rate were the dependent variables.

2.6.1. Analysis A: Transition and Congruency

The two independent variables were transition of target (repetition vs. switch) and congruency of stimuli (congruent vs. incongruent).

Transition 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 switched between trials (e.g. left - back). It should be mentioned that the distractors position was not relevant and therefore, not included in the transition analysis.

Congruency referred to the stimuli of target and distractor within one trial. The variable had two different levels (congruent vs. incongruent). The two stimuli could be congruent, which was the case when both digits were smaller than 5 or both greater than 5 (e.g. 2 and 4, 6 and 9), or they could be incongruent, which was the case when one digit was smaller and one greater than 5 (e.g. 1 and 7, 8 and 3).

2.6.2. Analysis B: Spatial combination of target and distractor

In a second analysis, the effect of different combinations of target's and distractor's location was studied. With eight different directions for target and distractor 56 different combinations were possible (target and distractor were never located in the same spot). For a manageable analysis, the combinations of target's direction and distractor's direction were categorized into five different classes (c.f. Figure 4). Classes were designed with respect to psychoacoustical and binaural criteria (ITD, ILD, Cones of Confusion, etc.). The first class included combinations where

the distractor was positioned as a mirror-image of the target relative to the median plane and was therefore called "Left–Right" (later also referred to as "L-R"). For the second class, the mirror was placed on the inter-aural axis. Target and distractor of this class were always on the same "Cone of Confusion", which gave the class its name (for abbreviated terms later also ref erred to as "Cone"). If target and distractor were "Next Neighbors", they belonged to the third class (short "Next"). All other combinations (with no specific binaural criterion) were divided into class four and five. The fourth class covered twelve combinations with angles of 90° or 180° between target and distractor that were not included in the first two classes. This class was therefore referred to as "90/180". All combinations with an angle difference of 135° were gathered in the fifth class (referred to as "135").

All trials were balanced over these five classes and combinations within one class were random.

3. Results

For the analysis of reaction times and error rates, the training sequence was removed from the data. The first trial in every block, trials where head movements were detected by the tracker as well as every trial with a reaction time exceeding ±3 standard deviations from the individual's mean reaction time were also excluded from the analysis. Additionally for the analysis of reaction times, every trial with an error and the following trial were eliminated, since these trials could not validly be defined as switch or repeat trials.

The figures presenting results (reaction times or error rates) show the mean and the standard error across participants.

3.1. Analysis A: Transition and Congruency

The reaction time and the error rates were submitted to two separate 3-way mixed analysis of variances (ANOVAs) with the variables of reproduction method (R, between subjects), transition (T), and congruency (C).

3.1.1. Reaction time

For reaction times, the ANOVA yielded a significant main effect of reproduction method [R: $F(3, 92) = 6.42$, $MSE = 2298,506$, $p < 0.05$, $\eta_p^2 = 0.173$]. A post-hoc t-test (LSD) was performed (cf. Figure 5). Reproduction method I with real sources was significantly different from the reproduction methods III and IV indicating a higher reaction time for the methods with non-individual HRTFs (III and IV) which did not significantly differ from each other. The individual binaural reproduction method II was significantly different from reproduction method IV. Furthermore, the results obtained with method II unfolded only a marginally significant difference from those of method III [$p = 0.061$]. However, method II did not lead to results significantly different from those obtained with the reproduction method I. Overall, reaction times decreased with the individuality of the reproduction method.

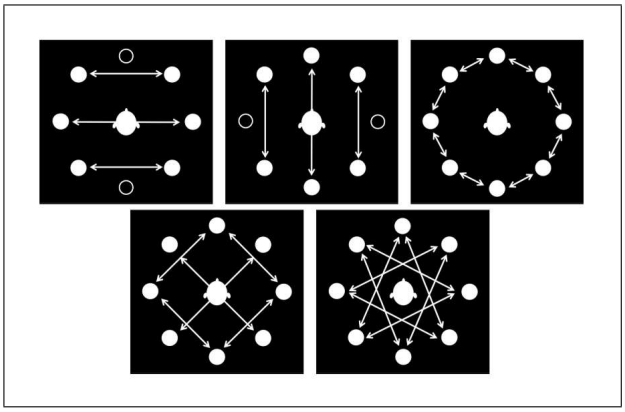

Figure 4. Groups of target and distractor combinations: "L-R", "Cone", "Next", "90/180" and "135"

