EVALUATION OF JOINT AUDITORY ATTENTION DECODING AND ADAPTIVE BINAURAL BEAMFORMING APPROACH FOR HEARING DEVICES WITH ATTENTION SWITCHING

Wenqiang Pu[1]*, Peng Zan*[2]*, Jinjun Xiao*[3]*, Tao Zhang*[3]*, Zhi-Quan Luo*[1]

[1] Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong, Shenzhen, China
[2] University of Maryland, College Park, Maryland, USA
[3] Starkey Hearing Technologies, Eden Prairie, Minnesota, USA

ABSTRACT

Beamforming is a common technique used to improve speech intelligibility and listening comfort of hearing aids users in a noisy environment. Traditional hearing aids beamforming algorithms require the a priori knowledge of the auditory of the listener, which may not be available in real applications. Recent advances in electroencephalography (EEG) offer a potential non-invasive solution to this problem. The listener's auditory is derived from the EEG signals through auditory decoding algorithms and can be used as an input to the beamforming algorithms. In [1], a joint auditory decoding and adaptive beamforming algorithm framework by correlating the envelope of beamforming output and the EEG signal was proposed to improve the beamformer's robustness against decoding error. Consistent performance improvement was demonstrated on an EEG database recorded on listeners with fixed . In this study, we present the evaluation results of this joint formulation on a new EEG dataset collected on subjects with dynamic switch. We demonstrate not only the joint framework's performance improvement against decoding errors, but also its ability to capture listener's dynamic switch.

Index Terms— EEG signals, auditory , microphone array signal processing, acoustic beamforming

1. INTRODUCTION

In modern hearing aids, speech intelligibility and listening comfort can be significantly improved by exploiting the spatial diversity provided by the microphone array [2, 3]. In this regard, various binaural beamforming techniques have been proposed in the past decades [4–6]. However, these beamforming techniques usually require the *a priori* knowledge of the auditory of listener, which may not be easily obtained in a real world environment with multiple talkers. Recent technology advance in electroencephalography (EEG) signal processing offers a potential non-invasive solution for tracking listener's auditory in a complex environment, such as a cocktail party scenario [7]. Based on the collected EEG signals from a scalp EEG system, many computational models [8–10] have been proposed and shown a reliable auditory decoding (AAD) performance in a multi-talker environment [11–17].

Several latest studies have started incorporating the listener's auditory inferred from EEG signals as an input to speech enhancement beamforming algorithms [16, 18]. However, directly utilizing the AAD results followed by beamforming algorithms requires additional source separation procedure, which itself is a very challenging problem in a multi-talker environment. Recently a joint optimization approach combining both AAD and beamforming has been proposed [1]. This approach aims to balance auditory alignment, target speech distortion, noise and interference suppression, and this joint approach eliminates the need to separate speech or its envelope of each talker. In [1], this joint approach was evaluated on an EEG database recorded on listeners without switch. It was shown that the joint approach can achieve significantly reduced temporal variation while maintaining a similar average beamforming performance when compared to the separate decoding and beamforming approach. However, the behavior of this approach in the case of switch remains to be explored.

In this study, we present the evaluation results of the joint approach [1] on a newly collected EEG dataset under the switch conditions. The EEG signals are collected in a competing-talker, noisy and reverberant environment where the listener dynamically switch between talkers. Results confirm that the joint approach is able to track switch in real time. Also, we use intelligibility-weighted signal-to-interference ratio improvement (IW-SIRI) and intelligibility-weighted spectral distortion (IW-SD) as performance metrics [19] to measure interference suppression and target speech distortion respectively.

2. PROBLEM FORMULATION

In this section, we briefly review the recently proposed joint approach [1]. Consider a noisy multi-talker scenario with K talkers, a listener equipped with a pair of binaural hearing aids with M microphones. One talker the listener attempts to listen to is referred as the *attended* talker, and the other $K-1$ talkers are referred as *unattended* (interfering) talkers. In terms of the short-time Fourier transform (STFT), microphone signals in the time-frequency domain can be expressed as $\boldsymbol{y}(\ell,\omega) = \sum_{k=1}^{K} \boldsymbol{h}_k(\omega) s_k(\ell,\omega) + \boldsymbol{n}(\ell,\omega)$, where $\boldsymbol{y}(\ell,\omega)$ is the microphone signal at frame ℓ and frequency band ω ($\omega = 1, 2, \cdots, \Omega$); $\boldsymbol{h}_k(\omega)$ is the acoustic transfer function (ATF) [2] between the k-th speech source and microphones and $s_k(\ell,\omega)$ is the corresponding speech signal in time-frequency domain; and $\boldsymbol{n}(\ell,\omega)$ is the background noise.

