EEG-BASED AUDITORY ATTENTION DECODING USING STEERABLE BINAURAL SUPERDIRECTIVE BEAMFORMER

Ali Aroudi, Daniel Marquardt, Simon Doclo

Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4All,
University of Oldenburg, Oldenburg, Germany
ali.aroudi@uni-oldenburg.de

ABSTRACT

During the last decades significant progress in multi-microphone speech enhancement algorithms has been made for hearing aids. However, the performance of many algorithms depends on identifying the target speaker to be enhanced. To identify the target speaker from single-trial EEG recordings in an acoustic scenario with two competing speakers, an auditory attention decoding (AAD) method was recently proposed. This AAD method however requires the clean speech signals of both the attended and the unattended speaker as reference signals for decoding. Since in practice only microphone signals, containing several undesired acoustic components, are available, in this paper we explore the potential of using steerable binaural superdirective beamformer for generating appropriate reference signals for decoding. The experimental results show that using steerable superdirective beamformer output signals improves the decoding performance compared to using the noisy microphone signals as reference signals.

Index Terms— auditory attention decoding, steerable beamformer, noisy signal, speech envelope, noise reduction, EEG signal, brain computer interface

1. INTRODUCTION

In complex listening conditions the human auditory system has a remarkable ability to separate a speaker of interest from a mixture of speakers or to tune out interfering sounds in a noisy environment, known as the cocktail-party paradigm [1]. However, this auditory scene analysis task may be very challenging for listeners with hearing impairment. During the last decades significant progress in multi-microphone speech enhancement algorithms has been made to improve speech intelligibility for hearing-impaired listeners [2–4]. In an acoustic scenario with multiple speakers, these speech enhancement algorithms however need to know which is the target speaker to be enhanced. In hearing aid applications, the target speaker is typically identified based on direction (e.g., the speaker is in front of the hearing aid user) or energy (i.e., the loudest speaker). However, in many realistic scenarios, e.g., two speakers with similar energy in front of the hearing aid user, this is not possible.

In [5] an auditory attention decoding (AAD) method was proposed to identify the attended speaker from single-trial EEG recordings in an acoustic scenario with two competing speakers. This AAD method has recently attracted a lot of attention, e.g., for controlling hearing aid processing [6–19]. The AAD method proposed in [5] aims to reconstruct the attended speech envelope from EEG recordings using a spatio-temporal filter. During the training step, the filter coefficients are computed by minimizing the least-squares error between the attended speech envelope and the reconstructed envelope. In the decoding step, the clean speech signals of both the attended and the unattended speaker are required as reference signals. However, in practice only the microphone signals, containing several undesired acoustic components, are available.

The feasibility of AAD using the noisy microphone signals as reference signals was shown in [14, 15], although the obtained decoding performance was significantly lower than when using the clean speech signals as reference signals. To generate better reference signals from the microphone signals, in [11, 13] a noise reduction algorithm using multi-channel Wiener filtering and in [16] a source separation algorithm using deep neural networks was proposed. In this paper, we explore the potential of using steerable superdirective beamformer for generating appropriate reference signals for decoding. A superdirective beamformer aims to pass the signals arriving from a certain direction without any distortion while minimizing the diffuse noise power [3, 20]. We evaluate the performance of steerable superdirective beamformer for AAD in anechoic and noisy conditions by considering a binaural configuration which uses the microphone signals of the left and the right hearing aid simultaneously.

For an acoustic scenario comprising two competing speakers and diffuse noise at different SNRs, 64-channel EEG responses with 18 participants were recorded. The experimental results show that using steerable superdirective beamformer output signals improves the decoding performance compared to using the noisy microphone signals as reference signals.

2. AUDITORY ATTENTION DECODING USING STEERABLE SUPERDIRECTIVE BEAMFORMER

In this section the least-squares-based AAD method using a steerable superdirective beamformer is presented. In Section 2.1 the acoustic scenario and the notation are defined. In Section 2.2 the binaural superdirective beamformer used for generating the reference signals are presented. In Section 2.3 the decoding and the training step of the AAD method are discussed.

