Hearing aid-controlled beamformer for binaural speech enhancement using a model-based approach

Kavalekalam, Mathew Shaji; Nielsen, Jesper Kjær; Christensen, Mads Græsbøll; Boldt, Jesper

Published in:
2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

DOI (link to publication from Publisher):
10.1109/ICASSP.2019.8683662

Publication date:
2019

HEARING AID-CONTROLLED BEAMFORMER FOR BINAURAL SPEECH ENHANCEMENT USING A MODEL-BASED APPROACH

Mathew Shaji Kavalekalam[1], Jesper K. Nielsen[1], Mads G. Christensen,[1] and Jesper B. Boldt[2]

[1]Audio Analysis Lab, CREATE, Aalborg University, Denmark {msk, jkn, mgc}@create.aau.dk
[2]GN Hearing A/S, DK 2750, Ballerup, Denmark {jboldt}@gnresound.com

ABSTRACT

The understanding of speech from a particular speaker in the presence of other interfering speakers can be severely degraded for a hearing impaired person. Beamforming techniques have been proven to be effective to improve the speech understanding in such scenarios. However, the number of microphones in a hearing aid (HA) is limited due to the space and power constraints present in the HA. In this paper, we propose to use an external device e.g., a microphone array, that can communicate with the HA to overcome this limitation. We propose a method to control this external device based on the look direction of the HA user. We show, by means of simulations, the robustness of the proposed method at very low SNRs in a reverberant scenario. Moreover, we have also conducted experiments that show the benefit of using this framework for binaural and monaural enhancement.

Index Terms— Speech enhancement, autoregressive models, beamforming.

1. INTRODUCTION

The ability of a hearing impaired person to understand speech is severely degraded in situations such as the cocktail party scenario, and this can subsequently lead to social isolation of the hearing impaired person. Therefore, it is crucial in such scenarios to perform speech enhancement. The speech enhancement algorithms can be broadly categorised into single and multi-channel methods [1]. In comparison to single channel algorithms which can exploit the temporal and spectral information, multi-channel algorithms are more effective as they can also exploit the spatial information. This property is useful in the cases where the speech and noise sources are spatially separated [2]. Beamforming, which forms a class of multi-channel enhancement algorithms, has been proven to be useful as it can exploit the spatial information to selectively attenuate the interferers in comparison to the speaker of interest [2]. The state of the art HAs are equipped with multiple microphones present at each ear which enables the use of beamforming algorithms. However, the space and power constraints on the HAs limit the number of microphones that can be used within a HA which consequently limits the performance of the beamformer within the HA to focus on the speaker of interest and attenuate the competing speakers. These limitations can be overcome by using an external device which can communicate wirelessly with the HA. In [3], it was investigated on how the speech intelligibility can be improved when the target speaker wears a microphone which picks up the speech signal uttered by the target speaker and transmits it wirelessly to the HA. The transmitted signal can then be binaurally saptialised according to the target speaker's location [4, 5, 6]. However, this solution has the constraint that the speaker of interest wears the microphone and that the listener is interested only in that speaker. In this paper, we try to relax this constraint by using a microphone array as the external device. Fig. 1 shows the scenario that we are interested in where the HA user is perhaps interested in listening to any of the speakers located at the table. As the external device is equipped with more microphones, it may be used to better exploit the spatial information for focusing on the speaker the HA user is listening to. However, a problem that can be encountered in this setup is that the external microphone array needs to know the direction of arrival of the source that the HA user is interested in. In this work, we propose a model-based method to estimate the direction of arrival via a collaboration between the HA and the external device. We use a model that we proposed in a previous paper [7] to represent the signal received at the HA as well as the external device. The estimated model parameters are subsequently used to estimate the direction of arrival by measuring the similarity between the model parameters. Using a model to represent the signal facilitates a low dimensional representation of the signal, which leads to less information being transmitted from the HA to the external device which is critical as power is a limiting factor in the HAs. To the best of author's knowledge, such an approach to control an external device based on the look direction of the HA user has not been done before.

The remainder of the paper is organised as follows. Section 2 introduces the setup, the signal model and the problem mathematically. The solution to the defined problem is then explained in Section 3 followed by the results and conclusion in Sections 4 and 5, respectively.