The EEG signals of the listener are also recorded which will be used to guide the beamformer design. By exploiting the information of the listener from the recorded EEG signals, our goal is to design a so-called beamformer $\boldsymbol{w}(\omega) \in \mathbb{C}^M$ for each frequency band ω, such that the final beamforming output $z(\ell,\omega) = \boldsymbol{w}^H(\omega)\boldsymbol{y}(\ell,\omega)$ preserving the attended talker's speech signal and suppressing other interfering speech signals and noise. The illustration of the considered EEG-assisted beamforming system diagram is shown in Fig.

This work is partially supported by the leading talents of Guangdong province Program (No. 00201501), the National Natural Science Foundation of China (No. 61731018), the Development and Reform Commission of Shenzhen Municipality, and the Shenzhen Fundamental Research Fund (No. KQTD201503311441545).

1. Notice that for adaptive beamforming, beamformer $\boldsymbol{w}(\omega)$ will change with time. Here for presentation simplicity, we first consider a time period that $\boldsymbol{w}(\omega)$ does not change and the adaptive rule for updating $\boldsymbol{w}(\omega)$ is presented in Algorithm 1.

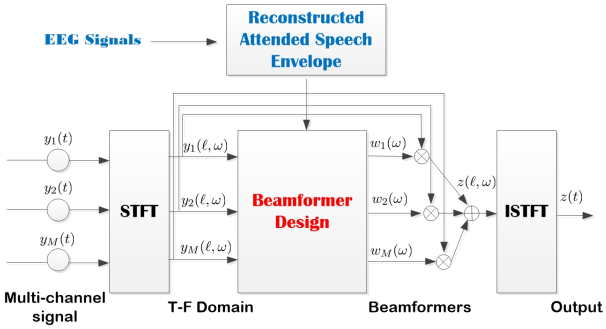


Fig. 1. Illustration of the EEG assisted beamforming system.

The key issue for the above-mentioned EEG-assisted beamforming system is the way of incorporating EEG signals. Recent advances in EEG signal processing and its applications to AAD have shown that low-frequency EEG signals (1-12Hz) have stronger correlation with speech envelope of attended speech than the unattended speech [8]. Based on this experimental observation, several AAD models [10,20,21] were proposed to reconstruct the attended speech envelope from EEG signals. In this regard, a potential way for designing the joint AAD and beamforming system is to establish connections between EEG signals and microphone signals through the attended speech envelope.

Ideally, suppose the beamformer preserves the attended speech and suppresses the noise and interferences; then the envelope of beamforming output approximates the attended speech envelope, which should have a strong correlation with the reconstructed attended speech envelope. This leads to one of the criteria for designing the beamformer by maximizing the Pearson correlation between the reconstructed speech envelope (from EEG signals) and the envelope of beamforming output. By balancing the noise reduction and assignment, the joint AAD and beamforming formulation is given as

$$\min_{\{\boldsymbol{w}(\omega)\},\boldsymbol{\alpha}} \underbrace{\sum_{\omega=1}^{\Omega} \boldsymbol{w}(\omega)^H \boldsymbol{R}(\omega)\boldsymbol{w}(\omega)}_{\text{Noise Reduction}} \underbrace{-\mu\kappa(\{\boldsymbol{w}(\omega)\})}_{\text{Attention Assignment}} \underbrace{-\gamma\|\boldsymbol{\alpha}\|^2}_{\text{Sparsity}}$$

$$\text{s.t.} \quad \boldsymbol{w}(\omega)^H \boldsymbol{h}_k(\omega) = \alpha_k, \ \forall k, \omega, \tag{1a}$$

$$\mathbf{1}^T \boldsymbol{\alpha} = 1, \ \alpha_k \geq 0, \ \forall k. \tag{1b}$$