2.1. Acoustic scenario and notation

Consider an acoustic scenario comprising two competing speakers and background noise component (Fig. 1), where $s^1[n]$ and $s^2[n]$ denote the clean speech signal, with n the discrete time index. The ongoing EEG responses of a listener to these acoustic stimuli are

This work was supported in part by the Cluster of Excellence 1077 Hearing4All, funded by the German Research Foundation (DFG).

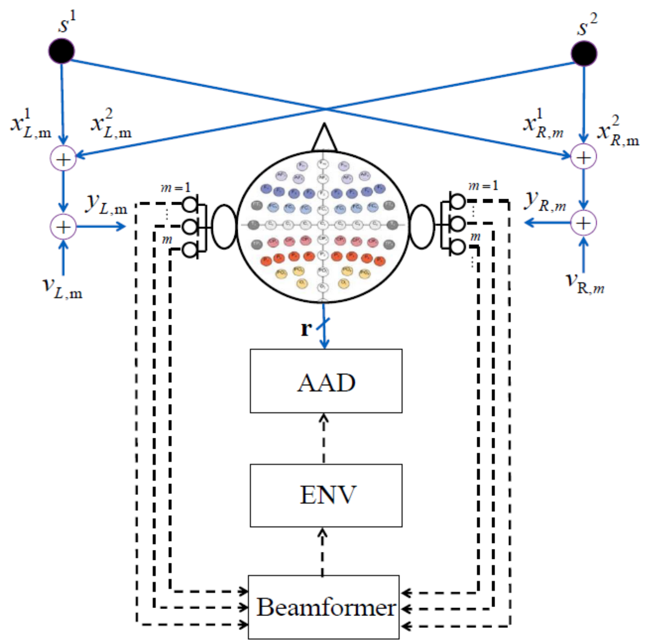

Fig. 1. The schematic of the AAD method using steerable binaural superdirective beamformer.

recorded and used for AAD (see Section 2.3). The m-th microphone signal at the left hearing aid can be written as

$$y_{L,m}[n] = x_{L,m}^1[n] + x_{L,m}^2[n] + v_{L,m}[n], \tag{1}$$

with $x_{L,m}^1[n]$ and $x_{L,m}^2[n]$ the anechoic speech components corresponding to the speakers, and $v_{L,m}[n]$ the background noise. The m-th microphone signal at the right hearing aid $y_{R,m}[n]$ is defined similarly as in (1).

In the short-time Fourier transform (STFT) domain, the m-th microphone signal at the left hearing aid can be written as

$$Y_{L,m}(k,l) = X_{L,m}^1(k,l) + X_{L,m}^2(k,l) + V_{L,m}(k,l), \tag{2}$$

with k denoting the frequency index and l denoting the frame index. The m-th microphone signal at the right hearing aid $Y_{R,m}(k,l)$ is defined similarly. For notational conciseness the frequency index k, the frame index l, and the time index n will be omitted in the remainder of this paper. We also define the M-dimensional signal vectors $\mathbf{Y}_L = [Y_{L,1}, \ldots, Y_{L,M}]^T$ and $\mathbf{Y}_R = [Y_{R,1}, \ldots, Y_{R,M}]^T$, and the $2M$-dimensional stacked signal vector $\overline{\mathbf{Y}} = \left[\mathbf{Y}_L^T, \mathbf{Y}_R^T\right]^T$.

