EEG-based Attention-Driven Speech Enhancement For Noisy Speech Mixtures Using N-fold Multi-Channel Wiener Filters

Neetha Das*†, Simon Van Eyndhoven*, Tom Francart† and Alexander Bertrand*
*Dept. Electrical Engineering (ESAT), KU Leuven, Belgium
†Dept. Neurosciences, ExpORL, KU Leuven, Belgium

***Abstract*—Hearing prostheses have built-in algorithms to perform acoustic noise reduction and improve speech intelligibility. However, in a multi-speaker scenario the noise reduction algorithm has to determine which speaker the listener is focusing on, in order to enhance it while suppressing the other interfering sources. Recently, it has been demonstrated that it is possible to detect auditory attention using electroencephalography (EEG). In this paper, we use multi-channel Wiener filters (MWFs), to filter out each speech stream from the speech mixtures in the microphones of a binaural hearing aid, while also reducing background noise. From the demixed and denoised speech streams, we extract envelopes for an EEG-based auditory attention detection (AAD) algorithm. The AAD module can then select the output of the MWF corresponding to the attended speaker. We evaluate our algorithm in a two-speaker scenario in the presence of babble noise and compare it to a previously proposed algorithm. Our algorithm is observed to provide speech envelopes that yield better AAD accuracies, and is more robust to variations in speaker positions and diffuse background noise.**

I. Introduction

Signal processing algorithms in hearing aids and cochlear implants allow to suppress background noise for improved speech intelligibility for the hearing impaired. By using multiple microphones, beamforming techniques can be applied to filter out sound from a target direction, and to suppress the noise from other directions. Adaptive beamformers are particularly powerful, as they allow to change and optimize their beam pattern to the acoustic scenario [1], [2]. However, in a multi-speaker scenario, a fundamental challenge is to determine which speaker the listener actually aims to focus on. Therefore, incorporating a brain-computer interface to infer the auditory attention of the listener opens up an interesting field of research aiming to build smarter hearing prostheses [3].

Various recent studies have demonstrated that it is possible to perform auditory attention detection (AAD) based on neural measurements such as EEG [4]–[7], and that differential tracking of the attended and unattended speech streams, necessary for AAD, is present also in hearing-impaired listeners [8]. Supported by discreet EEG recording technology [9]–[11], such AAD algorithms could work hand in hand with noise suppression systems in hearing aids to form neuro-steered hearing prostheses. A proof of concept for this idea was presented in [3] where a pre-trained AAD decoder reconstructed the attended speech stream's envelope from EEG recordings, which was then correlated with speech envelopes extracted from microphone signals using a blind envelope demixing algorithm. The envelope with the higher correlation with the reconstructed attended speech stream was considered to belong to the attended speaker. The voice activity pattern was then extracted from this envelope, and was used to drive a multi-channel Wiener filter (MWF) [1] filtering out the attended speech stream. This work demonstrated that MWF-based speech enhancement can rely on EEG-based attention detection to extract the attended speaker from a set of microphone signals, boosting signal to noise ratios (SNRs) even in noisy environments, and without prior knowledge of the clean speech envelopes to perform AAD. Nevertheless a significant reduction in AAD performance was observed compared to the case where clean speech envelopes are available (the effect of uncorrelated noise on speech envelopes for AAD has been investigated in detail in [6]). It was found that the AAD performance was robust to noise in the envelopes introduced by the demixing algorithm as well as uncorrelated babble noise in the microphone signals, although large variability in performance (52-98% accuracy) was observed over different subjects.

In [12], an adaptive version of the same algorithm was presented analyzing behavior of such a system in real-time when there is a switch in attention. In this work, it was found that EEG-informed AAD can steer an adaptive MWF towards the attended speaker, provided that the AAD accuracy is high. In this case, when a switch in attention is detected, the MWF has to first forget all its old statistics and re-converge to a new filter that targets the new attended speaker, which adds an extra adaptation delay in the algorithm (in addition to the intrinsic delay associated with detecting the attention switch).

