COGNITIVE-DRIVEN BINAURAL LCMV BEAMFORMER USING EEG-BASED AUDITORY ATTENTION DECODING

Ali Aroudi, Simon Doclo

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

ABSTRACT

Identifying the target speaker in hearing aid applications is an essential ingredient to improve speech intelligibility. 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. Aiming at enhancing the target speaker and suppressing the interfering speaker and ambient noise, in this paper we propose a cognitive-driven speech enhancement system, consisting of a direction-of-arrival (DOA) estimator, steerable beamformers and AAD. To preserve the spatial impression of the acoustic scene, which is important when intending to switch attention between speakers, the proposed system only partially suppresses the interfering speaker. The speech enhancement performance of the proposed system is evaluated in terms of the signal-to-interference-plus-noise ratio (SINR) improvement in anechoic and reverberant conditions. The experimental results show that the proposed system can obtain a considerably large SINR improvement (between 3.1 dB and 7.5 dB) in both conditions.

Index Terms— auditory attention decoding, steerable LCMV beamformer, speech enhancement, EEG signal, brain computer interface

1. INTRODUCTION

During the last decades significant advances have been made in multi-microphone speech enhancement algorithms for hearing aids. Although algorithms are available to reduce background noise or to perform source separation in multi-talker scenarios [1, 2], their performance in improving speech intelligibility depends on correctly identifying the target speaker to be enhanced. In hearing aid applications the target speaker is typically assumed to be located in front of the user or is assumed to be the loudest speaker. As in real-world conditions these assumptions are often violated, the performance of speech enhancement algorithms may substantially decrease.

Recently, a least-squares-based auditory attention decoding (AAD) method has been proposed to identify the attended speaker from single-trial EEG recordings in an acoustic scenario with two competing speakers [3–8]. This method aims at reconstructing the attended speech envelope from the EEG recordings using a trained spatio-temporal filter. In the training step, the clean speech signal of the attended speaker is used to train the spatio-temporal filter 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 used as reference signals. In practice, only the hearing aid microphone signals, containing reverberation, background noise and interference, are obviously available. In [5, 9] it has been shown that using the microphone signals as reference signals for decoding is feasible, however, results in a significantly decreased decoding performance. Aiming at generating appropriate reference signals from the microphone signals, noise reduction algorithms using multi-channel Wiener filtering [10, 11], a source separation algorithm using deep neural networks [12], and a steerable superdirective beamformer [13] were proposed. In [10] a neuro-steered multi-channel Wiener filter (MWF) was proposed, which enhances the attended speaker and strongly suppresses the interfering unattended speaker (in an anechoic condition). While strongly suppressing the unattended speaker is desired to improve intelligibility, it may deprive the user to switch attention. In addition, the MWF in [10] changes the spatial impression of the acoustic scene since all sources are perceived as coming from the direction of the attended speaker, which may lead to a confusion between acoustical and visual information.

Aiming at enhancing the attended speaker and controlling the suppression of the unattended speaker while preserving the spatial impression of the acoustic scene, in this paper we propose a cognitive-driven binaural LCMV beamformer system (see Fig. 1). First, the DOA of both speakers is estimated from the microphone signals. Based on the estimated DOA of the speakers, two LCMV beamformers generate reference signals for auditory attention decoding. The AAD method then identifies the DOA of the attended and the unattended speaker to steer a binaural LCMV beamformer [14]. To preserve the spatial impression of the acoustic scene, this binaural beamformer only partially suppresses the signal coming from the unattended DOA while preserving the signal coming from the attended DOA.

For an acoustic scenario with two competing speakers and diffuse babble noise, 64-channel EEG responses with 18 participants were recorded for two reverberation conditions (anechoic and moderate reverberation time $T_{60} = 0.5$ s). The experimental results show that the proposed cognitive-driven LCMV beamformer considerably improves the binaural SINR for both conditions. Moreover, the results show that for reverberant condition the binaural SINR improvement is larger when using (oracle) reverberant relative transfer functions (RTFs) instead of (estimated) anechoic RTFs.

