2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)
Banff Center, Banff, Canada, October 5-8, 2017

EEG-based Auditory Attention Decoding: Impact of Reverberation, Noise and Interference Reduction*

Ali Aroudi, Simon Doclo

***Abstract*—To identify the attended speaker from single-trial EEG recordings in an acoustic scenario with two competing speakers, an auditory attention decoding (AAD) method has recently been proposed. The AAD method requires the clean speech signals of both the attended and the unattended speaker as reference signals for decoding. However, in practice only the binaural signals, containing several undesired acoustic components (reverberation, background noise and interference), and influenced by anechoic head-related transfer functions (HRTFs), are available. To generate appropriate reference signals for decoding from the binaural signals, it is important to understand the impact of these acoustic components on the AAD performance. In this paper, we investigate this impact for decoding several acoustic conditions (anechoic, reverberant, noisy, and reverberant-noisy) by using simulated speech signals in which different acoustic components have been reduced. The experimental results show that for obtaining a good decoding performance the joint suppression of reverberation, background noise and interference as undesired acoustic components is of great importance.**

***Keywords*—auditory attention decoding; noisy and reverberant signal; speech envelope; noise reduction; dereverberation; EEG signal; brain computer interface**

I. Introduction

During the last decades significant advances in acoustic signal processing algorithms have been achieved to improve speech intelligibility for hearing-impaired listeners. Nevertheless understanding speech in complex listening conditions, particularly in multi-talker acoustic scenarios, is still a challenging problem since many acoustic signal processing algorithms need to rely on predefined assumptions about the target speaker to be enhanced. For example, it is typically assumed that the target speaker is in front of the hearing aid user or is the loudest speaker. As such assumptions are mostly violated in real-world conditions, the performance of these algorithms may dramatically decrease. Therefore, identifying the target speaker in hearing aid applications is an essential ingredient to successfully improve speech understanding.

Recently, an auditory attention decoding (AAD) method has been proposed for identifying the attended speaker from single-trial EEG recordings [1]. The AAD method 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 binaural signals, containing several undesired acoustic components (reverberation, background noise and interference), and influenced by head-related transfer functions (HRTFs), are available. In [2], [3] it was shown that the considered AAD method was to some extent robust to (simulated) residual noise at the output of a source separation algorithm, although it should be realized that this noise was not presented to the listeners during the EEG recordings. The feasibility of AAD using unprocessed, i.e. reverberant and noisy, binaural signals as reference signals was shown in [4], although the obtained decoding performance was significantly lower than when using the clean speech signals as reference signals. In this paper we study the case where noisy and reverberant binaural signals are presented to listeners during EEG recordings while using processed binaural signals as reference signals for decoding.

Many acoustic signal processing algorithms are available to reduce background noise, reverberation and interference sources [5], [6]. However, for most algorithms there is typically a trade-off between reducing each of these undesired acoustic components [7]. In order to use the most appropriate acoustic signal processing strategy for generating reference signals, the impact of reducing each undesired acoustic component on the AAD performance needs to be determined. In this paper we address this issue for different acoustic conditions (anechoic, reverberant, noisy, and reverberant-noisy) by using simulated speech signals in which different undesired acoustic components have been reduced.

For an acoustic scenario comprising two competing speakers and diffuse noise at different SNRs and reverberation times, 64-channel EEG responses with 18 participants were recorded. The experimental results show that in order to obtain a sufficient decoding performance the joint suppression of reverberation, background noise and interference as undesired acoustic components is of great importance.

II. Auditory Attention Decoding

In this section the least-squares method used for decoding auditory attention is presented. In Section II-A the different acoustic conditions used for recording EEG responses are defined. In Section II-B the training and evaluation steps are discussed.

*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).