2. PROBLEM FORMULATION

2.1. Scenario of interest

Fig. 1 shows an example of the scenario of interest. This situation can be encountered when the HA user is participating in a meeting with colleagues or sitting at a dinner table with family or friends. From the figure, it can be seen that the HA user is listening to speaker A. It will be assumed in this work that the HA user is looking at the source of interest. From the perspective of the HA user, speaker A is the target whereas speakers B and C are interferers. The objective is to focus the beamformer present in the external device (uniform circular array (UCA) in this case) towards speaker A as the HA user is looking towards speaker A. This requires a communication link between the HA and the external device. To compute the data to be transmitted from the HA to the external device, a conventional beamformer focusing on the nose direction is used on the HA to form a

This work was supported by Innovations fund Denmark (Grant no: 99-2014-1).

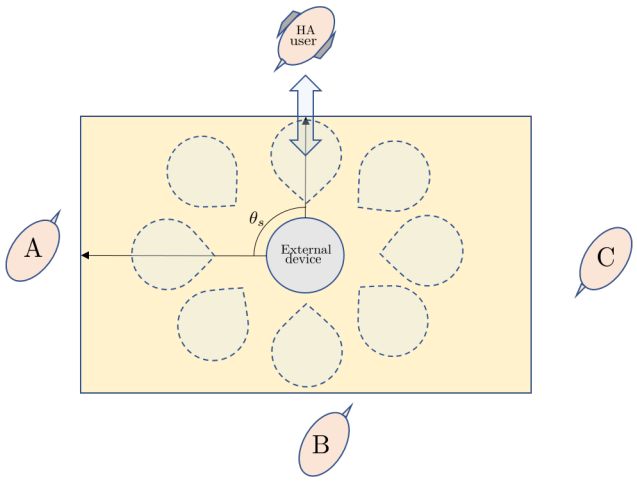

Fig. 1. Scenario of interest. HA user is listening to speaker A while speakers B and C are interferers.

preliminary enhanced signal which we denote as $x_{\text{HA}}(n)$. To estimate the source of interest, the UCA focuses its beam towards I different candidate directions ($I = 8$ in the figure) which are uniformly spaced along the azimuthal plane as shown in the figure. The set of beamformed signals from the different candidate directions are denoted as $\{x_{\text{CA}(\theta_i)}(n)\}_{i=1}^{I}$. Our objective here is to estimate θ_s (see Fig. 1) such that the UCA focuses on the source of interest, using the beamformed signal at the HA, $x_{\text{HA}}(n)$, and the set of beamformed signals from the candidate directions, $\{x_{\text{CA}(\theta_i)}(n)\}_{i=1}^{I}$.

2.2. Signal Model

We now introduce the signal model [7] that is used to represent the signals $x_{\text{HA}}(n)$ and $\{x_{\text{CA}(\theta_i)}(n)\}_{i=1}^{I}$. A frame of the signal of interest in the time domain denoted as $\mathbf{x} = [x(0) \dots x(N-1)]^T$ is modelled as a sum of U autoregressive (AR) processes as

$$\mathbf{x} = \sum_{u=1}^{U} \mathbf{c}_u \ . \tag{1}$$

Each of the AR process is expressed as a multivariate Gaussian [8, 9], i.e. ,

$$\mathbf{c}_u \sim \mathcal{N}(\mathbf{0},\, \sigma_u^2 \mathbf{Q}_u), \tag{2}$$

where σ_u^2 is the excitation variance and $\mathbf{Q}_u$ is the gain normalised covariance matrix. $\mathbf{Q}_u$ can be asymptotically approximated as a circulant matrix which can be diagonalised using the Fourier transform as [10]

$$\mathbf{Q}_u = \mathbf{F}\mathbf{D}_u\mathbf{F}^H \tag{3}$$

where $\mathbf{F}$ is the DFT matrix defined as $[\mathbf{F}]_{k,n} = \frac{1}{\sqrt{N}} \exp(j2\pi nk/N)$ $n, k = 0 \dots N-1$ and

$$\mathbf{D}_u = (\mathbf{\Lambda}_u^H \mathbf{\Lambda}_u)^{-1}, \qquad \mathbf{\Lambda}_u = \text{diag}(\sqrt{N}\mathbf{F}^H \begin{bmatrix} \mathbf{a}_u \\ \mathbf{0} \end{bmatrix}) \tag{4}$$