2. COGNITIVE-DRIVEN BINAURAL LCMV BEAMFORMER

2.1. Configuration and notation

Consider an acoustic scenario comprising two competing speakers with DOAs θ_1 and θ_2 and background noise in a reverberant environment (see Fig. 1). The clean signal of speaker 1 is denoted as $s_1[n]$, while the clean signal of speaker 2 is denoted as $s_2[n]$, with n the discrete time index. The m-th microphone signal from the left hearing aid can be written as

$$y_{L,m}[n] = x_{1,L,m}[n] + x_{2,L,m}[n] + v_{L,m}[n], \quad (1)$$

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) and Cluster of Excellence 1077 Hearing4all.

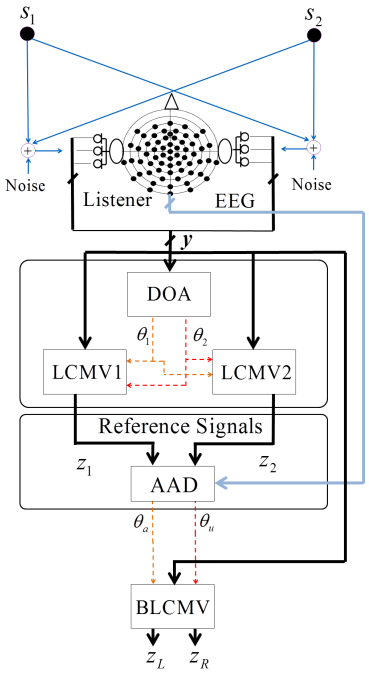


Fig. 1. The block diagram of the proposed cognitive-driven LCMV beamformer system.

where $x_{1,L,m}[n]$ and $x_{2,L,m}[n]$ denote the reverberant speech component in the m-th microphone signal corresponding to speaker 1 and speaker 2, respectively, and $v_{L,m}[n]$ denotes the background noise component. The m-th microphone signal from the right hearing aid $y_{R,m}[n]$ is defined similarly as in (1).

In the short-time Fourier transform (STFT) domain, the stacked microphone signals from the left and the right hearing aid can be written as

$$\boldsymbol{y}(k,l)=[Y_{L,1}(k,l)\ldots Y_{L,M}(k,l)\ Y_{R,1}(k,l)\ldots Y_{R,M}(k,l)]^T, \quad (2)$$

where k denotes the frequency index, l denotes the frame index, and M denotes the number of microphones per hearing aid. For notational conciseness the indices k, l and n will be omitted in the remainder of this paper.

2.2. DOA estimation

To estimate the DOA of multiple speakers from binaural microphone signals, several methods have been proposed [15–17]. In this paper, we will use the DOA estimation algorithm in [15], which estimates the source presence probability for different DOAs using support-vector machine (SVM) classifiers followed by a generalized linear model (GLM). The SVMs are trained using short-term generalized cross-correlation with phase transform (GCC-PHAT) features [18] to distinguish between the presence of a source for a certain direction and the absence for all other directions. The decision value of each SVM is then mapped to a source presence probability for each direction using a GLM. To increase the robustness against background noise, the probabilities are smoothed across time using recursive averaging with a time constant τ. The estimated DOAs of speakers 1 and 2 are denoted as $\hat{\theta}_1$ and $\hat{\theta}_2$, respectively.

2.3. Reference signal generation using LCMV beamformers

To generate appropriate reference signals from microphone signals for AAD, we propose to use two LCMV beamformers on the microphone signals from the left and the right hearing aid. An LCMV beamformer aims at 1) minimizing the output power of the noise component while 2) passing signals arriving from the target angle θ_t without distortion and 3) suppressing signals arriving from the interfering angle θ_i [14]. The corresponding constrained optimization problem is given by

$$\min_{\mathbf{w}} \underbrace{\mathbf{w}^H\boldsymbol{\Phi}_{\mathbf{v}}\mathbf{w}}_{\text{noise output power}} \text{ subject to } \underbrace{\mathbf{w}^H\mathbf{a}(\theta_t)=1}_{\text{target}}, \underbrace{\mathbf{w}^H\mathbf{a}(\theta_i)=0}_{\text{interference}}, \quad (3)$$