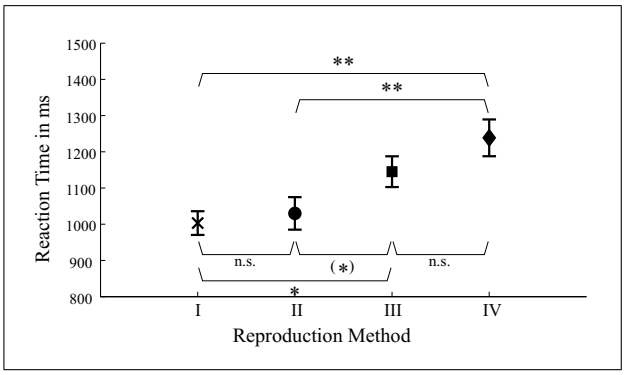

Figure 5. Reaction time (in ms) as a function of reproduction method. Error bars indicate standard errors. Marginal significance: $^{(*)} \mathrel{\widehat{=}} p = 0.061$. ANOVA: n.s. $\mathrel{\widehat{=}} p > 0.05$, $* \mathrel{\widehat{=}} p < 0.05$, $** \mathrel{\widehat{=}} p < 0.001$. I. $\mathrel{\widehat{=}}$ Real Sources, II. $\mathrel{\widehat{=}}$ Ind. stimuli via headphones, III. $\mathrel{\widehat{=}}$ Non-ind. stimuli via headphones, IV. $\mathrel{\widehat{=}}$ Non-ind. stimuli via CTC

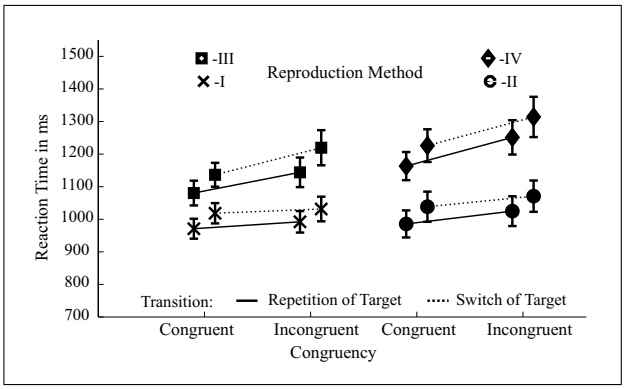

Figure 6. Reaction time (in ms) as a function of reproduction method, transition and congruency ($R \times T \times C$). Error bars indicate standard errors. I. $\mathrel{\widehat{=}}$ Real Sources, II. $\mathrel{\widehat{=}}$ Ind. stimuli via headphones, III. $\mathrel{\widehat{=}}$ Non-ind. stimuli via headphones, IV. $\mathrel{\widehat{=}}$ Non-ind. stimuli via CTC

The main effect of transition (T) on reaction time was significant and indicated a higher reaction time for switches than for repetitions (c.f. Figure 6) [T: $F(1, 92) = 96.94$, $MSE = 39,401$, $p < 0.001$, $\eta_p^2 = 0.513$]. The switch costs (difference between reaction times of switch

trials and those of repetition trials) amounted on average to 55 ± 10 ms.

The ANOVA also yielded a significant main effect of congruency (C), indicating higher reaction times for incongruent stimuli than for congruent stimuli [C: $F(1, 92) = 39.35$, $MSE = 91,129$, $p < 0.001$, $\eta_p^2 = 0.300$].

The ANOVA yielded no significant interaction of reproduction method and transition [$R{\times}T$: $F < 1$] (c.f. Figure 6). However, the ANOVA yielded a significant interaction of reproduction method and congruency ($R{\times}C$), indicating a difference between reproduction method I with real sources and all other reproduction methods [$R{\times}C$: $F(3, 92) = 3.68$, $MSE = 91,129$, $p < 0.05$, $\eta_p^2 = 0.107$]. The congruency effect was not significant with real sources (I: 17 ms) in contrast to the other three reproduction methods (II: 36 ms, III: 73 ms, IV: 88 ms). The interaction of transition and congruency ($T{\times}C$) was not significant [$T{\times}C$: $F < 1$]. Finally, the ANOVA yielded no significant interaction of reproduction method, transition, and congruency [$R{\times}T{\times}C$: $F < 1$].