We propose an improved algorithm where multiple MWFs receive different speaker-dependent voice activity information from the speaker envelopes extracted by the blind envelope demixing algorithm (unlike in [3] where a single MWF received the speaker voice activity track chosen by the AAD module). Also, instead of using the envelopes extracted from the demixing algorithm directly for attention detection, we use envelopes extracted from the output of the multiple MWFs. This allows the AAD to use cleaner speech envelopes, which improves and robustifies the attention detection. Another dif-

This work was carried out at the ESAT and ExpORL Laboratory of KU Leuven, and was supported by a research gift of Starkey Hearing Technologies, and has received funding from KU Leuven Special Research Fund BOF/STG-14-005, OT/14/119, C14/16/057 and CoE PFV/10/002 (OPTEC), IUAP nr. P7/19 (DYSCO), and the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No 637424). The scientific responsibility is assumed by its authors. The experiments were approved by the KU Leuven ethical committee and all subjects (and/or their legal guardian) signed an informed consent form.

ISBN 978-0-9928626-7-1 © EURASIP 2017

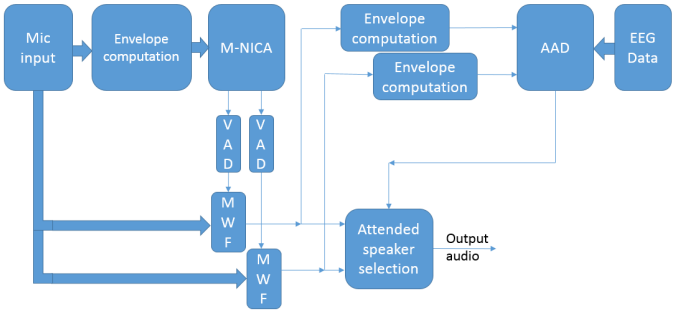

Fig. 1. Block diagram showing the different modules that constitute the proposed neuro-steered speech enhancement algorithm, for the special case of a two-speaker scenario.

ference in this algorithm is that, the use of multiple parallel MWFs (i.e., one for each speaker) eliminates the need to reset the MWF and let it re-converge to another speaker if an attention switch occurs (as was done in [12]). The use of multiple MWFs also gives us the flexibility to build a more realistic output by weighted averaging the multiple MWF outputs based on attention decisions.

Our new algorithm is benchmarked against the algorithm in [3] by computing AAD accuracies over a range of SNRs, and a range of source positions, all in the presence of diffuse babble noise.

II. Neuro-steered speech enhancement algorithm

A. Data model and problem statement

We consider a binaural hearing prosthesis equipped with M microphones capturing audio signals $\mathbf{y}[t] = [y_1[t]...y_M[t]]^T$ in an N-speaker scenario[1]. In the frequency domain, these microphone signals can be represented as

$$\mathbf{y}(\omega) = \sum_{i=1}^{N} \mathbf{x}_i(\omega) + \mathbf{n}(\omega) = \sum_{i=1}^{N} \mathbf{H}_i(\omega) s_i(\omega) + \mathbf{n}(\omega) \quad (1)$$

where for a frequency ω, $\mathbf{y}(\omega) = [y_1(\omega)...y_M(\omega)]^T$ denotes the vector in which all M microphone signals are stacked, $\mathbf{x}_i(\omega)$ denotes the M stacked signal components corresponding to the speaker s_i as it is observed by the M microphones, $\mathbf{n}(\omega)$ denotes the stacked noise components, picked up by the M microphones, and $\mathbf{H}_i(\omega)$ denotes the frequency domain representation of head-related transfer functions (HRTFs) that model the acoustic propagation path between the source s_i and the M microphones. We aim to enhance the speech component of one speaker, while suppressing that of the other speakers and noise. As described in [3], [12], this can be achieved using a multi-channel Wiener filter (MWF) $\mathbf{w}(\omega)$, that extracts the attended speech stream $\tilde{s}_{att}(\omega) = \mathbf{w}(\omega)^H \mathbf{y}(\omega)$ provided that we have the knowledge of the times at which the attended speaker is active (superscript H denotes the conjugate transpose operator).

[1]For the sake of generality, we describe the algorithm for an N-speaker scenario. However, the experiments in Section III focus on a 2-speaker scenario, as in most of the literature on AAD.