In (1), $\boldsymbol{R}(\omega)$ is the correlation matrix of noise at frequency band ω, which is estimated from noise only time period, $\kappa(\{\boldsymbol{w}(\omega)\})$ is a smooth nonlinear function w.r.t. beamformer across all frequency bands, which represents the Pearson correlation between the envelope of beamforming output and the reconstructed attended speech envelope (from EEG signal). The detailed mathematical expression of $\kappa(\cdot)$ can be found in [1] and we omit here due to space limitation. Notice that there are two types of optimization variables in (1), one is beamforming coefficients $\{\boldsymbol{w}(\omega)\}$ and the other is coefficients $\{\alpha_k\}$. These variables are coupled in linear constraints (1a), which implies each talker's speech signal at beamforming output is restricted to be scaled by the corresponding α_k. $\mu, \gamma \geq 0$ are user-defined trade-off parameters for balancing noise reduction and auditory assignment, the term $-\gamma\|\boldsymbol{\alpha}\|$ together with constraint (1b) enforces a sparse regularization for $\boldsymbol{\alpha}$. We also remark that ATFs $\{\boldsymbol{h}_k(\omega)\}$ in (1a) can be replaced by relative transfer functions (RTFs), which implies the beamforming output refers to one reference microphone.

Problem (1) is a smooth nonconvex optimization problem, i.e., $\kappa(\{\boldsymbol{w}(\omega)\})$ and $-\gamma\|\boldsymbol{\alpha}\|$ are nonconvex. By exploiting the separable structure property of constraints in (1a), a computationally efficient algorithm based on gradient projection method (GPM) is proposed to solve it [1]. To adaptively update beamformer according the latest EEG signals in real time, an adaptive implementation scheme is also presented in [1]. The adaptive scheme is based on a sliding window of fixed length, i.e., 20 seconds, and EEG signals within the sliding window are used for updating the beamformer. Detailed steps of the GPM based adaptive beamforming scheme is given in Algorithm 1.

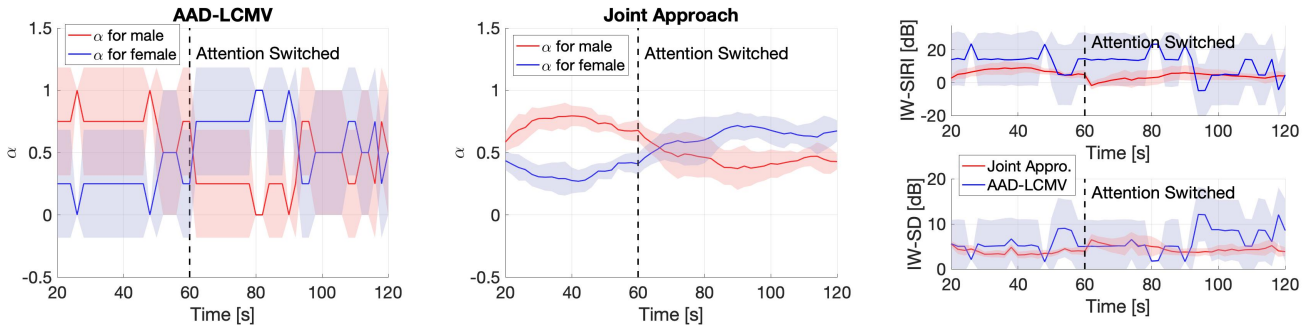
Algorithm 1 GPM Based Adaptive Beamforming Scheme

1: **for** Sliding window index $t = 0, 1, \ldots,$ **do**
2: Reconstruct attended speech envelope from EEG signals;
3: Update the objective function of problem (1);
4: Specify the previous sliding window result as initial point;
5: Fixed number of GPM iterations;
6: Update $\{\boldsymbol{w}(\omega)\}$ and $\boldsymbol{\alpha}$.
7: **end for**