2.2. Steerable superdirective beamformer

Assuming diffuse nosie, a superdirective beamformer aims at minimizing the output power of the noise component subject to a unit response constraint for the steering angle θ_s [3, 20, 21]. In this paper, we will consider a binaural configuration which uses the microphone signals of the left and the right hearing aid simultaneously. The $2M$-dimensional binaural beamformer steered towards θ_s can be computed as

$$\mathbf{W}(\theta_s) = \frac{(\mathbf{\Gamma} + \alpha \mathbf{I})^{-1} \mathbf{d}(\theta_s)}{\mathbf{d}^H(\theta_s)(\mathbf{\Gamma} + \alpha \mathbf{I})^{-1} \mathbf{d}(\theta_s)}, \tag{3}$$

with $\mathbf{\Gamma}$ the binaural spatial coherence matrix of a diffuse noise field, $\mathbf{d}(\theta_s)$ the (anechoic) relative transfer function (RTF) vector corresponding to θ_s, which can be obtained from anechoic measurements, $\mathbf{I}$ the identity matrix, and $\alpha = 10^{-2}$ a regularization parameter [20]. The output signal of the binaural beamformer in the time-domain is computed as

$$z(\theta_s) = \text{ISTFT}\left\{\mathbf{W}^H(\theta_s)\mathbf{Y}\right\}, \tag{4}$$

with ISTFT denoting the inverse short-time Fourier transform.

To investigate the potential of using steerable superdirective beamformer for AAD, we will consider the beamformer output signals $z(\theta_L)$ and $z(\theta_R)$, with steering angles towards the left side of the listener θ_L and the right side of the listener θ_R, as reference signals.

2.3. Auditory attention decoding

2.3.1. Decoding step

To decode auditory attention from C-channel EEG recordings $r_c[i]$, with $c = 1 \ldots C$ and i the sub-sampled time index, using a trained spatio-temporal filter $\mathbf{g}$ (see Training step 2.3.2), an estimate of the attended speech envelope $\hat{e}^a$ is reconstructed as, i.e.,

$$\hat{e}^a[i] = \mathbf{g}^T \mathbf{r}[i], \tag{5}$$

with $\mathbf{g} = \left[\mathbf{g}_1^T\, \mathbf{g}_2^T \ldots \mathbf{g}_C^T\right]^T$, $\mathbf{g}_c = [g_{c,0}\, g_{c,1} \ldots g_{c,J-1}]^T$, $\mathbf{r}[i] = \left[\mathbf{r}_1^T[i]\; \mathbf{r}_2^T[i] \ldots \mathbf{r}_C^T[i]\right]^T$, and $\mathbf{r}_c[i] = [r_c[i+\Delta]\; r_c[i+\Delta+1] \ldots r_c[i+\Delta+J-1]]^T$, where $g_{c,j}$ denotes the j-th filter coefficient in the c-th channel, J the number of filter coefficients per channel, and Δ models the latency of the attentional effect in the EEG responses to the acoustic stimuli.

Based on the correlation coefficients between the reconstructed speech envelope $\hat{e}^a[i]$ and the speech envelopes $e^1[i]$ and $e^2[i]$ of the first and the second speaker, respectively, i.e.,

$$\rho^1 = \rho\left(e^1[i], \hat{e}^a[i]\right), \; \rho^2 = \rho\left(e^2[i], \hat{e}^a[i]\right), \tag{6}$$

it is then decided that the listener was attending to, e.g., the first speaker if $\rho^1 > \rho^2$. Accordingly, the difference between the correlation coefficients ρ^1 and ρ^2 (correlation difference) is indicative of the reliability of the AAD decision.

The complete set of EEG recordings is segmented into $t = 1 \ldots T_{tr}$ trials and the filter $\mathbf{g}_t$ corresponding to trial t is used for decoding this trial. The decoding performance is defined as the percentage of correctly decoded trials over all considered trials and over all participants.

In previous work, e.g., [5,22], it has often been assumed that the clean speech signals s^1 and s^2 are available as reference signals for computing the speech envelopes $e^1[i]$ and $e^2[i]$ in (6), which is quite unrealistic in practice. In this paper we address this issue by using the output signals of steerable superdirective beamformer $z(\theta_L)$ and $z(\theta_R)$ as reference signals.