with $\boldsymbol{\Phi}_{\mathbf{v}}$ denoting the noise covariance matrix and $\mathbf{a}(\theta_t)$ and $\mathbf{a}(\theta_i)$ denoting the relative transfer function (RTF) vectors [1] corresponding to the target angle θ_t and the interfering angle θ_i, respectively. The LCMV beamformer with target angle θ_t and interfering angle θ_i is given as [14]

$$\mathbf{w}_{LCMV}(\theta_t,\theta_i)=\boldsymbol{\Phi}_{\mathbf{v}}^{-1}\mathbf{C}\left(\mathbf{C}^H\boldsymbol{\Phi}_{\mathbf{v}}^{-1}\mathbf{C}\right)^{-1}\mathbf{b}, \quad (4)$$

with

$$\mathbf{C}=[\mathbf{a}(\theta_t)\ \ \mathbf{a}(\theta_i)],\ \mathbf{b}=[1\ \ 0]^T. \quad (5)$$

We now consider two LCMV beamformers: 1) an LCMV beamformer with steering angles $\theta_t=\hat{\theta}_1$ and $\theta_i=\hat{\theta}_2$ to generate the reference signal for speaker 1 2) an LCMV beamformer with steering angles $\theta_t=\hat{\theta}_2$ and $\theta_i=\hat{\theta}_1$ to generate the reference signal for speaker 2, i.e.,

$$z_1=\text{ISTFT}\left\{\mathbf{w}_{LCMV}^H\left(\theta_t=\hat{\theta}_1,\ \theta_i=\hat{\theta}_2\right)\boldsymbol{y}\right\}, \quad (6)$$

$$z_2=\text{ISTFT}\left\{\mathbf{w}_{LCMV}^H\left(\theta_t=\hat{\theta}_2,\ \theta_i=\hat{\theta}_1\right)\boldsymbol{y}\right\}, \quad (7)$$

where ISTFT denotes the inverse short-time Fourier transform.

2.4. Auditory attention decoding

To decode auditory attention from C-channel EEG recordings $r_c[i]$, with $c=1\ldots C$ and i the sub-sampled time index, it has been proposed in [3] to reconstruct an estimate of the attended speech envelope $\hat{e}_a$ using a trained spatio-temporal filter, i.e.,

$$\hat{e}_a[i]=\sum_{c=1}^{C}\sum_{j=0}^{J-1}g_{c,j}\,r_c[i+j+\Delta], \quad (8)$$

with $g_{c,j}$ the j-th filter coefficient in the c-th channel, J the number of filter coefficients per channel, and Δ modeling the latency of the attentional effect in the EEG responses to acoustic stimuli.

Based on the correlation coefficients between the reconstructed speech envelope $\hat{e}_a[i]$ and the envelopes $e_1[i]$ and $e_2[i]$ of the reference signals z_1 and z_2, i.e.,

$$\rho_1=\rho(e_1[i],\hat{e}_a[i]),\ \rho_2=\rho(e_2[i],\hat{e}_a[i]), \quad (9)$$

it is then decided that the listener attended to speaker 1 if $\rho_1 > \rho_2$ or attended to speaker 2 otherwise. The DOAs of the attended and the unattended speaker are hence identified based on the correlation coefficients, i.e.,

$$\begin{cases} \hat{\theta}_a=\hat{\theta}_1,\ \hat{\theta}_u=\hat{\theta}_2 & \text{if } \rho_1>\rho_2 \\ \hat{\theta}_a=\hat{\theta}_2,\ \hat{\theta}_u=\hat{\theta}_1 & \text{otherwise.} \end{cases} \quad (10)$$

Prior to the decoding step, the spatio-temporal filter $\mathbf{g}$ in (8) needs to be trained. During the training step the attended speaker is assumed to be known and the filter $\mathbf{g}$ is computed by minimizing the least-squares error between the attended speech envelope and the reconstructed envelope.

