Neural tracking on the go: Auditory attended speaker decoding and saliency detection using mobile EEG

Straetmans, L.[1], Holtze, B. [1], Debener, S.[1,2], Jaeger, M.[1,3], & Mirkovic, B.[1]

[1] Neuropsychology Lab, Department of Psychology, Carl von Ossietzky University of Oldenburg, Oldenburg, Germany.

[2] Cluster of Excellence Hearing4all, University of Oldenburg, Oldenburg, Germany.

[3] Fraunhofer Institute for Digital Media Technology, IDMT, Division Hearing, Speech and Audio Technology, Oldenburg, Germany

Introduction: When we are listening to someone within a noisy environment, our auditory system allows us to follow the attended speaker despite concurrently ongoing sounds like other conversations (Cherry, 1953). Interestingly, while we are processing the speech of interest, it has been shown that neural activity tracks changes in the attended speech signal more strongly than those in the ignored speech (Mesgarani & Chang, 2012; Kerlin et al., 2010). Based on this finding, methods have been developed to decode the attended speaker within a multi-speaker environment (Alickovic et al., 2019; Ding & Simon, 2012). This offers opportunities for promising future applications in assistive devices (Slaney et al., 2020) such as neuro-steered hearing aids (Geirnart et al., 2021) or other brain-computer interfaces (Gao et al, 2014), especially in complex, uncontrolled natural auditory scenes. However, so far, these methods have been tested exclusively inside the lab, under controlled conditions. It remains unknown if the attended speaker can still be decoded when the neural data has been collected in more ecologically valid situations.

Methods: The present study takes a commonly used attended speaker decoding method out of the lab into a more realistic, ecologically valid scenario by addressing two issues: First, we determined whether attended speaker decoding is possible during leisure walking. Second, we investigated how bottom-up distraction impacts the top-down driven neural impulse response to an attended speech stream. We used a well-established two-competitive speaker paradigm (O'Sullivan et al., 2014; Mirkovic et al., 2016) to investigate the dynamics of auditory attention. Participants (N=19; 4 male, mean age 24.2 years) were instructed to attend to one of two simultaneously presented, spatially separated, continuous speech streams. To-be-attended and to-be-ignored speech streams consisted of short stories of an audio book, narrated by a male speaker and were presented via in-ear headphones. In a third auditory stream various natural environmental sounds served as transient salient events. Participants alternately sat on a chair or walked along an indoor route. Each block (sitting, walking) lasted approximately 5 minutes and was repeated three times. Neurophysiological responses were recorded using 24 channel mobile EEG system (SMARTING, mBrainTrain, Belgrade, Serbia). The first objective of this study was to compare the decoding accuracy of a representative, commonly used backward auditory attention decoding (AAD) model (Crosse et al., 2016; O'Sullivan et al., 2014) between the mobile and stationary condition. As artifact attenuation is an integral part of mobile EEG pre-processing, the influence of two different EEG artifact attenuation methods (Artifact Subspace Reconstruction (ASR), Independent Component Analysis (ICA)) on model performance was investigated as well. Additionally, the effect of trial length was investigated investigating AAD model performance in trials of different window lengths (5-s vs 60 s). The second objective was to investigate the effect of salient distractor events on the neural tracking of the to-be-attended and to-be-ignored speech stream by using ICA-attenuated data, 5 seconds before and 5 seconds after a salient event.

Results: Results confirmed that decoding accuracy (O'Sullivan et al.,2014, Crosse et al., 2016) was higher for 60-s long trials than for 5-s long trials (no attenuation: $t = -11.9$, $p < 10^{-10}$; ASR: $t = -7.6$, $p < 10^{-7}$; ICA: $t = -16.1$, $p < 10^{-12}$). Accuracy was above chance level across both tested trial lengths (5-s, 60-s) in both movement conditions (Figure 1), although higher in sitting compared to walking

condition (60-s: $t = 4.01$, $p < 0.001$; 5-s: $t = 3.42$, $p = 0.003$). In a second analysis, a forward model was trained to get a better estimate of neuropsychological processes in moments of distraction (Crosse et al., 2016). Results show that the magnitude of the neural impulse response to the to-be-attended speaker dropped significantly after the occurrence of a salient event ($t = 4.85$, $p < 10^{-4}$). This was found in both movement conditions. Interestingly, the magnitude of the neural impulse response to the to-be-ignored speaker also decreased significantly after the occurrence of a salient event, although to a lesser extent ($t = 2.6$, $p = 0.018$).

Conclusion: Our study demonstrates, for the first time, successful auditory attention decoding while listeners were walking freely. Even with AAD evaluation periods as short as 5 seconds, predicting the attended speaker was possible. This finding holds for artifact attenuated as well as uncorrected data. Furthermore, we confirmed the effect of transient salient events on sustained attention, using the neural impulse responses to attended speech. Limited attentional resources appear to be recruited by salient events, before they can be re-directed to the task at hand, and this appears to be the case in stationary as well as mobile scenarios.

