# Cue-Related Evoked Potentials Capture Auditory Attention Switches

Lukas J. Vollmer, Leonie Schmitz, Karin Loh, Janina Fels
*Institute for Hearing Technology and Acoustics, RWTH Aachen University, 52074 Aachen, Deutschland*
*Email: lukas.vollmer@akustik.rwth-aachen.de*

## Abstract
Successful communication in acoustically challenging scenarios requires targeted focusing and switching of auditory attention between the conversational partners. When three or more people are involved, the relevant speakers must be selected, and their speech must be extracted from irrelevant acoustic signals. In a cued attention-switching paradigm, auditory evoked potentials of cues and test stimuli were measured to assess attention switches on the neurological level. The cues indicated the spatial location of the subsequent target stimulus, which had to be distinguished from a simultaneously presented distractor originating from a different location. Two consecutive cues indicating a spatial change of the target result in an auditory selective attention switch in space. Test stimuli comprised digits (1, 4, 6, and 9) spoken by a male and a female speaker. Participants had to judge whether the target digit was smaller or larger than five. Behavioral results showed expected switching costs in reaction times, i.e., responses were slower when the target position changed. On the evoked potential level, differences between trials requiring an attention switch and those without were observed.

## Introduction
Selectively focusing auditory attention on a target speaker is a well-known ability of human auditory processing [1]. In conversational settings with more than one interlocutor, switching auditory attention between talkers requires recognition and interpretation of turn-taking cues [2]. With these turn-taking cues, active listeners can switch their auditory selective attention from one talker to the next, ideally before they start speaking. While cued auditory attention switching is well understood behaviorally [3–5], only few studies included electrophysiological measures, e.g., the electroencephalogram (EEG), to investigate the neural basis of the auditory attention switch [6]. In this exploratory study, event-related potentials (ERPs) were measured for switch-inducing auditory cues and target/distractor number word pairs.

## Methods
The experiment is based on a simplified version of the spatial attention switching paradigm by Oberem *et al.* [4]. In brief, an auditory cue is presented from one of four spatial locations around the participant. Matching to the cue's position, its verbal content reflects the same information using direction words, e.g., "front-left". The cue indicates the presentation location of a target digit which must be distinguished from a distractor digit. The cue-stimulus interval had a duration of 1635 ms consisting of 1135 ms cue presentation time and 500 ms silence. Target digits (1, 4, 6, and 9) could be uttered by either a male or female voice and had a duration of 600 ms. Participants were asked to judge whether the target digit was smaller or larger than five. It was recorded whether responses were correct and response time was measured from the onset of the target/distractor digit pair.

Intentional attention switches were induced by changing the cue position, and therefore that of the target, between subsequent trials. If the cue position does not change between trials, no attention switch is required. Trials of the former type are counted as *switch* trials, while the latter are *repetition* trials. A second factor that was investigated is target/distractor congruence, i.e., whether both target and distractor digit stem from the same category.

The EEG was recorded using a Brain Products actiCHamp Plus amplifier system with 32 active electrodes following the international 10-20 electrode layout system. EEG data were preprocessed by downsampling from 2500 Hz to 256 Hz, average re-referencing, and high-pass filtered at 1 Hz. Artifact subspace reconstruction [7] was applied to detect "bad channels", which were removed before using independent component analysis [8] to remove eye-blinks and lateral eye movements. Afterwards, removed channels were reconstructed by spherical interpolation. All processing steps were performed with the EEGLAB 2025.0.0 toolbox [9] for Matlab 2024b.

For the behavioral analysis, error rates were obtained after removing the first trial of each experiment block (because attention switch/repetition is undefined for these trials) and removing trials immediately following error trials, to avoid error-induced carry-over effects. Response times were evaluated for remaining trials with correct responses. Both were analyzed in R 4.3.3 [10] using the package glmmTMB 1.1.10 [11]. Error rates were statistically modeled using a logistic mixed effects regression, while a linear mixed effects model was fit to response speeds ($= 1/\textit{response time}$). Post-hoc tests were performed with emmeans 1.10.7 [12].

