Examining Different Cue Types in a Listening Test on Auditory Selective Attention

Josefa Oberem[1], Iring Koch[2], Janina Fels[1]

[1] *Medical Acoustics Group, Institute of Technical Acoustics, RWTH Aachen University, Email: job@akustik.rwth-aachen.de*

[2] *Institute of Psychology, RWTH Aachen University*

Introduction

Several new effects have been revealed using the spatially binaural paradigm to investigate the intentional switching of auditory selective attention in realistic environments (e.g. [3, 5, 1, 4, 7]).

It has been shown how participants rarely move within the binaural scene during the experiment [6]. However, in real-life scenarios conversational partners often move towards each other. To be able to examine the participant's motional behavior in greater detail the existing paradigm is being extended towards a paradigm offering more flexibility in free movements within the binaural scene.

One possibility to implement this, is the change in cue presentation. To be able to compare results using the original paradigm and the newly extended paradigm the influence of the cue type and modality must be well-understood. The original paradigm employs an endogenous visual cue consisting of a top-view sketch of the spatial arrangement of the participant and possible speaker positions. It is presented for 500 ms and the stimulus is presented immediately after the cue's disappearing. In the present investigation different enodogenous and exogenous, visual and auditory cue types are compared.

Definition of endogenous and exogenous Cues

In tasks of visual selective attention endogenous and exogenous cues have been extensively examined (for an overview: [2]). Exogenous cues are mainly peripheral/positioned at the spatial location of the actual event and therefore, automatically draw the attention to this location in space (bottom-up). In contrast, endogenous cues are mainly presented centrally and point to the spatial location of the actual event and therefore, the attention has to be actively moved to the event's location (top-down). For short cue-stimulus-intervals (visual: <500 ms [2]) exogenous cues result in shorter reaction times than endogenous cues. To our knowledge, there has been little research about auditory endogenous and exogenous cues [2].

Experimental Task

To analyze the intentional switching of auditory selective attention in spatial environments a spatially binaural paradigm was introduced by Oberem and colleagues [5] which was based on a preliminary dichotic version firstly introduced by Koch and colleagues [3]. A target and a distracting speaker simultaneously present spoken German stimuli from four possible locations around the listener (front, back, left, right). To distinguish between target and distractor the target' spatial location is cued in advance. In the present investigation the cue modality and type is changed between blocks of experimental trials. The used stimuli were spoken digits from 1 to 9, excluding 5 with a total length of 600 ms each. The listener's task is to categorize the target's digit into greater or smaller than 5. Compare Figure 1 for the temporal procedure of an experimental trial.

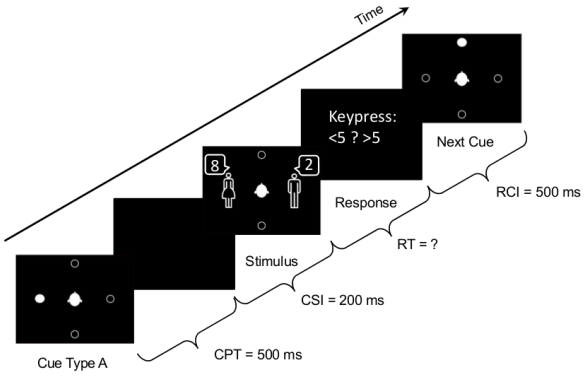


Figure 1: Procedure of a trial exemplary using the cue type A (endogenous, visual sketch) indicating the target direction for a cue-presentation-time (CPT) of 500 ms, a cue-stimulus-interval (CSI) of 200 ms, the synchronous presentation of the stimuli, reaction time between onset of stimulus and the response of the participant, and the response-cue-interval (RCI) of 500 ms.

Experimental Design

Three different independent variables are used in the present investigation, which are defined in following.

Auditory Attention Switch

The auditory attention switch refers to the target's spatial position in two consecutive trials, which can either be repeated from one trial to another (e.g. left - left) or switched between trials (e.g. left - front).

Congruency

The congruency effect is an implicit performance measure of attending the task-irrelevant information. The stimuli of target and distracting speaker can either be congruent (both digits are smaller than five or both are greater than five) or incongruent (one number word is smaller and one is greater than five) within one trial.