2.5. Binaural LCMV beamformer

The estimated DOAs in (10) are then used to steer a binaural LCMV beamformer [14], passing the signal from the attended DOA $\hat{\theta}_a$ and suppressing the signal from the unattended DOA $\hat{\theta}_a$. Since we aim at controlling the amount of suppression for the unattended speaker and preserving the spatial impression of the acoustic scene, we consider the same interference suppression factor for the left and the right hearing aid. The binaural LCMV beamformer coefficients for the left hearing aid can be computed as

$$\mathbf{w}_{BLCMV,L}\left(\hat{\theta}_a,\hat{\theta}_u\right)=\mathbf{\Phi}_{\mathbf{v}}^{-1}\mathbf{C}_L\left(\mathbf{C}_L^H\mathbf{\Phi}_{\mathbf{v}}^{-1}\mathbf{C}_L\right)^{-1}\mathbf{b}, \tag{11}$$

with

$$\mathbf{C}_L=\left[\mathbf{a}_L\left(\hat{\theta}_a\right)\ \mathbf{a}_L\left(\hat{\theta}_u\right)\right],\ \mathbf{b}=\begin{bmatrix}1 & \delta\end{bmatrix}^T, \tag{12}$$

where $\mathbf{a}_L\left(\hat{\theta}_a\right)$ and $\mathbf{a}_L\left(\hat{\theta}_u\right)$ denote the RTF vectors of the left hearing aid for the attended DOA and the unattended DOA, respectively, and $\delta > 0$ denotes the interference suppression factor which controls the amount of suppression and the binaural cue preservation of the signals arriving from the unattended DOA [14]. The LCMV beamformer coefficients for the right hearing aid $\mathbf{w}_{BLCMV,R}\left(\hat{\theta}_a,\hat{\theta}_u\right)$ can be computed similarly as in (11). Please note that $\delta = 0$ corresponds to a complete suppression of the unattended speaker (and unpredictable binaural cue distortion due to small RTF estimation errors [14]), while a large δ leads to a small suppression of the unattended speaker.

The output signals of the binaural LCMV beamformer for the left and the right hearing aid can be computed as

$$z_L=\text{ISTFT}\left\{\mathbf{w}_{BLCMV,L}^H\left(\hat{\theta}_a,\hat{\theta}_u\right)\boldsymbol{y}\right\}, \tag{13}$$

$$z_R=\text{ISTFT}\left\{\mathbf{w}_{BLCMV,R}^H\left(\hat{\theta}_a,\hat{\theta}_u\right)\boldsymbol{y}\right\}. \tag{14}$$

These signals can be decomposed as

$$z_L=z_{a,L}+z_{u,L}+z_{v,L}, \tag{15}$$

$$z_R=z_{a,R}+z_{u,R}+z_{v,R}, \tag{16}$$

where $z_{a,L}$ and $z_{a,R}$ denote the output speech component corresponding to the (oracle) attended speaker for the left and the right hearing aid, respectively, and $z_{u,L}$ and $z_{u,R}$ denote the output speech component corresponding to the (oracle) unattended speaker for the left and the right hearing aid, respectively, and $z_{v,L}$ and $z_{v,R}$ denote the output noise component.

3. EXPERIMENTAL SETUP

3.1. Acoustic stimuli

EEG responses were recorded for 18 native German-speaking participants. Two German audio 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., the audio stories, with (non-individualized) binaural impulse responses from [19], and adding diffuse noise, generated according to [20]. The left and the right speaker were simulated at $\theta_1 = -45^\circ$ and $\theta_2 = 45^\circ$. Four acoustic conditions were considered: two conditions with $\text{SNR} = 9.0$ dB and $\text{SNR} = 4.0$ dB, and two reverberant conditions (reverberation time $T_{60} = 0.5$ s with the same SNR). For experimental analysis, the acoustic conditions were grouped together based on reverberation time, resulting in two experimental analysis conditions, i.e., anechoic-noisy and reverberant-noisy. Among all participants, 8 participants were instructed to attend to the left speaker, while 10 participants were instructed to attend to the right speaker. Two participants were excluded from the analysis, one participant due to poor attentional performance and the other one due to a technical hardware problem.