where $\mathbf{a}_u = [1,\, a_u(1), \dots a_u(P)]^T$ represents the vector of AR coefficients corresponding to the u^{th} AR process and P is the AR order. The diagonal entries of the matrix $\mathbf{D}_u$ contains the eigenvalues of the matrix $\mathbf{Q}_u$ and these correspond to the power spectral density (PSD) of the u^{th} gain normalised AR process. The set of U PSDs can be arranged as the columns of a spectral basis matrix $\mathbf{D}$ as

$$\mathbf{D} = [\mathbf{d}_1 \dots \mathbf{d}_u \dots \mathbf{d}_U] \tag{5}$$

where $\mathbf{d}_u = [d_u(1) \dots d_u(k) \dots d_u(K)]^T$ and $d_u(k)$ is the k^{th} diagonal element of $\mathbf{D}_u$. Using the above model explained by (1) and (2), a frame of the beamformed signal at the HA denoted as $\mathbf{x}_{\text{HA}} = [x_{\text{HA}}(0) \dots x_{\text{HA}}(N-1)]^T$ is expressed as

$$\mathbf{x}_{\text{HA}} = \sum_{u=1}^{U} \mathbf{c}_{\text{HA}_u} \tag{6}$$

where $\mathbf{c}_{\text{HA}_u} \sim \mathcal{N}(\mathbf{0},\, \sigma^2_{\text{HA}_u}\mathbf{Q}_u)$. Denoting $\boldsymbol{\sigma}_{\text{HA}} = [\sigma^2_{\text{HA}_1} \dots \sigma^2_{\text{HA}_U}]^T$, the PSD of the modelled signal at the HA can be represented as $\mathbf{D}\boldsymbol{\sigma}_{\text{HA}}$. Similarly a frame of the beamformed signal at the UCA for the ith candidate direction denoted as $\mathbf{x}_{\text{CA}(\theta_i)} = [x_{\text{CA}(\theta_i)}(0), \dots$ $x_{\text{CA}(\theta_i)}(N-1)]^T$ can be modelled as

$$\mathbf{x}_{\text{CA}(\theta_i)} = \sum_{u=1}^{U} \mathbf{c}_{\text{CA}(\theta_i)_u} \tag{7}$$

where $\mathbf{c}_{\text{CA}(\theta_i)_u} \sim \mathcal{N}(\mathbf{0},\, \sigma^2_{\text{CA}(\theta_i)_u}\mathbf{Q}_u)$. Denoting $\boldsymbol{\sigma}_{\text{CA}(\theta_i)} = [\sigma^2_{\text{CA}(\theta_i)_1} \dots \sigma^2_{\text{CA}(\theta_i)_U}]^T$, the PSD of the modelled signal at the UCA for the ith candidate direction is obtained as $\mathbf{D}\boldsymbol{\sigma}_{\text{CA}(\theta_i)}$. In the case of observing $V > 1$ frames, the PSD of the modelled signal for each frame can be arranged as columns to form a matrix $\mathbf{D}\mathbf{\Sigma}_{\text{HA}}$ or $\mathbf{D}\mathbf{\Sigma}_{\text{CA}(\theta_i)}$ where $\mathbf{\Sigma}_{\text{HA}} = [\boldsymbol{\sigma}_{\text{HA}}(1) \dots \boldsymbol{\sigma}_{\text{HA}}(V)]$ and $\mathbf{\Sigma}_{\text{CA}(\theta_i)} = [\boldsymbol{\sigma}_{\text{CA}(\theta_i)}(1) \dots \boldsymbol{\sigma}_{\text{CA}(\theta_i)}(V)]$. $\mathbf{\Sigma}_{\text{HA}}$ and $\mathbf{\Sigma}_{\text{CA}(\theta_i)}$ will be henceforth denoted as activation coefficients.