B. Overview of the algorithm

This subsection briefly summarized the different sub-blocks of the proposed algorithm (figure 1), which will be further explained in subsections II-C to II-E. First, a blind envelope demixing algorithm extracts the per-speaker energy envelopes from the M microphone signals. From these envelopes, binary voice activity detection (VAD) signals are determined for each speaker. The N VAD tracks are used to design N MWFs that run in parallel, from hereon referred to as N-fold MWFs, to extract each speech stream exclusively. In order to detect to which of the speakers the listener is attending, an auditory attention detection (AAD) module reconstructs the attended speaker's envelope with the help of a pre-trained decoder applied to the listener's EEG. The reconstructed envelope is then correlated to the speech envelopes extracted from the outputs of the MWFs to determine which MWF output corresponds to the attended speaker (i.e., the one whose envelope correlates better with the reconstructed envelope).

In [3], [12], a different algorithm was presented with a single MWF, whose input VAD signal is switched between the speakers, based on the output of the AAD module. Such an algorithm suffers from several disadvantages. In [3], the speech envelopes that are fed to the AAD module were extracted using an energy envelope demixing algorithm [13]. In our new algorithm, we extract the envelopes from the outputs of the N-fold MWFs (referred to as MWF envelopes) and feed these to the AAD module instead. These envelopes have the advantage of the noise suppression brought in by the MWFs and are therefore expected to result in better AAD performance, particularly in cases with low input SNR. In [12], output SNR is susceptible to drop when the listener switches attention since, once the attention switch is detected, the MWF has to 'forget' all the speech and noise statistics so far, and 'relearn' these. With our proposed algorithm, the N-fold MWFs being fed with exclusive VAD signals for each speaker are, by design, more stable to learn speaker and noise statistics. This can ensure smoother tracking even when there is a switch in attention.

C. Envelope demixing for voice activity detection

The multiplicative non-negative independent component analysis (M-NICA) algorithm can be used to solve blind source separation problems where the underlying sources are independent, non-negative and well-grounded [14], which is the case for speech energy envelopes [13]. To apply the M-NICA algorithm, we convert the microphone signals to the energy domain by computing their short-term energies over windows of T samples, which for the ith microphone signal $y_i[t]$ (in the time domain) is given by

$$\mathbf{E}_i[n] = \frac{\sum_{w=1}^{T} y_i[nT+w]^2}{T}. \quad (2)$$

Note that this operation downsamples the signals with a factor T. The microphone energy signals are assumed to be linear mixtures of the original energy signals of the speech sources. The M-NICA algorithm then extracts the speech

ISBN 978-0-9928626-7-1 © EURASIP 2017

energy envelopes from the microphone energy envelopes based on multiplicative updates along with subspace projection. The algorithm exploits the differences in the energy distributions of the different sources across the (binaural) microphone array. By exploiting the non-negativity properties of the underlying sources, the algorithm utilizes only second order statistics, as opposed to many other source separation algorithms which rely on higher order statistics. Furthermore, as the algorithm operates on energy envelopes rather than raw microphone signals, it can operate at a much lower sampling rate. Both these aspects contribute to computational efficiency, which is a desirable factor to incorporate such a noise-suppression scheme in an actual hearing prosthesis.

In the next step, binary voice activity detection signals are generated from each of the energy envelopes extracted by simple thresholding (see section III for details).

D. Speech enhancement using multi-channel Wiener filters

Assume that we aim to estimate the n-th speech signal as it is observed in an arbitrary microphone r, which we refer to as the reference microphone. The MWF filter coefficients per frequency bin are computed such that the difference between the output $\tilde{x}_{n,r}(\omega) = \mathbf{w}_n(\omega)^H \mathbf{y}(\omega)$ and the n-th speaker contribution in reference microphone r is minimized in the linear minimal mean square error (LMMSE) sense, i.e.,

$$\hat{\mathbf{w}}_n = \arg\min_{\mathbf{w}_n} E\{|x_{n,r} - \tilde{x}_{n,r}|^2\} = \arg\min_{\mathbf{w}_n} E\{|x_{n,r} - \mathbf{w}_n^H \mathbf{y}|^2\} \quad (3)$$

where $E\{\cdot\}$ denotes the expectation operator and where $x_{n,r}$ denotes the signal of speaker n in microphone r. Note that we have omitted the frequency variable for the sake of simplicity. In practice, the MWF has to be computed for each frequency bin separately in the short-time Fourier transform (STFT) domain. The LMMSE design (3) results in an MWF given by [1]:

$$\hat{\mathbf{w}}_n = \mathbf{R}_{yy}^{-1} \mathbf{R}_{x_n x_n} \mathbf{e}_r \quad (4)$$

where $\mathbf{R}_{yy} = E\{\mathbf{y}(\omega)\mathbf{y}(\omega)^H\}$, $\mathbf{R}_{x_n x_n} = E\{\mathbf{x}_n(\omega)\mathbf{x}_n(\omega)^H\}$, and $\mathbf{e}_r$ denotes the r-th column of an $M \times M$ identity matrix, which selects the column of $\mathbf{R}_{x_n x_n}$ corresponding to the reference microphone. Matrix $\mathbf{R}_{yy}$ is estimated during the time periods when the first speaker is active (i.e. the 'speech plus interference' autocorrelation matrix). The matrix $\mathbf{R}_{x_n x_n}$ is not known but can be estimated as $\mathbf{R}_{x_n x_n} = \mathbf{R}_{yy} - \mathbf{R}_{vv}$ assuming independence between all sources, where $\mathbf{R}_{vv} = E\{\mathbf{n}(\omega)\mathbf{n}(\omega)^H\} + \sum_{i \neq n} \mathbf{P}_i \mathbf{H}_i(\omega)\mathbf{H}_i(\omega)^H$ where $\mathbf{P}_i = E\{|s_i(\omega)|^2\}$. Here, $\mathbf{R}_{vv}$ can be estimated by averaging over all STFT frames in which speaker n is not active. Note that in practice, the approximation $\mathbf{R}_{yy} - \mathbf{R}_{vv}$ often leads to poor filters, in particular if $\mathbf{R}_{vv}$ contains non-stationary sources. Therefore, we use a more robust estimation of $\mathbf{R}_{x_n x_n}$, based on a generalized eigenvalue decomposition of $\mathbf{R}_{yy}$ and $\mathbf{R}_{vv}$ (details in [2]).

As mentioned in subsection II-C, the information about the active and silent periods of each speaker is obtained from the VAD signals estimated from the speech energy envelopes extracted by the M-NICA module (M-NICA envelopes). Finally, multiple parallel MWFs $\mathbf{w}_n$ computed using (4) for the N different speakers are used to estimate the corresponding speakers' speech streams. Note that the N MWFs can share a large part of the computations, as they all rely on the same matrix inverse $\mathbf{R}_{yy}^{-1}$.

E. Auditory attention detection

We assume the listener's EEG is recorded while listening to the speech mixtures, which are used to detect to which of the speakers the subject aims to focus on. During a training phase the subject is asked to attend to one of the speakers. Using knowledge of the speech envelope of this speaker, a spatio-temporal decoder is trained on the recorded EEG data to reconstruct the attended speech stream's envelope using linear regression [4], [5], [7]. The decoder reconstructs the attended stream by linearly combining the EEG data over C channels over T time lags, given by

$$p[t] = \sum_{\tau=0}^{T-1} \sum_{c=1}^{C} d_c[\tau]\, r_c[t+\tau] \quad (5)$$

where the $d_c[\tau]$'s denote the decoder weights and $r_c[t]$ denotes the c-th EEG channel. For the sake of notation, we stack all the $d_c[\tau]$ in a vector $\mathbf{d} = [d_1[0], d_1[1], ..., d_1[T-1], d_2[0], ..., d_2[T-1], ..., d_C[0], ..., d_C[T-1]]^T$. The optimal decoder that minimizes the difference between $p[t]$ and the attended speech envelope in the least squares sense is given by

$$\hat{\mathbf{d}} = \mathbf{R}_{rr}^{-1} \mathbf{c}_{rs_{att}} \quad (6)$$

where $\mathbf{R}_{rr} = E\{\mathbf{r}[t]\ \mathbf{r}[t]^T\}$, $\mathbf{c}_{rs_{att}} = E\{\mathbf{r}[t] s_{att}[t]\}$, $\mathbf{r}[t] = [r_1[t], r_1[t+1], ..., r_1[t+T-1], r_2[t], ..., r_2[t+T-1], ..., r_C[t], ..., r_C[t+T-1]]^T$ represents the stacked EEG samples of all C channels for T time lags, and $s_{att}[t]$ represents the attended speech envelope used to train the decoder[2].