## Results
Behavioral results showed expected switching costs [3–5] in the back-transformed response times ($\Delta RT = 30.3\,\text{ms}$, $CI = [17.5, 43.2]\,\text{ms}$, $t = 4.621$, $p < .001$) as increased response times in *switch* trials ($RT = 868\,\text{ms}$, $CI = [753, 983]\,\text{ms}$) compared to *repetition* trials ($RT = 838\,\text{ms}$, $CI = [731, 945]\,\text{ms}$). Congruence effects were found for both error rates ($\Delta ER = 1.53\,\%$, $CI = [0.71, 2.35]\,\%$, $t = 3.652$, $p < .001$) and response times($\Delta RT = 18.4\,\text{ms}$, $CI = [6.85, 29.9]\,\text{ms}$, $t = 3.125$, $p = .002$), showing increased error counts and response

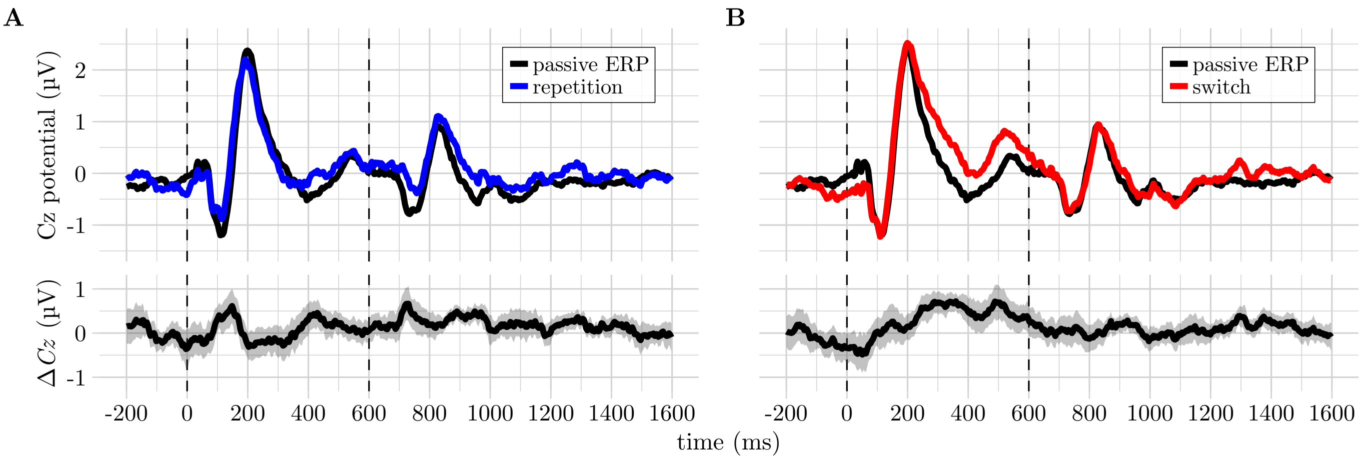


**Figure 1:** Cue-related potentials (top) and difference waves (bottom). Shaded areas around the difference waves indicate standard error. The dashed lines indicate the onset of the two words contained in the cue. (**A**) Comparison of *repetition* trials to cue ERPs measured during passive listening. (**B**) Comparison of *switch* trials to passive cue ERPs.

times for *incongruent* compared to *congruent* trials.

Figure 1 depicts the grand-average cue-related potentials measured during a passive listening session (black) and while participants performed the auditory selective attention task (blue, red). As reflected by the difference waves, there is no sustained activity difference between the passive ERP and those obtained in *repetition* trials, while in *switch* trials, a positive deflection of the ERP between 200 ms and 600 ms after cue onset can be observed. No differences were found for target/distractor-related potentials neither for *switch* compared to *repetition* nor for *incongruent* compared to *congruent* trials.

## Discussion and Conclusion

The results show the selective attention switching cost expected based on previous studies [3–5], though error rates in our results are lower and response times faster than reported. This is most likely explained by the simple task which used exogenous switching cues and included one distractor talker but no background noise.