A. Aroudi and S. Doclo are with the Department of Medical Physics and Acoustics, University of Oldenburg, Germany. (e-mail: {ali.aroudi, simon.doclo}@uni-oldenburg.de)

978-1-5386-1644-4/17/$31.00 ©2017 IEEE

A. Acoustic scenario

Consider an acoustic scenario comprising two competing speakers and background noise in a reverberant environment, where the ongoing EEG responses of a listener to these acoustic stimuli are recorded (cf. Fig. 1). The binaural signals at the ears hence consist of a mixture of both clean speech signals $s^j[n]$, with $j = a$ denoting the attended speaker and $j = u$ denoting the unattended speaker, incorporating head filtering effect, reverberation and background noise. The signal at the m-the ear $y_m[n]$, with $m = 1$ denoting the left ear and $m = 2$ denoting the right ear, at the discrete time index n can be written as

$$y_m[n] = \sum_{j=a,\,u} \underbrace{h_m^j[n] * s^j[n]}_{x_m^j[n]} + v_m[n], \tag{1}$$

with $h_m^j[n]$ the acoustic impulse response between the j-th speaker and the m-th ear, $*$ the convolution operation, $x_m^j[n]$ the reverberant speech signal of the j-th speaker at the m-th ear, and $v_m[n]$ the background noise component at the m-th ear. The reverberant speech signal $x_m^j[n]$ consists of the anechoic speech signal $x_m^{j,an}[n]$ (encompassing the head filtering effect with the anechoic HRTF [8]) and a reverberant component. For notational conciseness the time index n will be omitted in the remainder of this paper.

For the EEG recordings we will consider four different acoustic conditions, i.e. anechoic, reverberant, noisy and reverberant-noisy. Depending on the acoustic condition, the signal at the m-th ear obviously comprises different components, i.e. reverberation, background noise, and interference (for the attended speech signal the unattended speech signal is defined as interference, while for the unattended speech signal the attended speech signal is defined as interference). For the anechoic condition, it is referred to as anechoic speech signal with interference and equal to

$$x_m^{an} = \sum_{j=a,\,u} x_m^{j,an}, \tag{2}$$

for the *reverberant* condition it is referred to as reverberant speech signal with interference and equal to

$$x_m = \sum_{j=a,\,u} x_m^j, \tag{3}$$

for the *noisy* condition it is referred to as anechoic-noisy speech signal with interference and equal to

$$x_m^{no} = x_m^{an} + v_m, \tag{4}$$

and for the *reverberant-noisy* condition it is referred to as reverberant-noisy speech signal with interference and equal to

$$y_m = x_m + v_m. \tag{5}$$

For investigating the impact of reducing undesired acoustic components on the AAD performance we will consider several simulated signals, i.e. the reverberant speech signal with interference x_m in which noise has been reduced, the anechoic-noisy speech signal with interference x_m^{no} in which reverberation has been reduced, the anechoic speech signal with interference x_m^{an} in which noise and reverberation have jointly been reduced, the reverberant-noisy speech signal $x_m^{j,rn} = x_m^j + v_m$ in which interference has been reduced, the anechoic-noisy speech signal $x_m^{j,no}$ in which interference and reverberation have jointly been reduced, the reverberant speech signal x_m^j in which interference and noise have jointly been reduced, the anechoic speech signal $x_m^{j,an}$ in which interference, noise and reverberation have jointly been reduced, and the clean speech signal s^j in which interference, noise and reverberation have jointly been reduced and the head filtering effect has been canceled. A summary of all discussed signals is shown in Table I.

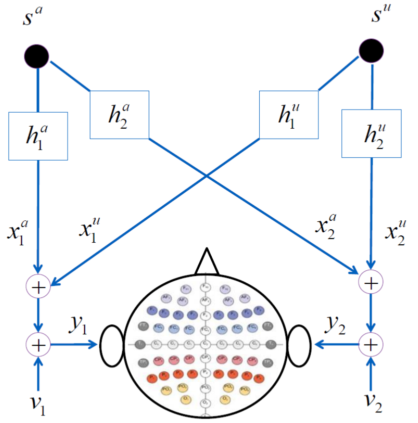


Figure 1. The binaural acoustic configuration used for stimuli presentation in different acoustic conditions.

Table I
SIGNALS USED FOR EXPERIMENTAL ANALYSIS.

Signal	Definition
Reverberant-noisy speech signal with interference	y_m
Reverberant speech signal with interference	x_m
Anechoic-noisy speech signal with interference	x_m^{no}
Anechoic speech signal with interference	x_m^{an}
Reverberant-noisy speech signal	$x_m^{j,rn}$
Anechoic-noisy speech signal	$x_m^{j,no}$
Reverberant speech signal	x_m^j
Anechoic speech signal	$x_m^{j,an}$
Clean speech signal	s^j

It should be noted that in the experiments (cf. Section III) 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). Due to the head filtering effect this implies that the broadband energy ratio between the attended speech component and the unattended speech component at the side of the attended speakers always larger than at the side of the unattended speaker. Therefore, we will consider the speech signals at the side of the attended speaker as the attended speech signals and the speech signals at the side of the unattended speaker as the unattended speech signals.