To perform attention detection, the trained decoder is then applied to the EEG data to compute the corresponding reconstructed attended stream $p[t]$. This $p[t]$ is then correlated with each of the speech envelopes from the different MWF outputs to find the one resulting in the higher correlation value. The output of the MWF that corresponded to this envelope can then be deemed to be the filtered attended speech stream.

III. Experiment

We use 64-channel EEG data collected from 16 normal-hearing subjects while they were listening to two simultaneously active speech signals. During the experiment, the subjects attended to one of the two speech sources, which were presented at 60dBA through insert phones. The experiment resulted in a total of 36 minutes of EEG recordings, where the subject was asked to switch attention between left and right ear across trials and the speech stimuli were filtered using head-related transfer functions (HRTFs) so that the subject

[2]As suggested in [7], a single decoder is computed using the entire training set instead of computing one for every trial period and then averaging (as in [4]). This has been shown to improve AAD performance and reduce or eliminate the need for regularization.

ISBN 978-0-9928626-7-1 © EURASIP 2017

perceived an auditory environment with 2 speakers located at 90 degrees to the left and right[3].

All data were split into 30 second trials and the EEG data were bandpass filtered between 1-9 Hz (the frequency range of most interest for tracking speech stimuli [16]–[18]), and down-sampled to 20 Hz. The EEG decoder was trained using all the recorded data, except the trial under test using the envelope extracted from the (known) attended speech stimulus. Envelope extraction was done by taking the absolute value and band-pass filtering the result between 1-9 Hz to be consistent with the EEG pre-processing.

We simulated microphone signals for 6 behind-the-ear microphones on a binaural hearing aid using publicly available HRTF coefficients recorded in an anechoic room [19]. Thus, the MWFs and the M-NICA module received $M = 6$ microphone signals. We simulated different speaker setups, each time consisting of 2 speakers placed at different angular positions at 3 meters distance with respect to the listener at the center, as well as uncorrelated babble noise sources positioned every 5 degrees around the listener at the same distance. The speech stimuli were the same as those used in the experiment in [7] whose EEG recordings we used for attention detection. In figure 2, the table shows the 12 analyzed speaker location pairs. The power spectral densities (PSD) of all the noise sources where adjusted to match the average of the PSDs of the two speakers to obtain speech-weighted noise spectra. For each speaker location pair, we analyzed the proposed algorithm for various input SNRs.

The microphone signals were down-sampled to 8 kHz and windows of $T = 400$ samples were used to compute the microphone energy envelopes using (2). The M-NICA module was applied in batch mode to generate demixed speech energy envelopes based on the microphone energy envelopes. Before computing (2) the microphone signals were low pass filtered with an 800 Hz cut-off before applying the demixing algorithm, since this step was found to be beneficial for effective demixing particularly for lower SNRs. VAD signals were extracted from the demixed energy envelopes by taking the 75th percentile value as a threshold, above which the speaker is considered to be active. Although this is a conservative threshold introducing many false negatives, for lower input SNR scenarios, where the extracted envelopes are increasingly corrupted by noise, this thresholding criterion has been empirically found to provide the best VAD signals that produced better MWFs in the next step. The signals are fed to the two MWFs that run in batch mode to generate the two noise-suppressed speech streams. The outputs of the MWFs also go through the same envelope extraction procedure as the audio signals in the training phase before correlating them with the reconstructed attended speech envelope from the EEG data.

[3]The data set for this experiment is a subset of a previously collected data set from [7] in which dry stimuli as well as HRTF-filtered speech stimuli were used [15]. Only the EEG data during HRTF-filtered speech streams were used here, in order to mimic a realistic listening scenario, and because it was shown that HRTF-filtered stimuli result in higher AAD accuracy [16].