3.2. EEG and AAD setup

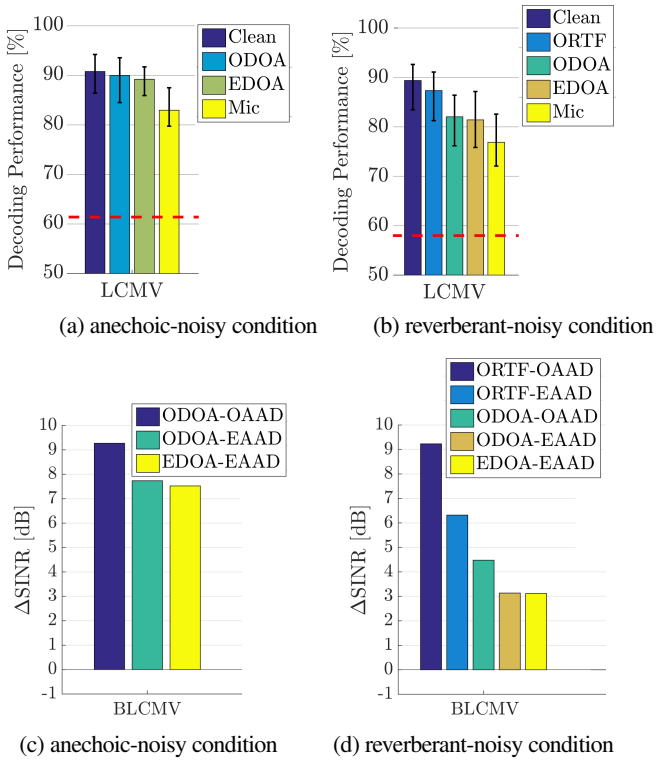
The EEG responses were recorded using $C = 64$ channels at a sampling frequency of 500 Hz, and referenced to the nose electrode. The EEG responses were re-referenced offline to a common average reference, band-pass filtered between 2 Hz and 8 Hz using a third-order Butterworth band-pass filter, and subsequently 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 AAD training and decoding steps (see Section 2.4), 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 of the spatio-temporal filter in (8) were set to fixed values as $\Delta = 8$ and $J = 8$ (corresponding to 125 ms). The decoding performance was computed by averaging the percentage of correctly decoded trials over all considered trials and all participants. Aiming at investigating the impact of AAD errors on the speech enhancement performance of the binaural LCMV beamformer, we considered oracle AAD (OAAD), i.e., $\hat{\theta}_a = \theta_a$ and $\hat{\theta}_u = \theta_u$, and estimated AAD (EAAD) where $\hat{\theta}_a$ and $\hat{\theta}_u$ are determined using (10).

3.3. DOA and LCMV beamformer setup

The hearing aid microphone signals were generated using measured impulse responses for a binaural hearing aid setup mounted on a dummy head from [19], where each hearing aid was equipped with 3 microphones. For the DOA estimation algorithm (see Section 2.2), the SVM classifiers were trained using simulated speech signals generated by convolving TIMIT training data with binaural anechoic Behind-The-Ear (BTE) impulse responses from the same hearing aid setup [19], and adding diffuse noise at SNRs of -20 dB to 20 dB in steps of 10 dB. The GCC-PHAT features were calculated using segment lengths of 10 ms with an overlap of 5 ms. The source presence probabilities were smoothed across time using the time constant $\tau = 1$ s. Aiming at investigating the impact of DOA estimation errors on the AAD performance and the speech enhancement performance, we considered the oracle DOAs (ODOA), i.e., $\hat{\theta}_1 = \theta_1$ and $\hat{\theta}_2 = \theta_2$, and the estimated DOAs (EDOA) where $\hat{\theta}_1$ and $\hat{\theta}_2$ are estimated using [15].