3. EEG DATA COLLECTION

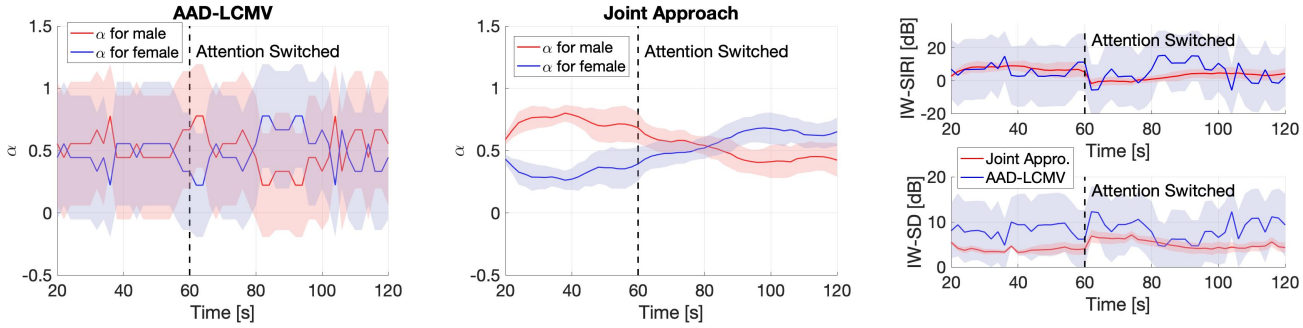
Seven listeners with normal hearing participated in the experiment and all signed a written consent form before the experiment. Auditory responses were recorded with a 64-channel Brain Vision EEG system at a sampling frequency of 2500 Hz. The recorded signal was filtered online by a high-pass filter with cutoff frequency of 0.1 Hz. Channel TP10 was set to be reference channel, and channels O1 and O2 were adjusted to record vertical and horizontal eye movements. The TP10 channel is positioned behind right ear, and channels O1 and O2 are positioned on the back side of the head. Recordings were done in a sound-treated and semi-electrically shielded sound booth. Same as in [1], a set of binaural audio stimuli were generated using a set of ATFs in a simulated noisy and reverberant room (0.6s reverberation time). The background babble noise was generated using sixteen loudspeakers distributed equally on the circle 2 meters away from the subject. Each hearing aid had 2 microphones with 7.5mm spacing. Auditory stimuli at a loud but comfortable level were presented by Presentation software (Neurobehavioral Systems Inc., Berkeley, CA, US) and delivered to the subject's ear through a set of Etymotic ER-3A insert earphones, and such a system has been shown to eliminate stimulus transduction artifact by grounded shielding of the electrical apparatus.

The experiment included two sets of experiments: the clean speech and the switch experiments. We included clean speech experiment to check if EEG recordings had robust auditory response. Specifically, DSS was used to extract the consistent response component, and the topographical map of the first DSS component was examined empirically to make sure that it matched typical auditory response topography. The stimuli included four non-overlapping audio segments, each of 2 min long. Two of the segments were narrated by male talker and the other two by female. For clean speech condition, the first audio segment narrated by male talker was presented for three times, simulated by ATFs to be a source from left. Then a second audio segment narrated by female was presented for three times, which was simulated to be a source from right. Both sources were 1 m away from the listener. A



(a) AAD-LCMV (b) Joint Approach (c) IW-SIRI and IW-SD

Fig. 2. Performance for trials with average accuracy 80%. (Attention switched from male to female at 60s)



(a) AAD-LCMV (b) Joint Approach (c) IW-SIRI and IW-SD

Fig. 3. Performance for trials with average accuracy 63%. (Attention switched from male to female at 60s)

double-choice detailed question was visually presented to subjects after each trial, and the subjects indicated their choice by pressing either left or right arrow on the keyboard. The other two segments was used to construct stimuli for switch condition. An audio mixture was created by mixing the audio segment by male, simulated to be the left source, and the audio segment by female simulated to be the right source. Then a babble noise at a sound level 5 dB below the speech signal was added to the mixture to create a cocktail party scenario. During each trial, subjects were instructed to focus the male talker for the first 60 s, switch their to the female talker after a 5-s visual instruction on the screen, and remain their for the rest of the trial. Three trials were presented for each subject. After each trial, in the same way as described in clean speech condition, subjects answered two comprehensive questions related to the segments they were instructed to pay to, as a way to keep them motivated to attend to the target talker.