Cue types

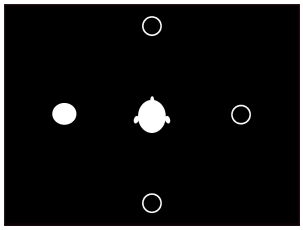
Four cue types are compared in the present examination. The modality of cue types A and B is visual in contrast to the modality of cue types C and D which are auditory. Furthermore, cue types A, B and C are endogenous compared to cue type D, which is exogenous.

A: Visual sketch: A top-view sketch of the participant-source-setup is shown on the monitor, where the source positions are depicted by circles. The target's position is represented by a filled circle (compare Figure 2(a)). Since the spherical arrangment of the target's position and the listener has to be translated into the real world situation, cue type A is endogenous.

B: Visual direction word: The target's spatial location from the participant's point of view is written in large capital letters on the screen. The direction words (front/vorne, back/hinten, left/links and right/rechts) are in German language (compare Figure 2(b)). Cue type B is also endogenous since the direction word has to be cognitively processed to direct the attention to the correct location in space.

C: Auditory direction word: Cue type C only differs in the modality from cue type B. Cue type C is an auditory cue and therefore the German direction word is diotically presented to the participant. With a time stretching algorithm that maintains the original frequencies of the recording, stimuli for the auditory cue are shortened or extended to 500 ms.

D: Auditory pulse: The only exogenous cue type consists of an auditory spatially binaural pulse chain with a total length of 500 ms. The white noise pulse chain consists of two pulses of 200 ms each separated by a break of 100 ms. The pulses are convolved with the same HRTFs as the listening test stimuli itself and therefore appear to be in the same spatial location.



(a) Cue type A: Visual sketch (endogen).

(b) Cue type B: Visual direction word (endogen).

Figure 2: Visual demonstration of cue types A and B.

Subjects

A number of 24 paid (12 Euros), student participants aged between 19 and 30 years (mean age: 23.2±2.3 years) participated voluntarily in the experiment. Nine of them were female participants and 15 were male participants. Listeners are screened to ensure that they have normal hearing (within 20 dB) for frequencies between 250 Hz and 10 kHz. All listeners can be considered as non-expert listeners since they have never participated in a listening test on auditory selective attention.

Results

Reaction times and error rates are measured as dependent variables in every trial and analyzed separately.

Reaction Time

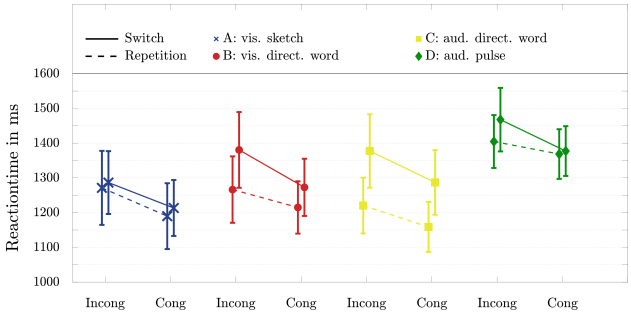


Figure 3: Reaction times (in ms) as a function of cue type, attention switch and congruency (CT × AS × C). Error bars indicate standard errors.

In reaction times, the ANOVA yields a significant main effect of cue type, $[CT\colon F(3,69) = 3.84, p < .05, \eta_p^2 = .14]$. Post-hoc tests show a significant difference between cue type A and cue type D, indicating higher reaction times for the exogeneous auditory pulse cue. The main effect of attention switch is significant, $[AS\colon F(1,23) = 9.77, p < .05, \eta_p^2 = .30]$, indicating longer reaction times in switches than in repetitions. The switch costs amounted on average to 71 ms. The main effect of congruency is also significant, $[C\colon F(1,23) = 11.57, p < .05, \eta_p^2 = .34]$, indicating longer reaction times in incongruent trials than in congruent trials. A significant interaction between the cue type and the attention switch is found $[CT \times AS\colon F(1,69) = 4.49, p < .05, \eta_p^2 = .16]$, indicating no effect of attention switch for cue type A, the endogenous visual sketch. All other interactions are not significant (compare Figure 3).