2.3. Mathematical problem

Our objective here is to estimate the direction of arrival of the speaker talking to the HA user relative to the UCA. In this paper, we propose to solve this problem by measuring the similarity between the beamformed signal received at the HA and the beamformed signal at the UCA for different candidate directions. This can be done in different ways. The first method, denoted as IS based method, is by using the spectral similarity to estimate θ_s as

$$\hat{\theta}_s = \underset{\theta_i \in \{\theta_i\}_{i=1}^{I}}{\arg\min} \quad d_{\text{IS}}(\mathbf{D}\mathbf{\Sigma}_{\text{HA}} | \mathbf{D}\mathbf{\Sigma}_{\text{CA}(\theta_i)}), \tag{8}$$

where $d_{\text{IS}}(\cdot,\cdot)$ is the sum of Itakura-Saito divergence [11] calculated over all the elements in the matrix. The second approach proposed here is to use the correlation between the estimated activation coefficients at the HA and UCA to estimate θ_s as

$$\hat{\theta}_s = \underset{\theta_i \in \{\theta_i\}_{i=1}^{I}}{\arg\max} \quad \text{corr}(\mathbf{\Sigma}_{\text{HA}} | \mathbf{\Sigma}_{\text{CA}(\theta_i)}) \tag{9}$$

where

$$\text{corr}(\mathbf{\Sigma}_{\text{HA}} | \mathbf{\Sigma}_{\text{CA}(\theta_i)}) = \frac{1}{UV} \sum_{v=1}^{V} \sum_{u=1}^{U} \left(\frac{\mathbf{\Sigma}_{\text{HA}}(u,v) - \mu_{\text{HA}}}{\epsilon_{\text{HA}}} \right) \times \left(\frac{\mathbf{\Sigma}_{\text{CA}(\theta_i)}(u,v) - \mu_{\text{CA}(\theta_i)}}{\epsilon_{\text{CA}(\theta_i)}} \right) \tag{10}$$

where $\mu_{\text{HA/CA}(\theta_i)}$ and $\epsilon_{\text{HA/CA}(\theta_i)}$ are the sample mean and standard deviation, respectively.

Algorithm 1 Main steps involved in the proposed framework

1: **while** new time-frames are available **do**
2: Apply beamforming in the HA as well as the UCA for different candidate directions to obtain $\mathbf{x}_{\text{HA}}$ and $\mathbf{x}_{\text{CA}(\theta_i)}$
3: Assuming the spectral basis matrix $\mathbf{D}$ is trained a priori, estimate $\mathbf{\Sigma}_{\text{HA}}$ and $\mathbf{\Sigma}_{\text{CA}(\theta_i)}$ using (11)
4: Transmit the estimated activation coefficients $\hat{\mathbf{\Sigma}}_{\text{HA}}$ from the HA to external device
5: Estimate θ_s using (8) or (9)
6: Use the beamformed signal from the UCA as a reference signal for performing binaural enhancement
7: **end while**

3. ESTIMATION OF THE MODEL PARAMETERS

As explained previously in Section 2.2, a frame of the signal is modelled as a sum of U AR processes with AR coefficients $\mathbf{a}_u$. In this work, the set of U AR coefficients are trained a priori using a standard vector quantisation technique used in speech coding applications. During the training stage, a speech codebook is first computed using the generalised Lloyd algorithm (GLA) [12, 8]. The speech codebook contains AR coefficients corresponding to the spectral envelopes of speech. During the training process, linear prediction coefficients (converted into line spectral frequency coefficients) are extracted from windowed frames, obtained from the training signal and passed as input to the vector quantiser. Once the speech codebook is created, the spectral envelopes corresponding to the AR coefficients $(\{\mathbf{a}_u\}_{u=1}^{U})$ are computed and arranged as columns of the spectral basis matrix $\mathbf{D}$ as explained by (4) and (5). Given the observed data and the spectral basis matrix $\mathbf{D}$, it has been shown in [7, 13] that the maximum likelihood estimation of the activation coefficients corresponds to minimising the IS divergence between the periodogram of the observed signal and the modelled PSD. Since there is no closed form solution for this, it is generally estimated iteratively using the multiplicative update (MU) rule [14] as

$$\hat{\mathbf{\Sigma}}_{\text{HA/CA}(\theta_i)} \leftarrow \hat{\mathbf{\Sigma}}_{\text{HA/CA}(\theta_i)} \odot \frac{\mathbf{D}^T\big((\mathbf{D}\hat{\mathbf{\Sigma}}_{\text{HA/CA}(\theta_i)})^{[-2]} \odot \mathbf{\Phi}_{\text{HA/CA}(\theta_i)}\big)}{\mathbf{D}^T(\mathbf{D}\hat{\mathbf{\Sigma}}_{\text{HA/CA}(\theta_i)})^{[-1]}}, \tag{11}$$

where $\mathbf{\Phi}_{\text{HA/CA}(\theta_i)}$ contains the periodogram of frames of the signals arranged as columns, $\odot$ is the element-wise product, $(\cdot)^{[-1]}$ is the element-wise inverse and the division is an element-wise division. The spectral basis matrix along with the estimated activation coefficients can be utilised as shown in (8) and (9) to estimate the direction of arrival to control the beam pattern of the UCA. It should be noted that the method proposed here to estimate θ_s requires only the transmission of $\hat{\mathbf{\Sigma}}_{\text{HA}}$ from the HA to the external device. This is generally much less than the amount of signal samples. The proposed algorithm is summarised in Algorithm 1.