4. EVALUATION RESULTS

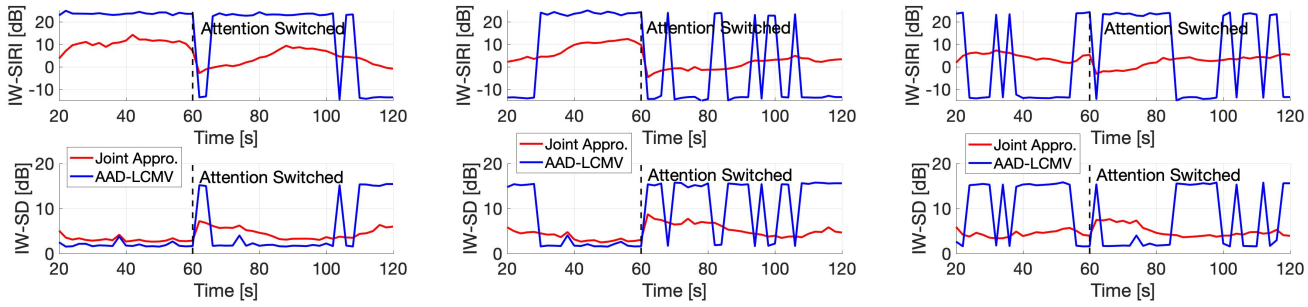
In this section, we present the evaluation results of switch experiment. The EEG signals collected under the babble noise condition in Section 3 are used in the evaluation. As a baseline approach for performance comparison purpose, we use auditory decoding followed by a linearly constrained minimum variance (LCMV) beamformer, or AAD-LCMV for brevity [16]. This approach firstly applies two different LCMV beamformers to separate the two speech signals and then decode the by comparing the Pearson correlations of the separated speech signals with respect to the reconstructed speech envelope. The decoded is finally used to perform a final LCMV beamforming to produce the output signal and we set the linear coefficients of the LCMV for the decoded attended and unattended talkers to be 1 and 0 respectively. Since the goal of evaluation is to study the behavior difference between the 'hard' (the AAD-LCMV approach)and 'soft' (the joint approach) approaches, we do not include any other smooth transition procedure [22] for the AAD-LCMV.

4.1. EEG Data Pre-processing and Decoder Training

The EEG recordings are firstly down-sampled to 1000 Hz, and processed by a denoising source separation (DSS) [23] approach to reduce noise and extract auditory components. A bias function defined by filtering signals into a frequency band of interest (2-12 Hz) and averaging over epochs is used to compute stimulus evoked response. Based on DSS analysis of magnetoencephalographic (MEG) recordings, the auditory response components are mostly contained in the first 6 DSS components and therefore we analyzed the first 6 components in our experiment [24, 25]. During the experiment, a same stimulus, e.g., a mixture of speech segments, was presented for 3 times, which we denoted as 3 trials. EEG time courses responding to the 3 trials of stimulus are averaged. The Hilbert envelope of clean speech segment is band-pass filtered to 1-12 Hz, and downsampled to 50 Hz. The boosting algorithm is used to estimate the decoder which is a linear mapping from response to speech envelope [20, 21]. For each trial, the first 6 DSS components are chosen to train the decoder by a 10-fold cross validation scheme.

4.2. Beamforming Setting

In the beamforming stage, a binaural hearing aid with two microphones on each side is considered. Audio signals at microphones are sampled at a 16 kHz sampling frequency and a 1024-point FFT



(a) Trial 1. (b) Trial 2. (c) Trial 3.

Fig. 4. Performance on representative trials (Trial 1: almost being tracked; Trials 2 and 3: not being well tracked).