To generate the reference signals using the LCMV beamformer (see Section 2.3), and the output signals of the left and the right hearing aid using binaural LCMV beamformer (see Section 2.5), the hearing aid microphone signals were processed using a weighted overlap-add (WOLA) framework with a frame size of 512 samples and an overlap of 50%. The sampling frequency was 16 kHz. The noise covariance matrix $\mathbf{\Phi}_{\mathbf{v}}$ was calculated using the diffuse noise assumption, i.e., by spatially averaging the auto- and cross-correlations of the anechoic acoustic transfer functions (ATFs) measured with a resolution of 5° [19]. The RTF vectors $\mathbf{a}(\theta_t)$ for the LCMV beamformers (see section 2.3) and the RTF vectors $\mathbf{a}_L(\theta_a)$ and $\mathbf{a}_R(\theta_a)$ for the binaural LCMV beamformer (see section 2.5) were calculated based on the (anechoic or reverberant) ATFs from [19] for the angle θ. Aiming at exploring the impact of different RTF vectors on the AAD performance and the speech enhancement performance, we considered anechoic RTF vectors, using either the oracle DOAs (ODOA) or the estimated DOAs (EDOA), for the anechoic-noisy and the reverberant-noisy condition, and oracle reverberant RTF vectors (ORTF) for the reverberant-noisy condition.

Aiming at preserving the spatial impression of the acoustic scene, we set the interference suppression factor $\delta = 0.2$ for the binaural LCMV



(a) anechoic-noisy condition (b) reverberant-noisy condition

(c) anechoic-noisy condition (d) reverberant-noisy condition

Fig. 2. Average decoding performance and binaural SINR improvement (ΔSINR) for the anechoic-noisy and reverberant-noisy conditions. The red dashed-line represents the upper boundary of the confidence interval corresponding to chance level based on a binomial test at the 5% significance level. The error bars represent the 95% bootstrap confidence interval.

beamformer to only partially suppress the unattended speaker while preserving its binaural cues.

The speech enhancement performance of the binaural LCMV beamformer is evaluated in terms of the binaural signal-to-interference-plus-noise ratio improvement (ΔSINR). The binaural input SINR is defined as

$$\text{SINR}_{in} = 10\log_{10}\frac{\varepsilon\{|x_{a,L,1}|^2\}+\varepsilon\{|x_{a,R,1}|^2\}}{\varepsilon\{|x_{u,L,1}+v_{L,1}|^2\}+\varepsilon\{|x_{u,R,1}+v_{R,1}|^2\}}, \quad (17)$$

with $x_{a,L,1}$ and $x_{a,R,1}$ the (oracle) attended reverberant speech component at the left and the right hearing aid, $x_{u,L,1}$ and $x_{u,R,1}$ the (oracle) unattended reverberant speech component at the left and the right hearing aid, and $\varepsilon\{\cdot\}$ the expectation operator. The binaural output SINR is defined as

$$\text{BSINR}_{out} = 10\log_{10}\frac{\varepsilon\{|z_{a,L}|^2\}+\varepsilon\{|z_{a,R}|^2\}}{\varepsilon\{|z_{u,L}+z_{v,L}|^2\}+\varepsilon\{|z_{u,R}+z_{v,R}|^2\}}. \quad (18)$$

with $z_{a,L}$, $z_{u,L}$, $z_{v,L}$ defined in (15), and $z_{a,R}$, $z_{u,R}$, $z_{v,R}$ defined in (16).

4. EXPERIMENTAL RESULTS

In this section we evaluate the AAD performance and the speech enhancement performance of the proposed cognitive-driven binaural LCMV system for the anechoic-noisy and the reverberant-noisy condition. In Section 4.1 we investigate the impact of DOA estimation errors and different RTF vectors on the AAD performance. In Section 4.2 we investigate the impact of AAD errors, DOA estimation errors and different RTF vectors on the speech enhancement performance.

4.1. Decoding performance

For the anechoic-noisy and the reverberant-noisy condition, Fig. 2a and Fig. 2b present the average decoding performance when using the clean speech signals, the LCMV output signals, or the microphone signals as reference signals for decoding. It can be observed that when using the LCMV output signals as reference signals the decoding performance for both acoustic conditions is improved compared to when using the microphone signals as reference signals. When using either the estimated DOAs (**EDOA**) or the oracle DOAs (**ODOA**), a similar decoding performance for both acoustic conditions is obtained, implying that the AAD performance is robust to DOA estimation errors for the considered acoustic scenario. When using the anechoic RTF vectors (ODOA) for the reverberant-noisy condition, the decoding performance decreases compared to when using the oracle reverberant RTF vectors (**ORTF**) .