4. EXPERIMENTS

4.1. Experimental setup

This section describes the experimental results obtained for the proposed method. The setup used for carrying out the experiments will be explained in this section followed by the results. Fig. 2 shows a portion of the experimental setup in a room of dimensions $12 \times 6 \times 6$ m with a reverberation time of 0.4 seconds. The room impulse responses were generated using [15]. In this figure, the HA user is trying to listen to the target speaker denoted by a green dot. Along with the speaker of interest, there are 3 other interferers located around the table as shown in the figure. The HA user is simulated with a ULA of 5 microphones with a span of 0.24 m which is a typical ear to ear distance of a human head. The external device that we have considered here is a UCA of radius of 0.13 m with 10 microphones. The signals used for testing consisted of speech signals spoken by two males and two females taken from the CHIME database [16]. A codebook of size 64 entries was generated using the GLA using speech from the EUROM database [17]. We have used different databases for testing and training to test the robustness of the proposed method against the mismatches that maybe encountered in a practical scenario. The parameters used for the experiments have been summarised in Table 1. The beamformed signals at the HA and the external device can be obtained using any of the conventional beamforming algorithms. For the experiments conducted in this section, we have used the robust Capon beamforming (RCB) [18, 19], as this method has been shown to be robust to reverberation and uncertainty in the steering vector [19]. The number of candidate directions I has been chosen to be 8 in the experiments. It should be noted that speakers are not constrained to be at the candidate positions as the RCB takes into account uncertainties in the steering vector using the parameter ϵ [18, eq. (14)] which was chosen to be 3.5 in our experiments. The experiments we have conducted to validate the robustness of the proposed method is shown in the following section.

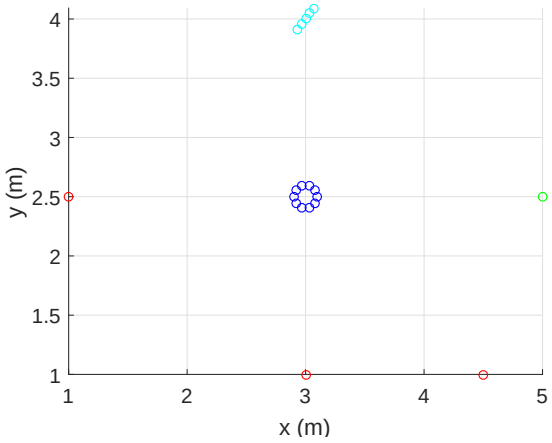

Fig. 2. Figure shows the experimental setup where the green dot indicates the source of interest and the red dot indicates the interferers.

4.2. Experimental Results

In this section we evaluate the accuracy of the proposed method in the experimental setup explained above. In addition to the case

Table 1. Parameters used for the experiments

Parameters	
sampling frequency	8000 Hz
Frame size (N)	200
Frame overlap	50%
AR order (P)	14
U	64
MU iterations	50
I	8

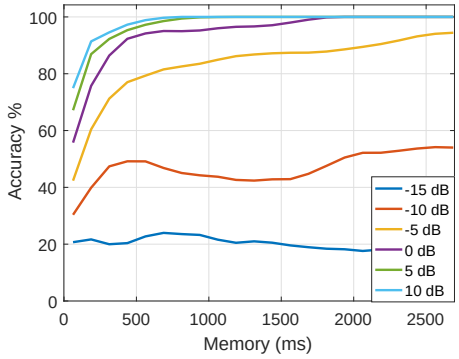

Fig. 3. Percentage of correctly detected direction for the correlation based method as a function of the SNR and memory.