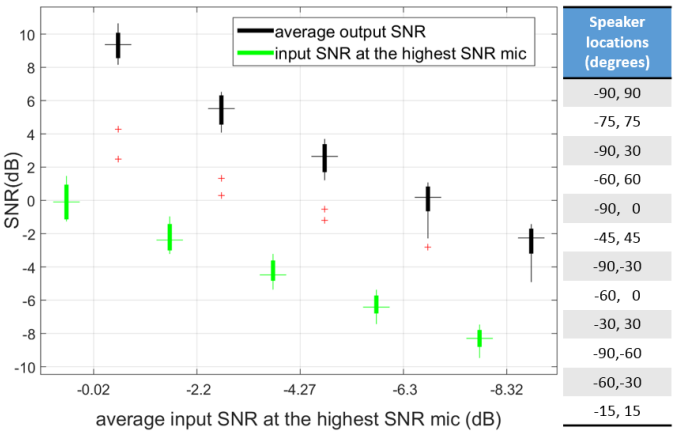

Fig. 2. Input and output SNRs for the 12 source locations. The table shows the investigated locations of the two speakers around the listener.

IV. Results

In this work, we have used the output SNR of the N-fold MWFs and accuracy of attention detection as the metrics to assess the performance of our proposed algorithm. SNR values are defined such that the attended speaker is treated as the signal, while the unattended speaker and the babble noise are treated as noise. Input SNRs are computed at the microphone with highest input SNR value.

For the scenarios which include babble noise, the noise reduction performance is plotted in figure 2 in relation to the input SNR at the highest-SNR microphone for each scenario. The box plots show the variation of the input/output SNR over the 12 speaker positions. For the scenario without diffuse babble noise, the MWF can almost perfectly cancel the unattended speaker, with an output SNR of 16 dB up to 32 dB (27 dB on average) over all scenarios. This scenario is omitted from figure 2 for the sake of intelligibility of the figure.

Figure 3 shows the boxplots of AAD accuracy obtained by both algorithms for the 12 source positions, where for each position the average AAD accuracy across 16 subjects was taken. The results are shown for various input SNRs, where the first one (2.45dB) corresponds to the noiseless case without babble noise sources. For a total of 52 trials of 30 seconds each per subject, our algorithm used pre-trained subject-specific decoders to decode attention using either M-NICA envelopes (as in [3]), or using MWF-envelopes. The average AAD accuracy using the clean speech envelopes was found to be 85.77%.

The boxplots in green represent the AAD performance when using M-NICA envelopes for attention decoding as in [3]. It can be seen that the performance has a large variance over the 12 source positions[4]. Also, the overall AAD performance shows a steady decrease as SNR drops. The boxplots in black represent the AAD performance for the new algorithm where

[4]Please note that the analysis over the different source positions applies only to the audio processing section of the algorithm, and not to the EEG signal processing, since we only have EEG data corresponding to the experiment conditions where the speech sources are at -90 and +90 degrees to the listener.

ISBN 978-0-9928626-7-1 © EURASIP 2017

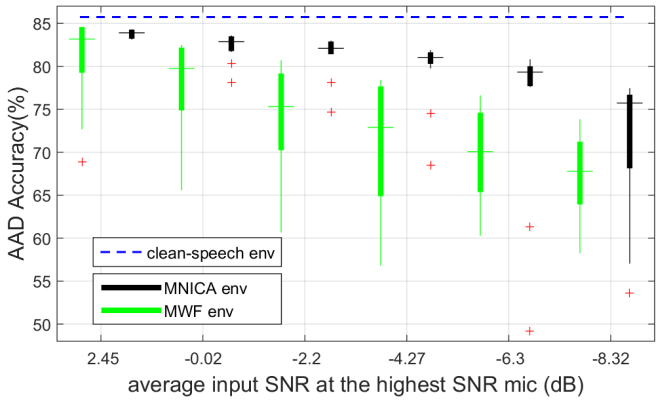

Fig. 3. A comparison of the proposed algorithm (MWF envelopes) with that of [3](M-NICA envelopes). The boxplots contain 12 points corresponding to average AAD accuracy (over 16 subjects) per source position.

MWF envelopes are used for attention detection. In this case, we observe that the variance over the 12 source positions has been greatly reduced, and the new algorithm ensures a significantly higher AAD accuracy, for almost all analyzed source positions, stable over the range of considered input SNRs. Except for low performance at some of the difficult positions ({-90,-30}, {-90,-60} and {-60,-30}) at low SNRs, and a relatively high variance for -8.32 dB SNR, the MWF envelopes consistently result in performance significantly better than, or equivalent to that of M-NICA envelopes.