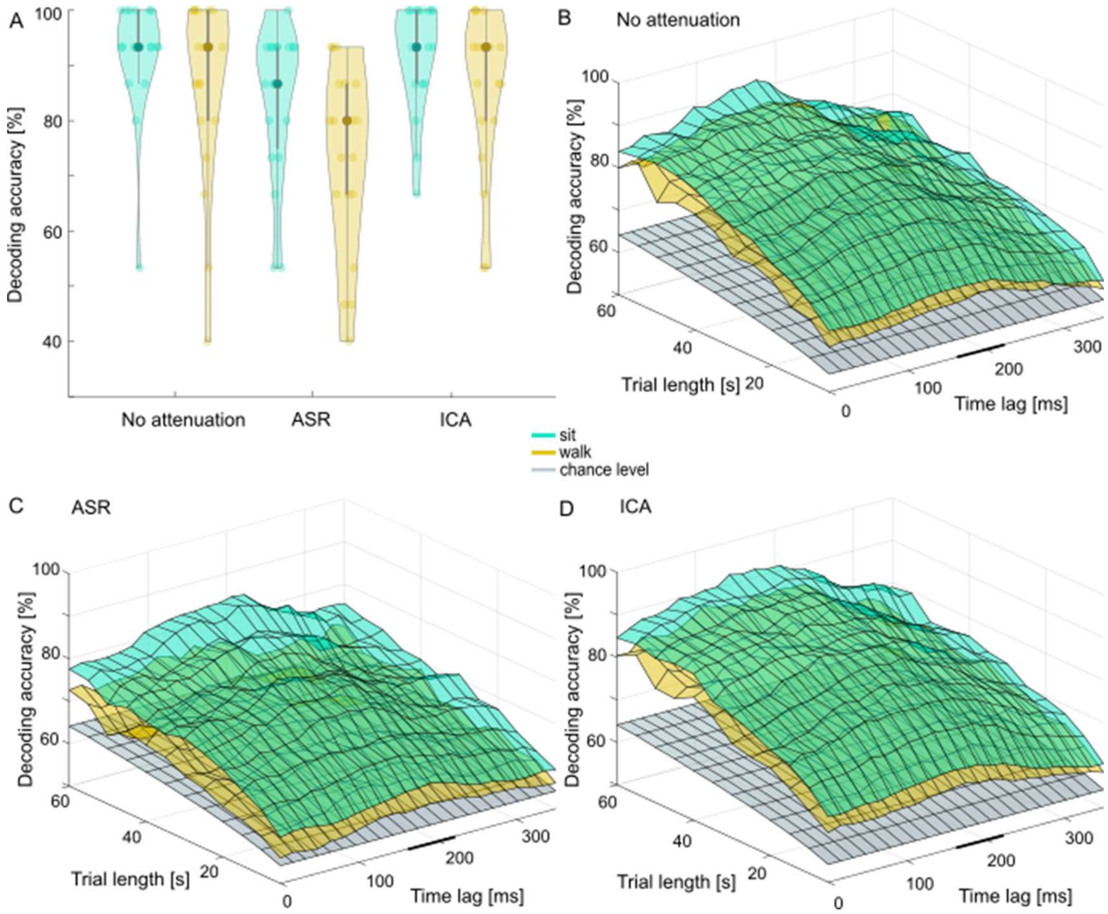

*Figure 1 | **A**: Distribution of decoding accuracy between 165-210ms after speech envelope in 60-s trials. **B**: Average decoding accuracy for all trial lengths (s) in latencies from 0 to 300 ms relative to speech envelope. Sitting (green) and walking (yellow) data without no artifact attenuation. Chance level for each trial length indicated in grey. **C:** Results for ASR attenuated data. **D:** Results for ICA attenuated data.*

Alickovic, E., Lunner, T., Gustafsson, F., & Ljung, Lennart. (2019). A Tutorial on Auditory Attention Identification Methods. *Front. Neurosci., 13:153*. doi: 10.3389/fnins.2019.00153

Cherry, E. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America, 25:5,* 975-979. doi: 10.1121/1.1907229

Crosse, M. J., Di Liberto, G.M., Bednar, A., & Lalor, E.C. (2016). The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Front. Hum. Neurosci., 10:604*. doi: 10.3389/fnhum.2016.00604

Ding, N., & Simon, J. (2012) E,ergence of neural encoding of auditory objects while listening to competing speakers. *PNAS, 109:20,* 11854-11859. doi: 10.1073/pnas.1205381109

Gao, S., Wang, Y., Gao, X., & Hong, B. (2014). Visual and Auditory Brain Computer Interfaces. IEEE Transactions in Biomedican Engineering, 61:5, 1436-1447. doi: 10.1109/tbme.2014.2300164

Geirnart, S., Vandecappelle, S., Alickovic, E., de Cheveigné, Lalor, E., Meyer, B.T., Miran, S., Francart, T., & Bertrand, A. (2021). EEG-based Auditory Attention Decoding – Towards Neuro-Steered Hearing Devices. *IEEE Signal Processing Magazine*. arXiv:2008.04569

Kerlin, J.R., Shahin, A.J., & Lee, M.M. (2010). Attentional Gain Control of Ongoing Cortical Speech Representations in a “Cocktail Party”. *The Journal of Neuroscience, 30:2,* 620-628. 10.1523/JNEUROSCI.3631-09.2010

Mesgarani, N., & Chang, E.F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature, 485,* 233-236. doi:10.1038/nature11020

Mirkovic, B., Bleichner, M.G., de Vos, M., and Debener, S. (2016). Target speaker detection with concealed EEG around the ear. *Front. Neurosci., 10:349*. doi:10.3389/fnins.2016.00349

O’Sullivan, J.A. Power, A.J. Mesgarani, N., Siddharth, R., Foxe, J.J. Shinn-Cunningham, B.G., Slaney, M., Shamma, S.A. & Lalor. E.C. (2015). Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trail EEG. *Cerebral Cortex, 25,* doi: 1697-1706. 10.1093/cercor/bht355

Slaney, M., Lyon, R.F., Garcia, R., Kemler, B., Gnegy, C., Wilson, K., Kanevsky, D., Savla., Sagar, & Cerf, Vinton (2020). Auditory measures for the next billion users. *Ear & Hearing, 41,* 131-139. doi: 10.1097/AUD.0000000000000955