B. Training step

For the training step, the attended speaker is assumed to be known and the attended speech envelope $e^a[i]$, with $i = 1 \dots I$ the sub-sampled time index, is used for filter training. The attended clean speech signal s^a is typically used for computing the attended speech envelope $e^a[i]$.

The AAD method proposed in [1] uses a spatio-temporal filter to estimate the attended speech envelope $\hat{e}^a[i]$ from C-channel EEG recordings $r_c[i]$ $(c = 1 \dots C)$ as

$$\hat{e}^a[i] = \sum_{c=1}^{C}\sum_{l=0}^{L-1} w_{c,l}\, r_c[i + \Delta + l], \tag{6}$$

with $w_{c,l}$ the l-th filter coefficient in the c-th channel, L the number of filter coefficients per channel, and Δ modeling the latency of the attentional effect in the EEG responses to the speech stimuli. In vector notation, (6) can be written as

$$\hat{e}^a[i] = \mathbf{w}^T \mathbf{r}[i], \tag{7}$$

with

$$\mathbf{w} = \left[\mathbf{w}_1^T\, \mathbf{w}_2^T \dots \mathbf{w}_C^T\right]^T, \tag{8}$$

$$\mathbf{w}_c = \left[w_{c,0}\, w_{c,1} \dots w_{c,L-1}\right]^T, \tag{9}$$

$$\mathbf{r}[i] = \left[\mathbf{r}_1^T[i]\; \mathbf{r}_2^T[i] \dots \mathbf{r}_C^T[i]\right]^T, \tag{10}$$

$$\mathbf{r}_c[i] = \left[r_c[i+\Delta]\;\, r_c[i+\Delta+1] \dots r_c[i+\Delta+L-1]\right]^T, \tag{11}$$

with $(.)^T$ the transpose operation. During the training step, the filter $\mathbf{w}$ 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 [1], i.e.

$$J(\mathbf{w}) = \frac{1}{I}\sum_{i=1}^{I}\left(e^a[i] - \mathbf{w}^T\mathbf{r}[i]\right)^2 + \beta \mathbf{w}^T \mathbf{D}\mathbf{w}, \tag{12}$$

with

$$\mathbf{D} = \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{bmatrix}, \tag{13}$$

and β the regularization parameter. The filter minimizing the regularized cost function in (12) is equal to [2], [9]

$$\mathbf{w} = \left(\mathbf{Q} + \beta\mathbf{D}\right)^{-1}\mathbf{q}, \tag{14}$$

with the correlation matrix $\mathbf{Q}$ and the the cross-correlation vector $\mathbf{q}$ equal to

$$\mathbf{Q} = \frac{1}{I}\sum_{i=1}^{I}\left(\mathbf{r}[i]\,\mathbf{r}^T[i]\right),\;\; \mathbf{q} = \frac{1}{I}\sum_{i=1}^{I}\left(\mathbf{r}[i]\, e^a[i]\right). \tag{15}$$

For each acoustic condition, the complete set of EEG recordings is segmented into T_{tr} trials. The correlation matrix and the cross-correlation vector corresponding to trial t are denoted as $\mathbf{Q}_t$ and $\mathbf{q}_t$, respectively, where t denotes the trial index. The filter to decode trial t is computed as

$$\tilde{\mathbf{w}}_t = \left(\tilde{\mathbf{Q}}_t + \beta\mathbf{D}\right)^{-1}\tilde{\mathbf{q}}_t, \tag{16}$$

with $\tilde{\mathbf{Q}}_t$ the average correlation matrix of trial t, computed by averaging all correlation matrices except $\mathbf{Q}_t$ (known as leave-one-out averaging), and $\tilde{\mathbf{q}}_t$ the average cross-correlation vector of trial t, computed by averaging all cross-correlation vectors except $\mathbf{q}_t$. Since EEG responses are recorded for different acoustic conditions, in this paper we will consider several training conditions (tc) for computing the filter $\tilde{\mathbf{w}}_t$, i.e. $tc = an$ using EEG responses in the *anechoic* condition, $tc = re$ using EEG responses in the *reverberant* condition, $tc = no$ using EEG responses in the *noisy* condition, and $tc = rn$ using EEG responses in the *reverberant-noisy* condition. For all considered training conditions, we will consider the anechoic attended speech signal as training signal.