3.1.2. Error rate

Error rates increased from reproduction method I to IV. The ANOVA yielded a significant main effect of the reproduction method (R) for error rates, indicating lower error rates for reproduction method I than for all reproduction methods with binaural synthesis, confirmed by a post-hoc t-test (LSD) (cf. Figure 7) [R: $F(1, 92) = 4.41$, $MSE = 0.000$, $p < 0.05$, $\eta_p^2 = 0.046$]. Differences between reproduction method II, III, and IV were not significant, except for the difference between individual binaural stimuli presented via headphones (method III) and the reproduction of non-individual binaural stimuli with CTC (method IV).

The main effect of transition (T) was not significant (c.f. Figure 8), but the ANOVA yielded a significant main effect of congruency (C), indicating higher error rates for incongruent stimuli than for congruent stimuli [T: $F(1, 92) = 1.60$, $MSE = 0.000$, $p > 0.05$, $\eta_p^2 = 0.017$], [C: $F(1, 92) = 438.13$, $MSE = 0.000$, $p < 0.001$, $\eta_p^2 = 0.826$].

The ANOVA yielded no significant interaction of reproduction method and transition ($R{\times}T$) for error rates $F < 1$ (c.f. Figure 8), but a significant interaction reproduction method and congruency ($R{\times}C$) could be observed, indicating a smaller effect of congruency for reproduction method I with real sources than for all other reproduction methods [$R{\times}T$: $F < 1$], [$R{\times}C$: $F(3, 92) = 16.87$, $MSE = 0.000$, $p < 0.001$, $\eta_p^2 = 0.355$]. The congruency effect was significantly smaller with real sources (I: 4 %) relative to the other three reproduction methods (II: 10 %, III: 12 %, IV: 13 %). The interaction of transition and congruency ($T{\times}C$) was significant, indicating generally larger switch costs on incongruent trials, but this effect did not differ significantly as a function of reproduction method ($R{\times}T{\times}C$); [$T{\times}C$: $F(3, 92) = 16.87$, $MSE = 0.000$, $p < 0.001$, $\eta_p^2 = 0.355$], [$R{\times}T{\times}C$: $F(3, 92) = 1.64$, $p > 0.05$, $\eta_p^2 = 0.051$].

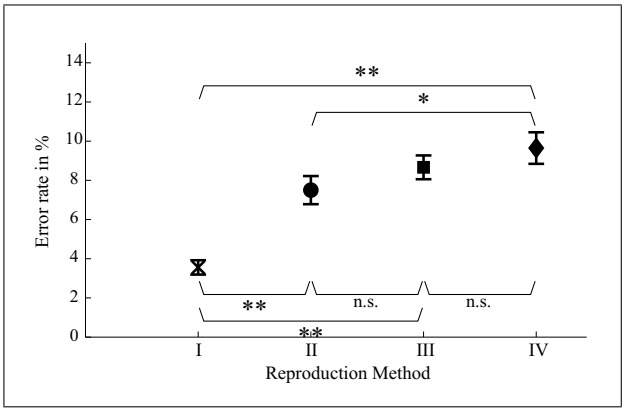

Figure 7. Error rate (in %) as a function of reproduction method. Error bars indicate standard errors. ANOVA: n.s. ≙ $p > 0.05$, * ≙ $p < 0.05$, ** ≙ $p < 0.001$. I. ≙ Real Sources, II. ≙ Ind. stimuli via headphones, III. ≙ Non-ind. stimuli via headphones, IV. ≙ Non-ind. stimuli via CTC