V. Discussion

The results obtained for the proposed algorithm clearly show that using envelopes extracted from per-speaker MWFs result in a more stable AAD performance, even at negative input SNRs. Unlike [3] where M-NICA envelopes were used for AAD, we extract cleaner envelopes from the output of the MWFs and provide these to the AAD module instead. The quality of these envelopes, for most SNRs we analyzed is superior to that of M-NICA envelopes. This is reflected in a higher AAD accuracy and a smaller variance in figure 3. The latter also means that the AAD accuracy is largely independent of the speaker positions, up to a few outliers. This is a clear improvement in design, since in a real auditory environment, a neuro-steered noise suppression algorithm should be able to provide good AAD performance irrespective of the position of the attended and the competing speakers, as well as in low SNR conditions. Although AAD accuracy decreases with decreasing SNR, the decrease is less steep when using MWF envelopes compared to M-NICA envelopes.

VI. Conclusion

We have demonstrated that, in a two-speaker scenario, using speech envelopes extracted from the output signals of multiple per-speaker MWFs, results in stable attention detection over a range of SNRs, and speaker positions. Moreover, with this improved design the MWFs can learn speaker and noise statistics without the interruption of a reset during an attention switch as was the case in [12].

References

[1] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Signal Processing*, vol. 50, no. 9, pp. 2230 – 2244, Sep. 2002.

[2] R. Serizel, M. Moonen, B. Van Dijk, and J. Wouters, "Low-rank approximation based multichannel Wiener filter algorithms for noise reduction with application in cochlear implants," *IEEE/ACM Trans. Audio, Speech, and Language Processing*, vol. 22, no. 4, pp. 785–799, 2014.

[3] 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.

[4] J. O'Sullivan et al., "Attentional selection in a cocktail party environment can be decoded from single-trial EEG," *Cerebral Cortex*, vol. 25, no. 7, pp. 1697–706, 2015.

[5] 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. 046007, 2015.

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

[7] 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.

[8] E. B. Petersen, M. Wöstmann, J. Obleser, and T. Lunner, "Neural tracking of attended versus ignored speech is differentially affected by hearing loss," *Journal of neurophysiology*, vol. 117, no. 1, pp. 18–27, 2017.

[9] S. Debener, R. Emkes, M. De Vos, and M. Bleichner, "Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear," *Scientific Reports*, vol. 5, no. 16743, 2015.

[10] D. Looney, P. Kidmose, C. Park, M. Ungstrup, M. Rank, K. Rosenkranz, and D. Mandic, "The in-the-ear recording concept: User-centered and wearable brain monitoring," *IEEE Pulse*, vol. 3, no. 6, pp. 32–42, Nov 2012.

[11] A. Bertrand, "Distributed signal processing for wireless EEG sensor networks," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 6, pp. 923–935, 2015.

[12] N. Das, S. Van Eyndhoven, T. Francart, and A. Bertrand, "Adaptive attention-driven speech enhancement for EEG-informed hearing prostheses," in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*. IEEE, 2016, pp. 77–80.

[13] A. Bertrand and M. Moonen, "Energy-based multi-speaker voice activity detection with an ad hoc microphone array," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Dallas, Texas USA, March 2010, pp. 85–88.

[14] ——, "Blind separation of non-negative source signals using multiplicative updates and subspace projection," *Signal Processing*, vol. 90, no. 10, pp. 2877–2890, 2010.

[15] T. Francart, A. Van Wieringen, and J. Wouters, "APEX 3: a multi-purpose test platform for auditory psychophysical experiments," *Journal of Neuroscience Methods*, vol. 172, no. 2, pp. 283–293, 2008.

[16] N. Das, W. Biesmans, A. Bertrand, and T. Francart, "The effect of head-related filtering and ear-specific decoding bias on auditory attention detection," *Journal of Neural Engineering*, vol. 13, no. 5, p. 056014, 2016.

[17] B. N. Pasley, S. V. David, N. Mesgarani, A. Flinker, S. A. Shamma, N. E. Crone, R. T. Knight, and E. F. Chang, "Reconstructing speech from human auditory cortex," *PLoS Biol*, vol. 10, no. 1, p. e1001251, 2012.

[18] 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, 2012.

[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, no. 1, pp. 1–10, 2009.

ISBN 978-0-9928626-7-1 © EURASIP 2017