C. Evaluation step

To decode to which speaker a listener attended during trial t, first an estimate of the the attended speech envelope $\hat{e}_t^a$ is computed using the (trained) filter $\tilde{\mathbf{w}}_t$ in (16), i.e.

$$\hat{e}_t^a = \tilde{\mathbf{w}}_t^T \mathbf{r}_t, \tag{17}$$

with $\mathbf{r}_t$ the EEG recordings of trial t. Based on the attended and the unattended correlation coefficients, i.e.

$$\rho_t^a = \rho\left(e_t^a, \hat{e}_t^a\right),\;\; \rho_t^u = \rho\left(e_t^u, \hat{e}_t^a\right), \tag{18}$$

with e_t^u the unattended speech envelope, it is then decided that auditory attention has been correctly decoded when $\rho_t^a > \rho_t^u$. Accordingly, a larger difference between the attended and the unattended correlation coefficient $\rho_t^a - \rho_t^u$ (correlation difference) is indicative of a more reliable AAD decision. The decoding performance is defined as the percentage of correctly decoded trials over all considered trials and over all participants.

In most previous work, e.g. [1], [10], it has been assumed that the attended and the unattended clean speech signals s^a and s^u are available as reference signals for computing the speech envelopes e_t^a and e_t^u in (18), which is quite unrealistic in practice. To generate reference signals from the binaural signals at the ears, the impact of reducing each acoustic components on the AAD performance needs to be determined. In this paper we address this issue by using the simulated (attended and unattended) signals (cf. Table I) as reference signals for several acoustic conditions ec, $ec \in \{an, re, no, rn\}$.

Table II
ACOUSTIC CONDITIONS USED FOR EXPERIMENTAL ANALYSIS AND STIMULI PRESENTATION.

Experimental Analysis	Acoustic Condition	SNR [dB]	T_{60} [s]
Anechoic	Anechoic [11]	∞	< 0.05
Reverberant	Reverberant I [11]	∞	0.5
	Reverberant II [12], [13]	∞	1.0
Noisy	Noisy I [11]	9.0	< 0.05
	Noisy II [11]	4.0	< 0.05
Reverberant-Noisy	*Reverberant-Noisy* I [11]	9.0	0.5
	Reverberant-Noisy II [11]	4.0	0.5
	Reverberant-Noisy III [12], [13]	9.0	1.0

Please note that all analyses in this paper are performed with the same training and evaluation conditions, i.e. $tc = ec$, such that the influence of acoustical differences between training and evaluation conditions are excluded. Investigating the influence of such acoustical differences on AAD is beyond the scope of this paper.

In [2] it has been shown that tuning the parameters involved in the filter design (L, Δ, β) plays a key role in optimizing the decoding performance. In order not to favour one specific acoustic condition, in this paper the parameters have been tuned to optimize the average decoding performance over all considered acoustic conditions (per participant).

III. ACOUSTIC AND EEG MEASUREMENT SETUP

Eighteen native German-speaking participants aged between 21 and 34 years with normal hearing took part in this study. Two stories in German, uttered by two different male speakers, were simultaneously presented to the participants using earphones at a sampling frequency of 48 kHz. Among all participants, 8 participants were instructed to attend to the left speaker, while 10 participants were instructed to attend to the right speaker. The stimuli were presented in 11 sessions, each of length 10 minutes, interrupted by short breaks. The participants were instructed to look ahead and minimize eye blinking. 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 presented stimuli at both ears were simulated by convolving the clean speech signals (stories) with (non-individualized) binaural acoustic impulse responses, either from [11] or [12], and adding diffuse noise, generated according to [14]. The left and the right speakers were simulated at $-45°$ and $45°$, respectively. Eight different acoustic conditions were considered (cf. Table II): anechoic, reverberant with a moderate and a large reverberation time ($T_{60} = 0.5$ s, $T_{60} = 1$ s), noisy with two different signal-to-noise ratios (SNR $= 9.0$ dB, SNR $= 4.0$ dB), and three combinations of reverberation and noise. For each participant, the anechoic condition was assigned to the first session and subsequently to every other third session (i.e. session 4, 7, and 10). Aiming at minimizing the influence of the speech material on AAD, the acoustic conditions (except for the anechoic condition) were randomly assigned to the other sessions. For experimental analysis, the acoustic conditions were grouped based on acoustic similarity as shown in Table II, resulting in four experimental analysis conditions, i.e. *anechoic*, *reverberant*, *noisy*, and *reverberant-noisy*.