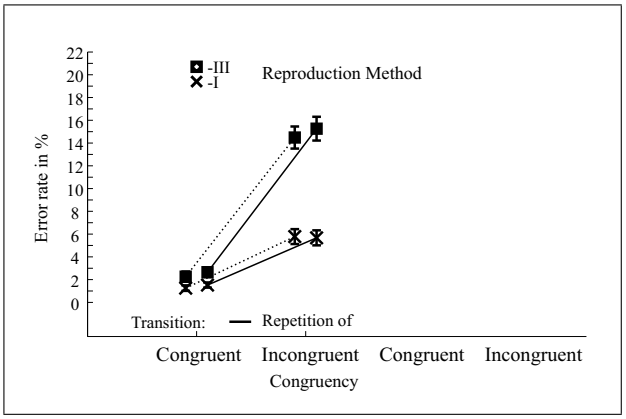

Figure 8. Error rate (in %) as a function of reproduction method, transition and congruency. Error bars indicate standard errors. I. ≙ Real Sources, II. ≙ Ind. stimuli via headphones, III. ≙ Non-ind. stimuli via headphones, IV. ≙ Non-ind. stimuli via CTC

3.2. Analysis B: Spatial combination of target and distractor

As shown in section 3.1, absolute reaction times and error rates obtained with the four reproduction methods differed in some cases. In this part of the analysis, the differences between categorized classes within one reproduction method were of interest. For a better comparison of the relative effects between reproduction methods, however, reaction times and error rates were referred to those of the class "Left–Right" (c.f. Figure 4). Figures 9 and 10 display the differences between a class and the values of the corresponding "Left–Right" class in percent.

For all reproduction methods, "Left–Right"-combinations show the smallest reaction times and error rates since all deviant reaction times and error rates were positive (c.f. Figure 9 and Figure 10).

3.2.1. Reaction time

The ANOVA yielded no significant interaction between the effects of reproduction method and spatial combina-

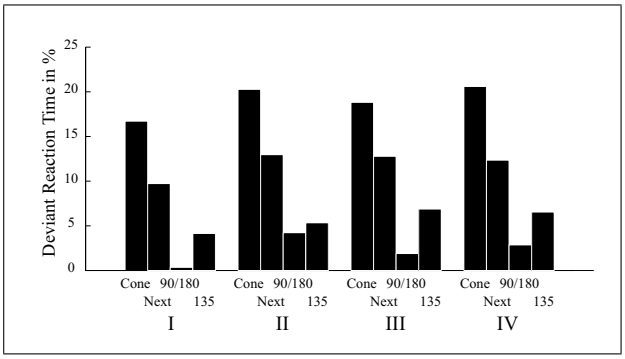

Figure 9. Deviant reaction time (in %) as a function of reproduction method and grouped combinations referenced to reaction times of class 'Left–Right'. I. ≙ Real Sources, II. ≙ Ind. stimuli via headphones, III. ≙ Non-ind. stimuli via headphones, IV. ≙ Non-ind. stimuli via CTC.

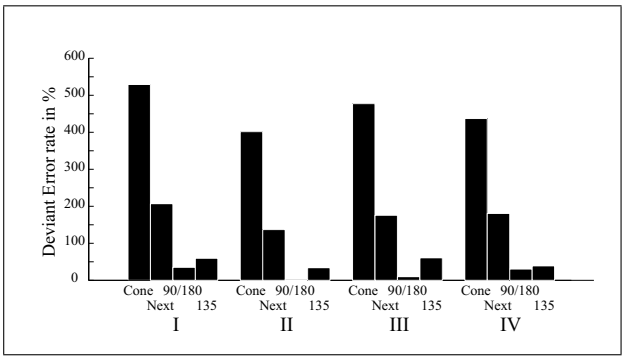

Figure 10. Deviant error rate (in %) as a function of reproduction method and grouped direction referenced to error rate of class 'Left–Right'. I. ≙ Real Sources, II. ≙ Ind. stimuli via headphones, III. ≙ Non-ind. stimuli via headphones, IV. ≙ Non-ind. stimuli via CTC

tions for reaction times (c.f. Figure 9) [$F < 1$]. Conform with binaural and psycho-acoustical criteria, highest reaction times could be observed for trials where target and distractor were located on the same cone of confusion. Post-hoc tests showed that the reaction times differed significantly from those of all other classes. Besides the combinations within a cone of confusion, trials with adjacent sources showed significantly higher reaction times than those of classes "Left–Right", "90/180", and "135". Apart from the reproduction method of individual binaural presentation via headphones (reproduction method II), classes "Left–Right" and "90/180" did not differ significantly in reaction times. Combinations close to front-back combinations (combinations in class "135" could be split in four combinations comparable to left-right-combinations (e.g. 90° and −135°) and four combinations comparable to front-back combinations (e.g. 0° and 135°)) increased the reaction times in class "135" compared to those of classes "Left–Right" and "90/180".