2.3.2. Training step

In the training step, the attended speaker is assumed to be known and the attended speech envelope $e^a[i]$ is used to compute the filter coefficients $g_{c,j}$. The attended clean speech signal s^a is typically used for computing $e^a[i]$. The filter $\mathbf{g}$ is computed by minimizing the least-squares error between the attended speech envelope $e^a[i]$

(assumed to be known) and the reconstructed envelope $\hat{e}^a[i]$, regularized with the squared l_2-norm of the derivatives of the filter coefficients to avoid over-fitting [5], i.e.,

$$F(\mathbf{g}) = \frac{1}{I}\sum_{i=1}^{I}\left(e^a[i] - \mathbf{g}^T\mathbf{r}[i]\right)^2 + \beta\mathbf{g}^T\Lambda\mathbf{g}, \tag{7}$$

with β the regularization parameter, Λ the derivative matrix, and I the total number of EEG samples or envelope samples considered for computing F. The filter minimizing the regularized cost function in (7) is equal to [7,9]

$$\mathbf{g} = (\mathbf{Q} + \beta\Lambda)^{-1}\mathbf{q}, \tag{8}$$

with $\mathbf{Q} = \frac{1}{I}\sum_{i=1}^{I}\left(\mathbf{r}[i]\,\mathbf{r}^T[i]\right)$ the correlation matrix, and $\mathbf{q} = \frac{1}{I}\sum_{i=1}^{I}\left(\mathbf{r}[i]\,e^a[i]\right)$ the cross-correlation vector. The correlation matrix and the the cross-correlation vector corresponding to trial $t = 1\cdots T_{tr}$ are defined as $\mathbf{Q}_t$ and $\mathbf{q}_t$, respectively. The filter to decode trial t is computed as $\mathbf{g}_t = \left(\tilde{\mathbf{Q}}_t+\beta\Lambda\right)^{-1}\tilde{\mathbf{q}}_t$ with $\tilde{\mathbf{Q}}_t$ computed by averaging all correlation matrices except $\mathbf{Q}_t$ (known as leave one-out averaging), and $\tilde{\mathbf{q}}_t$ computed by averaging all cross-correlation vectors except $\mathbf{q}_t$.

3. EXPERIMENTAL SETUP

EEG responses were recorded for 18 native German-speaking participants aged between 21 and 34 years with self-reported normal hearing. Two German stories, uttered by two different male speakers, were simultaneously presented to the participants using earphones. The presented stimuli at both ears were generated by convolving the clean speech signals, i.e., stories, with (non-individualized) binaural anechoic impulse responses from [23], and adding diffuse noise, generated according to [24]. The left and the right speakers were simulated at $-45°$ and $45°$, respectively. Two acoustic conditions were considered: noiseless, and noisy with two different average signal-to-noise ratios, i.e., $\mathrm{SNR} = 9.0$ dB and $\mathrm{SNR} = 4.0$ dB. The stimuli for the noiseless condition were presented in 4 sessions, each of length 10 minutes, while for the noisy condition were presented in 2 sessions, one per each SNR[1]. Among all participants, 8 participants were instructed to attend to the left speaker, while 10 participants were instructed to attend to the right speaker. During the breaks, the participants were asked to fill out a questionnaire consisting of 10 multiple-choice questions related to each story. Two participants were excluded from the analysis, one participant due to poor attentional performance (as revealed by the questionnaire results) and the other one due to a technical hardware problem.

The EEG responses were recorded using $C = 64$ channels at a sampling frequency of 500 Hz, and referenced to the nose electrode. The EEG responses were offline re-referenced to a common average reference, band-pass filtered between 2 Hz and 8 Hz using a third-order Butterworth band-pass filter, and subsequently downsampled to 64 Hz. The envelopes of the speech signals were obtained using a Hilbert transform, followed by low-pass filtering at 8 Hz and downsampling to 64 Hz. For the training and decoding steps (see Sections 2.3), the EEG recordings of each session were split into 20 trials, each of length 30 seconds. Each participant's own data were used for filter training and evaluation. The parameters involved in the filter design were set to fixed values as $\Delta = 125$ ms, $J = 8$ (corresponding to 125 ms), and $\beta = 10^4$.

