Influence of Noisy Reference Signals on Selective Attention Decoding

Ali Aroudi, *Member, IEEE*, Bojana Mirkovic, Maarten de Vos, *Member, IEEE*, and Simon Doclo, *Senior Member, IEEE*

***Abstract*— In a recently proposed method for selective attention decoding [1] using electroencephalography (EEG) recordings the clean speech signals of both attended and unattended competing speakers are required as reference signals. In this study, the influence of noisy reference signals on the decoding performance was investigated. The results show that the decoding performance is robust to noise up to a certain signal-to-noise ratio, depending on the acoustic noise type.**

I. Introduction and Problem Statement

In [1] a least-squares method has been proposed to decode selective attention from EEG recordings in a cocktail-party situation with two competing speakers. This method however requires the clean speech signals of both the attended and the unattended speaker to be available as reference signals, which is hard –if not impossible– to achieve using acoustical signal processing algorithms in practice [2].

Let us consider a cocktail-party situation, where the ongoing EEG responses $\boldsymbol{R}$ of a participant listening to two competing speakers have been recorded. The attended and unattended clean speech signals are denoted as $\boldsymbol{x}_a$ and $\boldsymbol{x}_u$, respectively, and their corresponding envelopes as $\boldsymbol{s}_a$ and $\boldsymbol{s}_u$, which are assumed to be available in the method proposed in [1]. To decode a participant's selective attention using a decoder $\boldsymbol{W}$, an estimate of the attended speech envelope is computed as $\hat{\boldsymbol{s}}_a = \boldsymbol{W}^T\boldsymbol{R}$, i.e. a linear combination of the EEG responses. Based on the correlation coefficients $\rho_a = \rho(\hat{\boldsymbol{s}}, \boldsymbol{s}_a)$ and $\rho_u = \rho(\hat{\boldsymbol{s}}, \boldsymbol{s}_u)$, it is then decided that the selective attention has been correctly decoded when $\rho_a > \rho_u$.

To investigate the influence of acoustic noise on the selective attention decoding performance of this method, we assume here that the reference signals used for decoding are not equal to the clean speech signals but have been corrupted by noise (e.g., residual noise of an acoustical pre-processing algorithm), i.e.

$$\tilde{\boldsymbol{x}}_a = \boldsymbol{x}_a + \alpha\boldsymbol{n}_{x_a}, \quad \tilde{\boldsymbol{x}}_u = \boldsymbol{x}_u + \alpha\boldsymbol{n}_{x_u} \tag{1}$$

where $\boldsymbol{n}_{x_a}$ and $\boldsymbol{n}_{x_u}$ denote different realizations of the same acoustic noise type, which are assumed to be mutually independent from $\boldsymbol{x}_a$ and $\boldsymbol{x}_u$, and α is a scalar, determining the amount of noise. The amount of noise is characterized by the signal-to-noise ratio (SNR), i.e. $\mathrm{SNR}_{\tilde{x}_a/\tilde{x}_u} = 10\log\left(P_{x_a/x_u} / \alpha^2 P_{n_{x_a}/n_{x_u}}\right)$ where P_{x_a/x_u} denotes the power of either $\boldsymbol{x}_a$ or $\boldsymbol{x}_u$ and $P_{n_{x_a}/n_{x_u}}$ denotes the noise power of either $\boldsymbol{n}_{x_a}$ and $\boldsymbol{n}_{x_u}$.

A. Aroudi, B. Mirkovic and S. Doclo are with the University of Oldenburg, Germany. M. De Vos is with the Institute of Biomedical Engineering, Oxford University, UK. This work was supported in part by the PhD Program "Signals and Cognition" and the Cluster of Excellence 1077 "Hearing4All", funded by the German Research Foundation (DFG).

II. Experimental Evaluation and Summary

Using earphones eight participants diotically listened to two stories uttered by male and female speakers, each of length 48 min, and their ongoing EEG responses were recorded using 96-channel electrodes [3]. The participants were instructed to attend to only one story during the whole experiment. For each participant the 48-minute EEG responses were split into 24 trials, each of length 2 min. The leave-one-out cross-validation approach was used for training and testing, i.e. 23 trials were used for training and 1 trial was used for testing. To evaluate the influence of noisy reference signals on the decoding performance (DP), two experimental setups were considered: (a) without noise ($\alpha = 0$), and (b) with noise (either white or speech-shaped noise) where $\mathrm{SNR} = \mathrm{SNR}_{\tilde{x}_a} = \mathrm{SNR}_{\tilde{x}_u}$.

Figure 1 depicts the DP of each experimental setup, averaged over all participants, for different SNRs ranging from –30 to 30dB. When no noise is present, the DP is equal to 91.1%. When noise is present, as expected the DP gradually decreases and drops below 88.0% for $\mathrm{SNR} \leq -4$dB (white noise) and $\mathrm{SNR} \leq -12$dB (speech-shaped noise).

In conclusion, these results show that the least squares-based selective attention decoding method is to some extent robust to reference signals corrupted by acoustic noise.

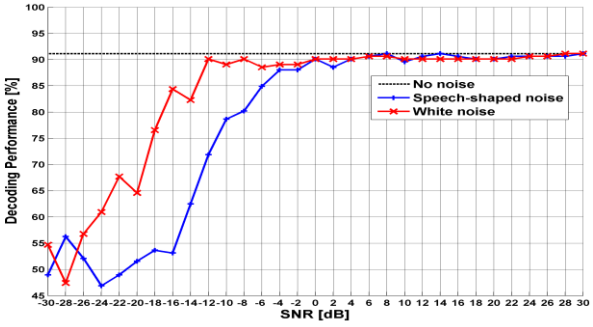

Figure 1. Decoding performance averaged across participants for different SNRs (white noise and speech-shaped noise)

References

[1] J. A. O'Sullivan, A. J. Power, N. Mesgarani, S. Rajaram, J. J. Foxe, B. G. Shinn-Cunningham, M. Slaney, S. A. Shamma, and E. C. Lalor, "Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG," *Cerebral Cortex*, 2014. doi:10.1093/cercor/bht355.

[2] S. Doclo, W. Kellermann, S. Makino, S. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 18-30, 2015.

[3] B. Mirkovic, S. Debener, M. Jaeger, and M. De Vos, "Decoding the attended speech stream with multi-channel EEG: implications for online, daily-life applications," *Journal of Neural Engineering*, vol. 12, no. 4, pp. 46007, 2015.