Error Rate

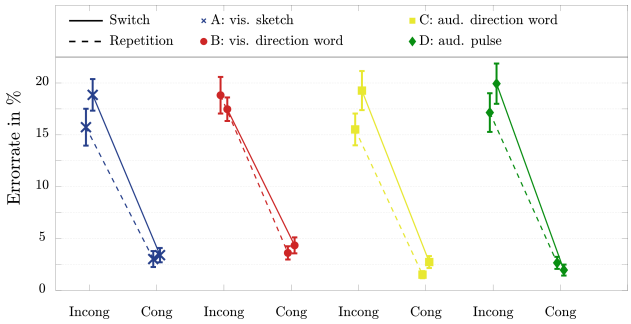


Figure 4: Error rates (in %) as a function of cue type, attention switch and congruency (CT × AS × C). Error bars indicate standard errors.

In error rates, the ANOVA yields no significant main effect of cue type, $[CT\colon F(3,69) = 1.13, p > .05, \eta_p^2 = .05]$. The main effect of attention switch is significant, $[AS\colon F(1,23) = 5.72, p < .05, \eta_p^2 = .20]$, indicating higher error rates in switches than in repetitions. The main effect of congruency is significant, $[C\colon F(1,23) = 206.14, p < .001, \eta_p^2 = .90]$, indicating higher error rates in incongruent trials than in congruent trials. The difference between congruent and incongruent trials amounted on average to 14.9 %. No significant interactions are found (compare Figure 4).

Discussion and Conclusion

The present investigation compares four different cue types used in an examination on the intentional switching of auditory selective attention. Visual and auditory modalities are compared as well as endogenous and exogenous cues.

While results in error rates show no significant difference between cues, the change in cue influences the effects and interactions in reaction times. Apparently, the exogenous cue results in longer reaction times than the endogenous cues. In visual examinations there is an effect called 'inhibition of return' which occurs for long cue-stimulus-intervals. Due to the principle of the attentional spotlight, exogenous cues with a long cue-stimulus-interval result in higher reaction times than those with a short cue-stimulus-interval (for more details [2]). The effect observed here might be due to an auditory version of the 'inhibition of return'.

More important than the absolute values of reaction times for the further use of tested cue types are interactions with the paradigm-specific variables of attention switch and congruency. The attention switch, which has been mainly pronounced in reaction times differs significantly between cue types. Inexplicably, the effect turns out not to be significant with the original cue type A, which has been used in all preceding investigations (e.g. [5, 1, 4, 7]).

However, this interactions shows that the cue type and modality has to be chosen carefully to prevent unexplainable paradigm-specific effects which are in fact due to the cue type or modality.

References

[1] J. Fels, J. Oberem, and I. Koch. Examining auditory selective attention in realistic, natural environments with an optimized paradigm. *PoMA*, 28(1):050001, 2017.

[2] Andrea Kiesel and Iring Koch. Wahrnehmung und Aufmerksamkeit. In Andrea Kiesel and Hans Spada, editors, *Lehrbuch Allgemeine Psychologie*, pages 35–120. Hogrefe & Huber Publishers, 2018.

[3] I. Koch, V. Lawo, J. Fels, and M. Vorländer. Switching in the cocktail party: Exploring intentional control of auditory selective attention. *J. Exp. Psychol. Human*, 37(4):1140–1147, 2011.

[4] J. Oberem, I. Koch, and J. Fels. Intentional switching in auditory selective attention: Exploring age-related effects in a spatial setup requiring speech perception. *Acta Psychol.*, 177:36–43, 2017.

[5] J. Oberem, V. Lawo, I. Koch, and J. Fels. Intentional switching in auditory selective attention: Exploring different binaural reproduction methods in an anechoic chamber. *Acta Acust. united Ac.*, 100(6):1139–1148, 2014.

[6] J. Oberem, J.C. Seibold, I. Koch, and J. Fels. Exploring influences on auditory selective attention by a static and a dynamic binaural reproduction. In *Fortschritte der Akustik: DAGA 2017*, pages 1154–1155, 2017.

[7] J. Oberem, J.C. Seibold, I. Koch, and J. Fels. Intentional switching in auditory selective attention: Exploring attention shifts with different reverberation times. *Hearing research*, 359:32–39, 2018.