The positions of the attended and the unattended speaker are not always the same, i.e., sometimes the attended speaker is on the left side of the listener (and the unattended speaker is on the right side), while sometimes the attended speaker is on the right side (and the unattended speaker is on the left side). In order to assess the AAD performance using the anechoic speech signals as reference signals, we consider the anechoic speech signals of the hearing aid at the side of the attended speaker as the anechoic attended speech signals and the anechoic speech signals of the hearing aid at the side of the unattended speaker as the anechoic unattended speech signals. To assess the AAD performance using the microphone signals as reference signals, we consider the microphone signals of the hearing aid at the side of the attended speaker as the microphone attended speech signals and the microphone signals of the hearing aid at the side of the unattended speaker as the microphone unattended speech signals. For both the anechoic and the microphone reference signals the first microphone of the left and the right hearing aid is used.

To generate reference signals using the superdirective beamformer with $M = 3$, the microphone signals are processed using a weighted overlap-add framework with a frame size of 512 samples and an overlap of 50%. The diffuse coherence matrix $\mathbf{\Gamma}$ is calculated using spatially averaged auto- and cross-correlations of the anechoic acoustic transfer functions (ATFs) measured with an angle resolution of $5°$ [23]. For the steering angle θ_s the RTF vector $\mathbf{D}(\theta_s)$ is computed from the same anechoic ATFs. For the RTF vector the first microphone of left hearing aid is arbitrarily selected as the reference microphone. The beamformer output signal is computed for the target angles $(\theta_L, \theta_R) = (-45°, 45°)$ where speakers are positioned. In order to assess a possible mismatch between steering and target angles the beamformer output signal is also computed for $(-75, 75°)$ and $(-90°, 90°)$. The interference and the noise reduction performance of the beamformer steered towards the left side of the listener ($\mathbf{W}(\theta_L)$) is evaluated in terms of the broadband signal-to-interference ratio SIR_L and signal-to-noise ratio SNR_L improvements between the output signal $z(\theta_L)$ and the first microphone signals of the left hearing aid $y_{L,1}[n]$. For the beamformer steered towards the right side of the listener ($\mathbf{W}(\theta_R)$) the broadband SIR_R and SNR_R improvement is similarly evaluated between $z(\theta_R)$ and $y_{R,1}[n]$. The average SIR and SNR improvements are computed as $\mathrm{SIR} = \frac{\mathrm{SIR}_L+\mathrm{SIR}_R}{2}$ and $\mathrm{SNR} = \frac{\mathrm{SNR}_L+\mathrm{SNR}_R}{2}$, respectively.

4. RESULTS AND DISCUSSION

For the binaural beamformer, Fig. 2 depicts the SIR and SNR improvements, computed over 30 trials, for different steering angles. It can be observed that for all steering angles both the SIR and the SNR can be improved. The best SIR and SNR improvements are obtained for the target steering angles $\pm 45°$ where the speakers are positioned. This result is expected since the beam of the superdirective beamformer is relatively narrow [3,20,21] and if it is steered towards different angles rather than the target angle some target source suppression will occur.

For the noiseless and the noisy condition, Fig. 3 a, b depicts the decoding performance when using either the microphone signals, the anechoic signals, or the output signals of the superdirective beamformer steered towards different angles. It can be observed that when using microphone signals as reference signals the decod-

[1]The data for this experiment are a subset of the EEG measurements in several acoustic conditions from [14,15].