3.2.2. Error rate

The distribution of error rates was comparable to those of the reaction times. The greatest error rates could be found in trials belonging to the class "Cone" and the smallest error rates for the class "Left–Right". Unlike the ratio between reaction times, which were in a range of $0 - 23\%$, error rates varied to larger scales ($0 - 550\%$). Due to differences in error rates of the class "Cone" between reproduction methods, the ANOVA yielded a significant interaction of reproduction method and grouped combinations for error rates (c.f. Figure 10) [$F(4, 89) = 11.90$, $MSE = 000$, $p < 0.001$, $\eta_p^2 = 0.280$]. Separate ANOVAs for every reproduction method showed significant differences between the class "Cone" and all other classes as well as between the class "Next" and all other classes.

4. Discussion

The main aim of this investigation was to compare different binaural reproduction methods by means of listening tests with a paradigm focusing on intentional switching in auditory selective attention. There was statistical evidence that absolute values of reaction times and error rates differed between reproduction methods (c.f. Section 3.1) in this investigation. By contrast with other investigations that compared reproduction methods in localization experiments, similarities and differences could be found.

As expected, reaction times for reproduction method I did not differ significantly from reaction times for reproduction method II, since the HRTFs were individual. However, a significant difference could be found in error rates. The difference between reproduction method I and II was the static presentation of the binaural synthesis in reproduction method II. In both reproduction methods subjects were able to perform small head movements (sounding) within the permitted area defined by the tracker. While subjects listening to real sources got a feedback in terms of changes in interaural level difference (ILD) and interaural time difference (ITD) from the movements of sounding, subjects listening to the binaural synthesis missed this additional localization information. The static presentation of individual binaural stimuli did not offer the additional localization information of head movements and therefore, it could be assumed that error rates were increased at least partly due to the lack of this advantage.

As shown in several investigations [25, 26, 28], localization suffers from non-individual binaural stimuli compared to stimulus material based on individual HRTFs. In this investigation, a marginally significant difference between reproduction method II and III could be found in reaction times and no significant difference in error rates. However, a tendency of worse results for non-individual reproduction could be observed.

Reproduction method IV achieved the results with highest reaction times and error rates. It seems reasonable that reaction times and error rates were comparable to or worse than results for reproduction method III because the applied HRTFs were identical across all subjects in these reproduction methods. In CTC evaluations [44, 45, 46, 47] concerning localization, limited sweet spots raised a challenge and affected performance. Since headmovements were supervised limitation due to the sweet spot did not severely affect the results of this reproduction method in

the present study. Expectations were met with highest reaction times and error rates in reproduction method IV.

It can be summarized that absolute values of reaction times and error rates increase, partly significantly, with the individuality of the reproduction method. However, in most psychological studies the absolute values are not of major importance, but subsequently discussed effects and interactions of variables are relevant.

Another important aim of the present investigation was to extend the dichotic-listening paradigm [8] to a more realistic binaural-listening paradigm. Results of the present investigation were therefore compared with results collected with the dichotic-listening paradigm by Koch *et al.* [8, 9, 10, 11].

The significant effect of transition, indicating that subjects responded more slowly when the target's direction was switched, could be observed with all reproduction methods as well as in the investigations of Lawo *et al.* [11] using dichotic listening. The switch cost provided an explicit measure of how well instructions to switch attention could be followed. In general, larger switch costs were found in dichotic-listening-experiments compared to the present investigation ($\sim$100 ms vs. 55 ms). A switch between only two possible directions (i.e. dichotic listening) was expected to be easier to detect than a switch to one of eight possible directions equally distributed on the horizontal plane. The angular distance between the target's positions could have been a reason for different complexity of switches. Besides the angular distance of target's positions, the visual cue (in ear-based dichotic experiments the visual cue was a letter (L/R) and therefore differed from the cue design of this investigation) as well as the reproduction method (binaural vs. dichotic) might have had an effect on the switch costs.