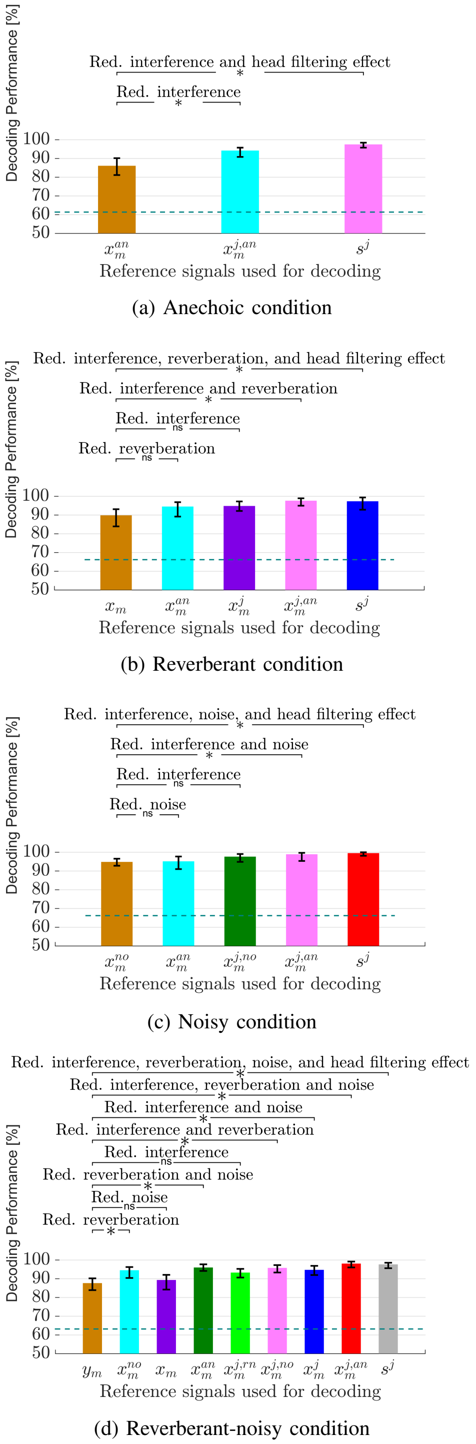
The EEG responses were recorded using $C = 64$ channels at a sampling frequency of 500 Hz, and referenced to the nose electrode. The EEG responses were offline re-referenced to a common average reference, band-pass filtered between 2 and 8 Hz using a third-order Butterworth band-pass filter, and subsequently downsampled to $f_s = 64$ Hz. The envelopes of the speech signals were obtained using a Hilbert transform, followed by low-pass filtering at 8 Hz and downsampling to $f_s = 64$ Hz. For the training and evaluation steps, the EEG recordings of each session were split into 10 trials, each of length 60 seconds. Each participant's own data were used for filter training and evaluation.

IV. RESULTS AND DISCUSSION

For all considered signals used as reference signals (cf. Table I), Fig. 2 presents the decoding performance for different acoustic conditions (cf. Table II). It is noted the number of bars in each acoustic condition corresponds to the number of available reference signals in the underlying acoustic condition. It can be observed that for all considered signals used as reference signals a good decoding performance (larger than 86%) can be obtained. These results are consistent with the previous findings in which either the clean speech signals (s^j), the anechoic speech signals ($x_m^{j,an}$) [1], [4], [2], [10], [15], [16] or the (unprocessed) binaural signals (x_m^{an}, x_m, x_m^{no}, y_m) [4] were used as reference signals for decoding.