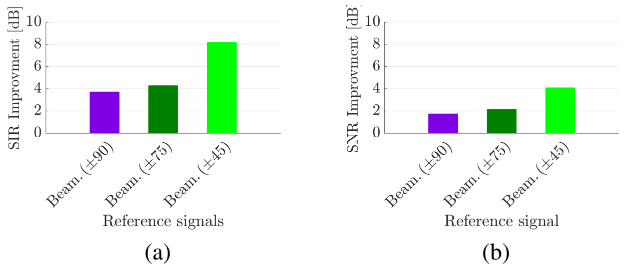

Fig. 2. SIR improvement in the noiseless and the noisy condition (a) and SNR improvement in the noisy condition (b) obtained by the superdirective beamformer steered at different angles.

ing performance for both conditions is large[2], however, substantially lower than when using anechoic signals as reference signals, consistent with previous findings in [14, 15]. When using the superdirective beamformer output signals, the decoding performance for all considered steering angles are considerably improved compared to when using microphone signals as reference signals. This can be explained by considering the impact of the SIR and SNR improvements on the correlation difference. Fig. 3 c, d depicts the correlation difference, averaged across trials and participants (noted that these average correlation coefficients are not directly used for decoding). It should be realized that although the obtained correlation differences are quite small, these differences are still statistically significant for making AAD decisions. It can be observed that the largest correlation difference is obtained when using anechoic signals as reference signals, which is related to the fact that these signals do not contain noise and interference. Comparing the correlation difference with the SIR and SNR improvements in Fig. 2 a, b, it can be observed that the correlation difference for all steering angles is considerably improved compared to when using microphone signals as reference signals.

In conclusion, these results show that using steerable superdirective beamformer for decoding auditory attention improves the decoding performance compared to when using the microphone signals as reference signals.

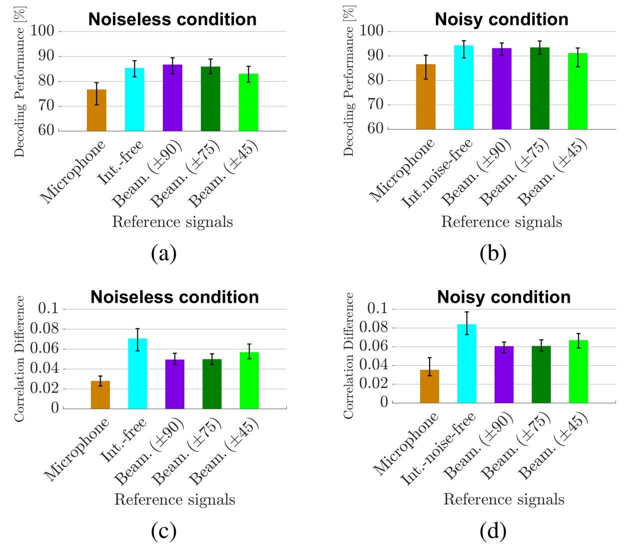

Fig. 3. Decoding performance (a) and (b) and correlation difference (c) and (d) in the noiseless and noisy conditions obtained by using microphone signals, interference-free signals, interference-noise-free signals, and the superdirective beamformer steered at different angles. Error bars represent the 95% bootstrap confidence interval.

5. ACKNOWLEDGMENT

The authors would like to thank Mareike Daeglau and Bojana Mirkovic for the EEG data collection.

6. REFERENCES

[1] E. C. Cherry, "Some experiments on the recognition of speech, with one and with 2 ears," *Journal of the Acoustical Society of America*, vol. 25, no. 5, pp. 975–979, 1953.

[2] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, *Handbook on Array Processing and Sensor Networks (S. Haykin, K. J. Ray Liu, eds.)*. Wiley, 2010, ch. Acoustic beamforming for hearing aid applications, pp. 269–302.

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

[4] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A consolidated perspective on multimicrophone speech enhancement and source separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 4, pp. 692–730, 2017.

[5] 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.

[6] A. Aroudi, B. Mirkovic, M. De Vos, and S. Doclo, "Influence of noisy reference signals on selective attention decoding," in *Proc. Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy*, 2015.

[7] A. Aroudi, B. Mirkovic, M. De Vos, and S. Doclo, "Auditory attention decoding with EEG recordings using noisy acoustic reference signals," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, Mar. 2016, pp. 694–698.