While the significant effect of congruency was less apparent for reaction times, the difference between congruent and incongruent trials was more pronounced for error rates. These findings were also confirmed by the previous dichotic investigations [8, 9, 10, 11]. The congruence effect could be taken as an implicit performance measure of attending to task-irrelevant information and filtering out the irrelevant information [8]. Results of the 3-way-ANOVA showed significant differences in the congruence between the reproduction method of real sources and the other tested methods. Thus, the distractor's information was less effectively ignored when the reproduction method was based on binaural synthesis relative to real sources. The loss of additional information in localization of small head movements due to a static reproduction could have been a reason for this effect.

In summary, the effects of transition of the speaker's location and the congruency of the stimuli showed the same patterns for binaural- and dichotic-listening. However, they were differently pronounced due to a more complex arrangement of possible speaker's positions in the binaural-listening paradigm. With the more complex binaural-listening paradigm further effects such as the spatial combination of target and distractor's location as presented in section 3.2 could be analyzed. Differences between the grouped combinations could be found showing greater reaction times and error rates for combinations of the classes "Cone of Confusion" and "Next Neighbors" compared to the other three classes.

In localization experiments with real sources, individual and non-individual reproduction via headphones by Møller *et al.* [28], errors accumulated in "Within Cone" and "Median Plane" conditions. Furthermore, it was observed that especially in the non-individual reproduction the percentage of errors in "Median Plane" conditions increased. In this investigation, highest error rates and highest reaction times also occurred in trials with target and distractor located on a Cone of Confusion, but there was no increased error for non-individual reproduction methods in this category. Overall, there was a higher rate of errors for reproduction methods based on binaural synthesis, but errors were equally spread over all categorized combinations.

The effect of spatial separation of sources in experiments focusing on selective attention was studied by Best *et al.* [19]. It was shown that auditory selective attention was exposed to a greater challenge when sources were not or only little spatially separated. These results were only based on error rates since Best at al.'s paradigm did not allow the measurement of reaction times. In the present study, comparatively large error rates and reaction times could be found in the combination class of "Next Neighbors". These findings confirmed Best at al.'s findings.

5. Conclusions

This investigation showed that the extension from a dichotic-listening paradigm to a binaural-listening paradigm was successful and offers more possibilities to analyze intentional switching in auditory attention. The comparison of reproduction methods showed that differences between absolute values of reaction time and error rates should not be neglected, but, in experiments where effects and interactions of different variables were to be compared, all reproduction methods yielded nearly the same results. Results of this study prove that findings with the used paradigm are comparable for all four reproduction methods, but similar results can be expected for other paradigms also focusing on auditory selective attention.

In terms of applicability, the binaural synthesis based on non-individual HRTFs reproduced via headphones (III) is the most convenient reproduction method, which is especially true for the use in non-acoustically equipped psychological facilities since no anechoic chamber is needed during the experiment. However, a loss of reality due to non-individual binaural stimuli must be accepted.

From a technical, acoustical point of view, the assumption that head movements could be the reason for the decrease in performance regarding error rates in binaural synthesis as well as the significant differences in the effect of congruency needs to be analyzed. Therefore, a dynamic binaural reproduction is needed. With a successful integration of head-movements in the experiment, reproduction

methods I and II should not lead to differences in the effect of congruency nor in the absolute reaction times and error rates. Under this assumption, an individual binaural reproduction of stimuli is advisable for future investigations for investigators who have the possibility to measure individual HRTFs in an anechoic chamber.

Since the general aim of this investigation was to analyze intentional switching in auditory selective attention in more realistic environments, further steps could be an extension of the binaural-listening paradigm towards room acoustics and multiple distracting sources.

Acknowledgments

The authors are grateful for the provided financing by DFG (Deutsche Forschungsgemeinschaft, FE1168/ 1-1 and KO2045/ 11-1). Albina Kamil, Marcia Lins, Anne Stockmann and Suliang Wang assisted with data collection. Two anonymous reviewers gave important feedback that greatly strengthened the paper.