First, we investigate the impact of individual reduction of undesired acoustic components on the AAD performance. For all acoustic conditions, when interference is reduced the decoding performance is larger than when using the binaural signals as reference signals, although the decoding performance difference is only significant ($p < 0.05$) for the anechoic condition. For the reverberant condition and the reverberant-noisy condition, when reverberation is reduced the decoding performance is larger than when using the binaural signals as reference signals, although the decoding performance difference is only significant ($p < 0.05$) for the reverberant-noisy condition. For the noisy condition and the reverberant-noisy condition, when background noise is reduced there is no significant difference ($p > 0.05$) in the decoding performance compared to when using the binaural signals as reference signals.

Secondly, we investigate the impact of the joint reduction of undesired acoustic components on the AAD performance. For all acoustic conditions, the decoding performance can significantly ($p < 0.05$) be improved when all undesired acoustic components are reduced, i.e. reducing interference for the anechoic condition, jointly reducing interference and noise



(a) Anechoic condition

(b) Reverberant condition

(c) Noisy condition

(d) Reverberant-noisy condition

Figure 2. Impact of reducing noise, reverberation, and interference on the AAD performance. Comparison of the decoding performance when different acoustic components are reduced in (a) the anechoic condition, (b) the reverberant condition, (c) the noisy condition, (d) the reverberant-noisy condition. The dashed lines represent the upper boundary confidence interval of chance level based on a binomial test at the 5% significance level, the error bars represent the bootstrap confidence interval at the 5% significance level, the asterisks represent the significant decoding performance difference ($p < 0.05$) using a Kruskal-Wallis test followed by post-hoc paired Wilcoxon signed rank test, and red. stands for reducing.

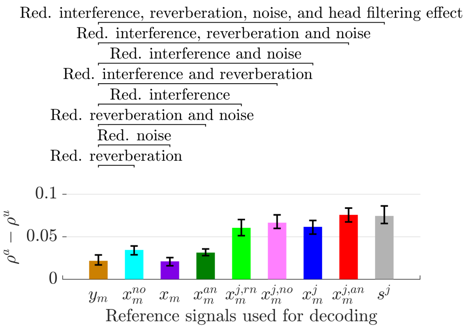


Figure 3. Impact of reducing noise, reverberation, and interference on the correlation difference (averaged across trials and participants) in the reverberant-noisy condition. The error bars represent the bootstrap confidence interval at the 5% significance level and red. stands for reducing.

for the noisy condition, jointly reducing interference and reverberation for the noisy condition, jointly reducing interference, noise and reverberation for the reverberant-noisy condition. This can be explained by considering the impact of the joint reduction of undesired acoustic components on the correlation difference ($\rho_t^a - \rho_t^u$). For all considered reference signals, Fig. 3 presents the resulting correlation difference, averaged across trials and participants, for the reverberant-noisy acoustic condition as the most challenging listening condition (note that these average correlation coefficients are not directly used for decoding). It can be observed that the largest correlation difference can be obtained when all undesired acoustic components are reduced, i.e., using either the clean speech signals (s^j) or the anechoic speech signals ($x^{j,an}$) as reference signals.

V. Conclusion

In this paper, we have investigated the impact of undesired acoustic components on the AAD performance for different acoustic conditions (anechoic, reverberant, noisy, and reverberant-noisy). The experimental results show that for obtaining a good decoding performance the joint suppression of reverberation, background noise and interference is of great importance.

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.

[2] 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.

[3] 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.

[4] 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 Koreas*, 2017.

[5] 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.

[6] 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.

[7] E. Habets, J. Benesty, and P. A. Naylor, “A speech distortion and interference rejection constraint beamformer,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 3, pp. 854–867, 2012.

[8] J. Blauert, *Spatial hearing : the psychophysics of human sound localization*. Cambridge, Mass. MIT Press, 1997.

[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] 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.

[11] 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.

[12] M. Jeub, M. Schäfer, and P. Vary, “A binaural room impulse response database for the evaluation of dereverberation algorithms,” in *Proc. of International Conference on Digital Signal Processing (DSP)*, Santorini, Greece, Jul. 2009, pp. 1–5.

[13] J. Thiemann and S. van de Par, “Multiple model high-spatial resolution HRTF measurements,” in *Proc. DAGA*, 2015.

[14] 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.

[15] 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.

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