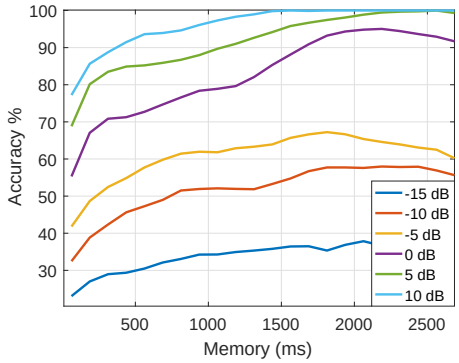

Fig. 4. Percentage of correctly detected direction for the IS based method as a function of the SNR and memory.

shown in Fig. 2, we have varied the source of interest, so the results shown in this section are averaged over all the speakers (4 in this case). In addition to the interferers, we also add spherically isotropic babble ambient noise generated using the implementation in [20]. Figures 3 and 4 show the average accuracy obtained over all the speakers for 10 iterations per speaker as a function of SNR and the memory for correlation based and IS based methods, respectively. Memory here is related to the number of frames V (used for the computation of the model parameters) that is needed to estimate θ_s. It can be seen from the figure that the correlation based method converges to 100 % accuracy for SNRs 0 dB and above when a memory greater than 1700 ms is used for computation of the model parameters, whereas the IS based method converges to 100 % accuracy for SNRs 5 dB and above when a memory of 2200 ms is used. It should be noted that the experiments conducted in this section, assumed the positions of the HA user, the target speaker and the interferers to be stationary. However, in practical scenarios it may be useful to update the result at much finer time scale, as the HA user may continously change the look direction. The influence of memory has also been investigated in figures 3 and 4 and it can be seen that as the SNR increases the memory required for the proposed method to obtain certain accuracy decreases, e.g., to reach 80 % accuracy, it requires a memory of approximately 100, 240 and 600 ms for SNRs 10, 0 and -5 dB, respectively for the correlation based method.

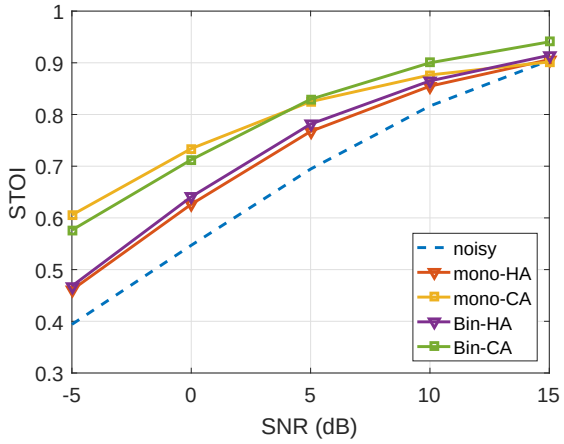

Fig. 5. Average STOI score obtained for the left and right channels for binaural and monaural configurations.

4.2.1. Enhancement performance

In this section, we show an example of how this setup can be used for monaural/binaural enhancement in HAs. One option is to wirelessly transmit the beamformed signal at the external device to the HA and play back that signal at both the ears. Playing back the monaural signal, however, may lead to distortion of binaural cues. Thus in this paper, we also perform binaural speech enhancement, where we consider the signals received at the left and rightmost microphones as the binaural noisy signal. To perform the enhancement we consider the binaural enhancement framework proposed in [21], which is based on the MMSE criterion. This method applies a common gain on the left and right channels which leads to the preservation of the binaural cues. The common gain applied in this case requires the estimation of the speech/noise statistics [21, eq. (17)]. In this work, we propose to use the beamformed signal at the UCA from candidate direction to be used as the reference signal to estimate the clean speech statistics. Fig. 5 shows the averaged short time objective intelligibilty (STOI) [22] scores for the left and right channels obtained for the different configurations. The beamformed signals at the HA and UCA are denoted as mono-HA and mono-CA, respectively. The binaural enhancement method where we use the beamformed signals at the HA and UCA to estimate the clean speech statistics is denoted as Bin-HA and Bin-CA, respectively. It can be seen that using the beamformed signal at the UCA shows an improvement in both the binaural and monaural configurations.

5. CONCLUSION

In this paper we have proposed a framework for improving the speech understanding for HA users in the presence of multiple interferers. The proposed system consisted of using an external microphone array whose beam-pattern is controlled by the look direction of the HA user using a model-based approach. The robustness of the proposed method at very low SNRs in a reverberant scenario has been shown by the means of simulations. Moreover, the benefits of using the external device in addition to the HA for performing binaural enhancement has been shown using an objective measure for intelligibility.