On the neural level, we found a positive deflection on the cue-related potential between 200 ms and 600 ms after cue onset, most likely reflecting attention switch-related activity. Previous studies using visual cues [6] found sustained activity longer than 600 ms after cue onset. The main reason why our results do not show similar sustained activity is the higher cut-off frequency of 1 Hz applied during EEG preprocessing compared to the 0.25 Hz used by Holmes *et al.* [6]. However, differences could also arise due to the cue being present through the full cue-stimulus interval in [6] or due to differences in the processing of the exogenous auditory cues applied here and the endogenous visual cues used by Holmes *et al.* [6].

In summary, cue-related potentials appear to reflect attention switching processes within a time frame of 200 ms to 600 ms after cue onset. A relevant aspect that could not be explored is error trials for which it is assumed that the attention switch failed. Future studies could increase the difficulty of the task to obtain higher error rates and aim to compare ERPs for switch trials with correct responses to those with erroneous responses. Further, time-frequency and topographical analyses of the attention switch process should be interpreted in the context of attention network theory [13].

## References


[1] E. C. Cherry, "Some Experiments on the Recognition of Speech, with One and with Two Ears," *The Journal of the Acoustical Society of America*, vol. 25, no. 5, pp. 975–979, 1953. DOI: 10.1121/1.1907229.

[2] G. Skantze, "Turn-taking in Conversational Systems and Human-Robot Interaction: A Review," *Computer Speech & Language*, vol. 67, p. 101 178, 2021. DOI: 10.1016/j.csl.2020.101178.

[3] I. Koch, V. Lawo, J. Fels, and M. Vorländer, "Switching in the cocktail party: Exploring intentional control of auditory selective attention.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 37, no. 4, pp. 1140–1147, 2011. DOI: 10.1037/a0022189.

[4] 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 Acustica united with Acustica*, vol. 100, no. 6, pp. 1139–1148, 2014. DOI: 10.3813/aaa.918793.

[5] K. Loh, E. Fintor, S. Nolden, and J. Fels, "Children's intentional switching of auditory selective attention in spatial and noisy acoustic environments in comparison to adults.," *Developmental Psychology*, vol. 58, no. 1, pp. 69–82, 2022. DOI: 10.1037/dev0001239.

[6] E. Holmes, P. T. Kitterick, and A. Q. Summerfield, "EEG activity evoked in preparation for multi-talker listening by adults and children," *Hearing Research*, vol. 336, pp. 83–100, 2016. DOI: 10.1016/j.heares.2016.04.007.

[7] C.-Y. Chang, S.-H. Hsu, L. Pion-Tonachini, and T.-P. Jung, "Evaluation of Artifact Subspace Reconstruction for Automatic EEG Artifact Re-

moval," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Honolulu, HI: IEEE, 2018, pp. 1242–1245. DOI: 10.1109/EMBC.2018.8512547.

[8] A. Delorme, T. Sejnowski, and S. Makeig, "Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis," *NeuroImage*, vol. 34, no. 4, pp. 1443–1449, 2007. DOI: 10.1016/j.neuroimage.2006.11.004.

[9] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, 2004. DOI: 10.1016/j.jneumeth.2003.10.009.

[10] R Core Team, *R: A Language and Environment for Statistical Computing*, version 4.3.3, Vienna, Austria: R Foundation for Statistical Computing, 2024. [Online]. Available: www.R-project.org.

[11] M. E. Brooks, K. Kristensen, K. J. van Benthem, A. Magnusson, C. W. Berg, A. Nielsen, H. Skaug J., M. Mächler, and B. Bolker M., "glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling," *The R Journal*, vol. 9, no. 2, p. 378, 2017. DOI: 10.32614/RJ-2017-066.

[12] R. V. Lenth, *Emmeans: Estimated Marginal Means, aka Least-Squares Means*, version 1.10.2, 2024. [Online]. Available: https://github.com/rvlenth/emmeans.

[13] S. Markett, D. Nothdurfter, A. Focsa, M. Reuter, and P. Jawinski, "Attention networks and the intrinsic network structure of the human brain," *Human Brain Mapping*, vol. 43, no. 4, pp. 1431–1448, 2022. DOI: 10.1002/hbm.25734.