References

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

[2] B. G. Shinn-Cunningham, V. Best: Selective attention in normal and impaired hearing. Trends in Amplification **12** (2008) 283–299.

[3] A. W. Bronkhorst: The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. Acta Acustica united with Acustica **86** (2000) 117–128.

[4] D. E. Broadbent: Perception and communication. Pergamon, Oxford, 1958.

[5] H. E. Pashler: The psychology of attention. MIT Press, 1999.

[6] A. Ihlefeld, B. G. Shinn-Cunningham: Spatial release from energetic and informational masking in a selective speech identification task. The Journal of the Acoustical Society of America **123** (2008) 4369–4379.

[7] K. Hugdahl: Fifty years of dichotic-listening research - still going and going and. . .. Brain and Cognition **76** (2011) 211–213.

[8] I. Koch, V. Lawo, J. Fels, M. Vorländer: Switching in the cocktail party: Exploring intentional control of auditory selective attention. Journal of Experimental Psychology / Human Perception and Performance **37** (2011) 1140–1147.

[9] V. Lawo, I. 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** (2014) 237–244.

[10] I. Koch, V. Lawo: Exploring temporal dissipation of attention settings in auditory task switching. Attention, Perception, & Psychophysics **76** (2013) 73–80.

[11] V. Lawo, J. Fels, J. Oberem, I. Koch: Intentional attention switching in selective listening: Exploring the efficiency of nonspatial and spatial selection. The Quarterly Journal of Experimental Psychology Section A (online first publication) (2014).

[12] A. Kiesel, M. Steinhauser, M. Wendt, M. Falkenstein, K. Jost, A.M. Philipp, I. Koch: Control and interference in task switching - A review. Psychological Bulletin **136(5)** (2010) 849–874.

[13] V. Best, E. J. Ozmeral, B. G. Shinn-Cunningham: Visually-guided attention enhances target identification in a complex auditory scene. JARO-Journal of the Association for Research in Otolaryngology **8** (2007) 294–304.

[14] V. Best, B. G. Shinn-Cunningham, E. J. Ozmeral, N. Kopčo: Exploring the benefit of auditory spatial continuity. The Journal of the Acoustical Society of America **127** (2010) EL258–EL264.

[15] G. Kidd, T. L. Arbogast, C. R. Mason, F. J. Gallun: The advantage of knowing where to listen. The Journal of the Acoustical Society of America **118** (2005) 3804.

[16] K. Allen, D. Alais, S. Carlile: Speech intelligibility reduces over distance from an attended location: Evidence for an auditory spatial gradient of attention. Perception & Psychophysics **71** (2009) 164–173.

[17] T. A. Mondor, R. J. Zatorre, N. A. Terrio: Constraints on the selection of auditory information. Journal of Experimental Psychology: Human Perception and Performance **24** (1998) 66.

[18] V. Best, E. J. Ozmeral, F. J. Gallun, K. Sen, B. G. Shinn-Cunningham: Spatial unmasking of birdsong in human listeners: Energetic and informational factors. The Journal of the Acoustical Society of America **118** (2005) 3766.

[19] V. Best, F. J. Gallun, A. Ihlefeld, B. G. Shinn-Cunningham: The influence of spatial separation on divided listening. The Journal of the Acoustical Society of America **120** (2006) 1506.

[20] A. S. Bregman: Auditory scene analysis: The perceptual organization of sound. MIT Press, Massachusetts, 1994 (pp. 58–148).

[21] D. Deutsch: Auditory illusions, handedness, and the spatial environment. Journal of the Audio Engineering Society **31** (1983) 606–620.

[22] G. Kidd, C. R. Mason, A. Brughera, W. M. Hartmann: The role of reverberation in release from masking due to spatial separation of sources for speech identification. Acta Acustica united with Acustica **91** (2005) 526–536.

[23] A. W. Bronkhorst: Localization of real and virtual sound sources. Journal of the Acoustical Society of America **98** (1995) 2542–2553.

[24] F. L. Wightman, D. J. Kistler: Headphone simulation of free-field listening. II: Psychophysical validation. The Journal of the Acoustical Society of America **85** (1989) 868–878.