with 50% overlap is used in STFT (Hann window). Correlation matrices $\{\mathbf{R}(\omega)\}$ are estimated by sample averaging from a 5-second noise only time period. Different from settings in [1], the sampling rate of the reconstructed speech envelope is adjusted from 20 Hz to 50 Hz, which achieves a smaller decoding error in the decoder training stage (Section 4.1). To match the sampling rate of the reconstructed speech envelope (50 Hz), beamforming output is down sampled from 16 kHz to 50 Hz to represent function $\kappa(\{\boldsymbol{w}(\omega)\})$. Both the joint and AAD-LCMV approaches use the relative transfer functions in equality constraints. To make the evaluation being consistent over all trials and avoid the relative impact of reference microphone selection, we present results that reference microphone is on the unattended talker side. EEG signals with a time window of length 20 seconds is used to assist the beamformer design. In each adaption, the time window is shifted every 2 seconds, followed by 10 iterations of GPM for updating $\{\boldsymbol{w}(\omega)\}$ and $\boldsymbol{\alpha}$. The parameters μ is fixed as $\mu = 100 \sum_\omega \text{trace}(\mathbf{R}(\omega))$ and the initial value of $\boldsymbol{\alpha}$ is set as $\alpha_1 = \alpha_2 = 0.5$. Since the two approaches perform similarly if γ is sufficiently large [1], we consider an extreme case that $\gamma = 0$ (no sparsity regularization) to study the their behavior differences. Further, the Armijo rule [26] is used in GPM per iteration.

4.3. Results

To study the behavior of the two approaches of tracking the subjects' switch under different EEG quality conditions, we select 4 trials with high decoding accuracy (80% average decoding accuracy across trials) and 8 trials with low decoding accuracy (63% average decoding accuracy across trials).

The average parameter α_k across time is compared in Figs. 2(a), 2(b), 3(a), and 3(b), where α_k of AAD-LCMV corresponds to the linear coefficients (0 or 1) used LCMV. The solid lines in figures correspond to the mean values and the light regions represent the standard deviation (which may exceed 0 ro 1) across trials. For the joint approach, the switch is captured in both high and low decoding accuracy cases, but with slightly longer time delay when compared to the AAD-LCMV approach. Comparing the two approaches in high and low decoding accuracy cases, the joint approach shows robustness against raw AAD error and achieves much smaller variance of α_k due to the joint optimization. We also remark that the switch is quickly captured in the selected 4 high accuracy trials (as shown in Figs. 2(a) and 2(b)) and the joint approach shows a slight time delay. We then study the beamforming performance of the two approaches in Figs. 2(c) and 3(c), where the IW-SIRI and IW-SD are calculated for the latest 2 seconds beamforming output. Same as in [1] on EEG database with fixed , the joint approach achieves smaller variation among segments. Further, the AAD-LCMV approach achieves a little larger average IW-SIRI than the joint approach and similar IW-SD in high accuracy case. But in low accuracy case, both approaches have similar average IW-SIRI and the joint approach significantly reduced speech distortion for the attended talker.

Finally, to understand the particular beamforming behavior of the two approaches, we also present IW-SIRI and IW-SD of 3 representative recordings in Fig. 4. When the decoding error rarely happens and the switch has been quickly decoded (Trial 1 in Fig. 4), the joint approach consistently enhance the attended speech before and after the switch. In the situation when the decoding error often happens (Trials 2 and 3 in Fig. 4), the joint approach performs soft enhancement of the attended speech. For the AAD-LCMV approach, no matter the decoding error is small or not, it is very susceptible to decoding error and produces large error in IW-SIRI and IW-SD.

In short, the joint approach seeks for a balance among robustness against the decoding error and average performance. When the decoding error rarely happens, the joint approach requires a little longer time delay for tracking switch and slightly lose the average IW-SIRI performance. However, in low AAD accuracy case which often happens in practice, the two approaches have similar time delay for tracking switch and average IW-SIRI, but the joint approach significantly reduces speech distortion for the attended talker. Further, no matter the decoding accuracy high or low, the joint formulation (soft decoding) has smoothed change of speech sources which is very important for listening comfort.

5. CONCLUSION

In this study, we evaluated the joint auditory decoding and adaptive binaural beamforming approach proposed in [1] on a EEG database newly collected on listeners with switch in a noisy and reverberated environment. Evaluation results confirm that the joint approach is able to track the switch in real time. Compared with the AAD-LCMV approach, the joint approach has shown robustness against raw decoding errors and reduced speech distortion for the attended talker. As part of the future work, we plan to extend the algorithm to the complete binaural beamforming application, where two beamformers for two ears will be simultaneously designed with the assistance of EEG signals. Design considerations in binaural beamforming including the binaural cue preservation [2] will be included in the optimization model.

6. REFERENCES

[1] W. Pu, J. Xiao, T. Zhang, and Z. Luo, "A joint auditory attention decoding and adaptive binaural beamforming algorithm for hearing devices," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 311–315.