6. REFERENCES

[1] J Benesty, S Makino, and J Chen, "Speech enhancement, ser. signals and communication technology," 2005.

[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] M. S. Lewis, C. C. Crandell, M. Valente, and J. E. Horn, "Speech perception in noise: Directional microphones versus frequency modulation (fm) systems," *Journal of the American Academy of Audiology*, vol. 15, no. 6, pp. 426–439, 2004.

[4] G. Courtois, P. Marmaroli, M. Lindberg, Y. Oesch, and W. Balande, "Implementation of a binaural localization algorithm in hearing aids: specifications and achievable solutions," in *Audio Engineering Society Convention 136*. Audio Engineering Society, 2014.

[5] J. M. Kates, K. H. Arehart, R. K. Muralimanohar, and K. Sommerfeldt, "Externalization of remote microphone signals using a structural binaural model of the head and pinna," *The Journal of the Acoustical Society of America*, vol. 143, no. 5, pp. 2666–2677, 2018.

[6] M. Farmani, M. S. Pedersen, and J. Jensen, "Sound source localization for hearing aid applications using wireless microphones," in *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*. IEEE, 2018, pp. 455–459.

[7] M. S. Kavalekalam, J. K. Nielsen, L. Shi, M. G. Christensen, and J. Boldt, "Online parametric nmf for speech enhancement," in *Proceedings of the European Signal Processing Conference*, 2018.

[8] S. Srinivasan, J. Samuelsson, and W. B. Kleijn, "Codebook-based bayesian speech enhancement for nonstationary environments," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 15, no. 2, pp. 441–452, 2007.

[9] J. K. Nielsen, M. S. Kavalekalam, M. G. Christensen, and J. B. Boldt, "Model-based noise psd estimation from speech in non-stationary noise," in *IEEE International Conference on Acoustics, Speech and Signal Processing. Proceedings*, 2018.

[10] R. M Gray et al., "Toeplitz and circulant matrices: A review," *Foundations and Trends® in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006.

[11] F. Itakura, "Analysis synthesis telephony based on the maximum likelihood method," in *The 6th international congress on acoustics, 1968*, 1968, pp. 280–292.

[12] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Communications*, vol. 28, no. 1, pp. 84–95, 1980.

[13] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis," *Neural computation*, vol. 21, no. 3, pp. 793–830, 2009.

[14] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in neural information processing systems*, 2001, pp. 556–562.

[15] E. A. P. Habets, "Room impulse response generator," *Technische Universiteit Eindhoven, Tech. Rep*, vol. 2, no. 2.4, pp. 1, 2006.

[16] J. Barker, R. Marxer, E. Vincent, and S. Watanabe, "The third 'chime' speech separation and recognition challenge: Dataset, task and baselines," *IEEE 2015 Automatic Speech Recognition and Understanding Workshop*, 2015.

[17] D. Chan, A. Fourcin, D. Gibbon, B Granstrom, et al., "Eurom-a spoken language resource for the eu," in *Proceedings of the 4th European Conference on Speech Communication and Speech Tecnology, Eurospeech'95*, 1995, pp. 867–880.

[18] J. Li, P. Stoica, and Z. Wang, "On robust capon beamforming and diagonal loading," *IEEE transactions on signal processing*, vol. 51, no. 7, pp. 1702–1715, 2003.

[19] Y. Zhao, J. R. Jensen, M. G. Christensen, S. Doclo, and J. Chen, "Experimental study of robust beamforming techniques for acoustic applications," in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2017 IEEE Workshop on*. IEEE, 2017, pp. 86–90.

[20] E. A. P. Habets and S. Gannot, "Generating sensor signals in isotropic noise fields," *The Journal of the Acoustical Society of America*, vol. 122, no. 6, pp. 3464–3470, 2007.

[21] G. Enzner, M. Azarpour, and J. Siska, "Cue-preserving mmse filter for binaural speech enhancement," in *Acoustic Signal Enhancement (IWAENC), 2016 IEEE International Workshop on*. IEEE, 2016, pp. 1–5.

[22] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time–frequency weighted noisy speech," *IEEE Trans. Audio, Speech, and Language Process.*, vol. 19, no. 7, pp. 2125–2136, 2011.