[8] W. Biesmans, N. Das, T. Francart, and A. Bertrand, "Auditory-inspired speech envelope extraction methods for improved EEG-Based auditory attention detection in a cocktail party scenario," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 5, pp. 402–412, May 2017.

[9] B. Mirkovic, M. G. Bleichner, M. De Vos, and S. Debener, "Target speaker detection with concealed EEG around the ear," *Frontiers in Neuroscience*, vol. 10, p. 349, 2016.

[10] R. Zink, A. Baptist, A. Bertrand, S. Van Huffel, and M. De Vos, "Online detection of auditory attention in a neurofeedback

[2]The difference in decoding performance between the noiseless and the noisy condition is due to the different speech materials used as stimuli in these conditions.

application," in *Proc. 8th International Workshop on Biosignal Interpretation*, Osaka, Japan, Nov. 2016, pp. 1–4.

[11] S. Van Eyndhoven, T. Francart, and A. Bertrand, "EEG-informed attended speaker extraction from recorded speech mixtures with application in neuro-steered hearing prostheses," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 5, pp. 1045–1056, 2017.

[12] S. Akram, J. Z. Simon, and B. Babadi, "Dynamic estimation of the auditory temporal response function from MEG in competing-speaker environments," *IEEE Transactions on Biomedical Engineering*, no. 99, p. 1, 2016.

[13] N. Das, S. Van Eyndhoven, T. Francart, and A. Bertrand, "EEG-based attention-driven speech enhancement for noisy speech mixtures using N-fold multi-channel Wiener filters," in *Proc. European Signal Processing Conference (EUSIPCO)*, Kos, Greece, Aug. 2017.

[14] A. Aroudi and S. Doclo, "EEG-based auditory attention decoding: Impact of reverberation, noise and interference reduction," in *Proc. IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Banff, Canada, 2017.

[15] ——, "EEG-based auditory attention decoding using unprocessed binaural signals in reverberant and noisy conditions," in *Proc. Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jeju, South Korea, 2017, pp. 484–488.

[16] J. O'Sullivan, Z. Chen, J. Herrero, G. M. McKhann, S. A. Sheth, A. D. Mehta, and N. Mesgarani, "Neural decoding of attentional selection in multi-speaker environments without access to clean sources," *Journal of Neural Engineering*, vol. 14, no. 5, 2017.

[17] L. Fiedler, M. Woestmann, C. Graversen, A. Brandmeyer, T. Lunner, and J. Obleser, "Single-channel in-Ear-EEG detects the focus of auditory attention to concurrent tone streams and mixed speech," *Journal of Neural Engineering*, vol. 14, no. 3, p. 036020, 2017.

[18] S. A. Fuglsang, T. Dau, and J. Hjortkjær, "Noise-robust cortical tracking of attended speech in real-world acoustic scenes," *NeuroImage*, Apr. 2017.

[19] S. Miran, S. Akram, A. Sheikhattar, J. Z. Simon, T. Zhang, and B. Babadi, "Real-time tracking of selective auditory attention from M/EEG: A bayesian filtering approach," *bioRxiv*, Jan. 2017.

[20] J. Bitzer and K. U. Simmer, "Superdirective microphone arrays," in *Microphone arrays*. Springer, 2001, pp. 19–38.

[21] D. Marquardt and S. Doclo, "Performance comparison of bilateral and binaural MVDR-based noise reduction algorithms in the presence of DOA estimation errors," in *Proc. ITG Symposium on Speech Communication*, 5-7 Oct. 2016, pp. 1–5.

[22] 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, p. 46007, 2015.

[23] H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, p. 6, 2009.

[24] E. Habets, I. Cohen, and S. Gannot, "Generating nonstationary multisensor signals under a spatial coherence constraint," *Journal of the Acoustical Society of America*, vol. 124, no. 5, pp. 2911–2917, Nov. 2008.