4.2. Speech enhancement performance

Fig. 2c and Fig. 2d present the binaural SINR improvement for the anechoic-noisy and the reverberant-noisy condition. It can be observed that a large binaural SINR improvement for all considered cases is obtained. When using the estimated AAD and oracle DOAs for the anechoic-noisy condition (ODOA–**EAAD**), the binaural SINR improvement decreases by about 1.5 dB compared to when using the oracle AAD (ODOA–**OAAD**). Similarly, when using the estimated AAD for the reverberant-noisy condition (either assuming oracle reverberant RTFs, i.e., ORTF–**EAAD**, or oracle DOAs, i.e., ODOA–**EAAD**) the binaural SINR improvement decreases compared to when using the oracle AAD (either ORTF–**OAAD** or ODOA–**OAAD**) showing the sensitivity to AAD errors. For both acoustic conditions, a similar binaural SINR improvement is obtained when using the estimated DOAs (**EDOA**–EAAD) or the oracle DOAs (**ODOA**–EAAD), showing the robustness to DOA estimation errors. When using the oracle anechoic RTF vectors (either **ODOA**–OAAD or **ODOA**–EAAD) for the reverberant-noisy condition, the binaural SINR improvement decreases compared to when using the oracle reverberant RTF vectors (either **ORTF**–OAAD or **ORTF**–EAAD).

5. CONCLUSION

In this paper, we proposed a cognitive-driven binaural LCMV beamformer system enhancing the attended speaker while controlling the amount of suppression for the unattended speaker and preserving the spatial impression of the acoustic scene. The experimental results showed that the proposed system can considerably improve the decoding performance and obtain a large binaural SINR improvement both for anechoic as well as reverberant conditions. In addition, the results showed that the binaural SINR improvement is robust to DOA estimation errors but sensitive to AAD errors.

6. REFERENCES

[1] 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.

[2] 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.

[3] 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.

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

[5] 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," *bioRxiv*, 2018. [Online]. Available: https://www.biorxiv.org/content/early/2018/07/03/360834

[6] 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," *Frontiers in Neuroscience*, vol. 12, p. 262, 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnins.2018.00262

[7] D. D. E. Wong, S. A. Fuglsang, J. Hjortkjr, E. Ceolini, M. Slaney, and A. de Cheveign, "A comparison of regularization methods in forward and backward models for auditory attention decoding," *Frontiers in Neuroscience*, vol. 12, p. 531, 2018. [Online]. Available: https://www.frontiersin.org/article/10.3389/fnins.2018.00531

[8] A. Cheveigné, D. D. Wong, G. M. D. Liberto, J. Hjortkjr, M. Slaney, and E. Lalor, "Decoding the auditory brain with canonical component analysis," *NeuroImage*, vol. 172, pp. 206 – 216, 2018.

[9] A. Aroudi and S. Doclo, "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.

[10] 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.

[11] 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.

[12] 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.

[13] A. Aroudi, D. Marquardt, and S. Doclo, "EEG-based auditory attention decoding using steerable binaural superdirective beamformer," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, Canada, Apr. 2018, pp. 851 – 855.

[14] E. Hadad, S. Doclo, and S. Gannot, "The binaural LCMV beamformer and its performance analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 3, pp. 543–558, March 2016.

[15] H. Kayser and J. Anemueller, "A discriminative learning approach to probabilistic acoustic source localization," in *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, Sep. 2014, pp. 99–103.

[16] S. Braun, W. Zhou, and E. A. P. Habets, "Narrowband direction-of-arrival estimation for binaural hearing aids using relative transfer functions," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct 2015, pp. 1–5.

[17] D. Marquardt and S. Doclo, "Noise power spectral density estimation for binaural noise reduction exploiting direction of arrival estimates," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct 2017, pp. 234–238.

[18] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, Aug. 1976.

[19] 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.

[20] 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.