[25] C. Searle, L. Braida, D. Cuddy, M. Davis: Binaural pinna disparity: another auditory localization cue. The Journal of the Acoustical Society of America **57** (1975) 448–455.

[26] R. A. Butler, K. Belendiuk: Spectral cues utilized in the localization of sound in the median sagittal plane. Journal of the Acoustical Society of America **61** (1977) 1264–1269.

[27] E. M. Wenzel, M. Arruda, D. J. Kistler, F. L. Wightman: Localization using nonindividualized head-related transfer functions. The Journal of the Acoustical Society of America **94** (1993) 111–123.

[28] H. Møller, C. B. Jensen, D. Hammershøi, M. F. Sørensen: Using a typical human subject for binaural recording. An Audio Engeneering preprint presented at the 100th convention, 1996.

[29] J. Fels, B. Masiero, J. Oberem, V. Lawo, I. Koch: Performance of binaural technology for auditory selective attention. The Journal of the Acoustical Society of America **131** (2012) 3317.

[30] B. C. J. Moore: An introduction to the psychology of hearing. 5 ed. Academic Press, San Diego and USA, 2003.

[31] J. Blauert: Spatial hearing - The psychophysics of human sound localization. MIT Press, Cambridge MA, USA, 1997.

[32] Institute of Technical Acoustics, RWTH Aachen: ITA-Toolbox. *www.ita-toolbox.org* 2013.

[33] P. Dietrich, B. Masiero, M. Vorländer: On the optimization of the multiple exponential sweep method. Journal of the Audio Engineering Society **61** (2013) 113–124.

[34] P. Majdak, P. Balazs, B. Laback: Multiple exponential sweep method for fast measurement of head-related transfer functions. Journal of Audio Engineering Society **55** (2007) 623–637.

[35] B. Masiero, J. Fels: Perceptually robust headphone equalization for binaural reproduction. Audio Engineering Society **8388** (2011) 7.

[36] A. Schmitz: Ein neues digitales Kunstkopfmesssystem. Acta Acustica united with Acustica **81** (1995) 416–420.

[37] P. Minnaar, S. K. Olesen, F. Christensen, H. Møller: Localization with binaural recordings from artificial and human heads. Journal of Audio Engineering Society **49** (2001) 323–336.

[38] B. S. Atal, M. R. Schröder: Apparent sound source translator. 1966.

[39] H. Møller: Reproduction of artificial-head recordings through loudspeakers. Journal of the Audio Engineering Society **37** (1989) 30–33.

[40] A. Schmitz: Naturgetreue Wiedergabe kopfbezogener Schallaufnahmen über zwei Lautsprecher mit Hilfe eines Übersprechkompensators. Dissertation. RWTH Aachen University, Aachen Germany, 1993.

[41] T. Lentz: Binaural technology for virtual reality. Dissertation. RWTH Aachen University, Aachen Germany, 2007.

[42] P. Majdak, T. Walder, B. Laback: Effect of long-term training on sound localization performance with spectrally warped and band-limited head-related transfer functions. The Journal of the Acoustical Society of America **134** (2013) 2148–2159.

[43] DIN 45631: Calculation of loudness level and loudness from the sound spectrum - Zwicker method - Amendment 1: Calculation of the loudness of time-variant sound. 03.2010.

[44] W. G. Gardner: 3-D audio using loudspeakers. Dissertation. Massachusetts Institute of Technology, Massachusetts USA, 1997.

[45] T. Takeuchi, P. A. Nelson, H. Hamada: Robustness to head misalignment of virtual sound imaging systems. The Journal of the Acoustical Society of America **109** (2001) 958–971.

[46] T. Lentz, I. Assenmacher, J. Sokoll: Performance of spatial audio using dynamic cross-talk cancellation. Proceedings of the 119th Audio Engineering Society Convention, New York, USA, 2005.

[47] M. R. Bai, C.-C. Lee: Objective and subjective analysis of effects of listening angle on crosstalk cancellation in spatial sound reproduction. The Journal of the Acoustical Society of America **120** (2006) 1976–1989.

[48] V. Best, E. J. Ozmeral, N. Kopčo, B. G. Shinn-Cunningham: Object continuity enhances selective auditory attention. Proceedings of the National Academy of Sciences **105** (2008) 13174–13178.