[2] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, "Acoustic beamforming for hearing aid applications," *Handbook on array processing and sensor networks*, pp. 269–302, 2008.

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

[4] D. Marquardt, V. Hohmann, and S. Doclo, "Interaural coherence preservation in multi-channel wiener filtering-based noise reduction for binaural hearing aids," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 23, no. 12, pp. 2162–2176, 2015.

[5] E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, "Theoretical analysis of binaural transfer function mvdr beamformers with interference cue preservation constraints," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2449–2464, Dec 2015.

[6] W. Pu, J. Xiao, T. Zhang, and Z. Luo, "A penalized inequality-constrained minimum variance beamformer with applications in hearing aids," in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct 2017, pp. 175–179.

[7] S. Haykin and Z. Chen, "The cocktail party problem," *Neural computation*, vol. 17, no. 9, pp. 1875–1902, 2005.

[8] 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*, vol. 25, no. 7, pp. 1697–1706, 2015.

[9] A. Aroudi, B. Mirkovic, M. De Vos, and S. Doclo, "Impact of different acoustic components on eeg-based auditory attention decoding in noisy and reverberant conditions," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 4, pp. 652–663, April 2019.

[10] S. Miran, S. Akram, A. Sheikhattar, J. Simon, T. Zhang, and B. Babadi, "Real-time tracking of selective auditory attention from m/eeg: A bayesian filtering approach," *Frontiers in neuroscience*, vol. 12, 2018.

[11] A. J. Power, J. J. Foxe, E. Forde, R. B. Reilly, and E. C. Lalor, "At what time is the cocktail party? a late locus of selective attention to natural speech," *European Journal of Neuroscience*, vol. 35, no. 9, pp. 1497–1503, 2012.

[12] 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. 046007, 2015.

[13] J. O'Sullivan, Z. Chen, J. Herrero, G. 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 5, pp. 056001, 2017.

[14] N. Das, A. Bertrand, and T. Francart, "EEG-based auditory attention detection: boundary conditions for background noise and speaker positions," *Journal of Neural Engineering*, vol. 15, no. 6, pp. 066017, 2018.

[15] 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, 2017.

[16] A. Aroudi and S. Doclo, "Cognitive-driven binaural lcmv beamformer using eeg-based auditory attention decoding," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 406–410.

[17] S. A. Fuglsang, T. Dau, and J. Hjortkjær, "Noise-robust cortical tracking of attended speech in real-world acoustic scenes," *Neuroimage*, vol. 156, pp. 435–444, 2017.

[18] 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.

[19] A. Spriet, M. Moonen, and J. Wouters, "Robustness analysis of multichannel wiener filtering and generalized sidelobe cancellation for multimicrophone noise reduction in hearing aid applications," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 4, pp. 487–503, July 2005.

[20] S. V. David, N. Mesgarani, and S. A. Shamma, "Estimating sparse spectro-temporal receptive fields with natural stimuli.," *Network*, vol. 18 3, pp. 191–212, 2007.

[21] N. Ding and J. Z. Simon, "Neural coding of continuous speech in auditory cortex during monaural and dichotic listening," *Journal of Neurophysiology*, vol. 107, no. 1, pp. 78–89, Jan 2012.

[22] S. Geirnaert, T. Francart, and A. Bertrand, "An interpretable performance metric for auditory attention decoding algorithms in a context of neuro-steered gain control," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 1, pp. 307–317, Jan 2020.

[23] A. Cheveigné and J. Z. Simon, "Denoising based on spatial filtering," *Journal of Neuroscience Methods*, vol. 171, no. 2, pp. 331 – 339, 2008.

[24] N. Ding and J. Z. Simon, "Adaptive temporal encoding leads to a background-insensitive cortical representation of speech," *J. Neurosci.; Journal of Neuroscience*, vol. 33, no. 13, pp. 5728–5735, 2013.

[25] N. Ding and J. Z. Simon, "Emergence of neural encoding of auditory objects while listening to competing speakers," *PNAS; Proceedings of the National Academy of Sciences*, vol. 109, no. 29, pp. 11854–11859, 2012.

[26] D. P. Bertsekas, *Nonlinear programming*, Athena scientific Belmont, 1999.