Model-based Beamforming for Wearable Microphone Arrays

Emilie d'Olne, Alastair H. Moore, and Patrick A. Naylor
Department of Electrical and Electronic Engineering, Imperial College London, UK
{emilie.dolne16, alastair.h.moore, p.naylor}@imperial.ac.uk

***Abstract*—Beamforming techniques for hearing aid applications are often evaluated using behind-the-ear (BTE) devices. However, the growing number of wearable devices with microphones has made it possible to consider new geometries for microphone array beamforming. In this paper, we examine the effect of array location and geometry on the performance of binaural minimum power distortionless response (BMPDR) beamformers. In addition to the classical adaptive BMPDR, we evaluate the benefit of a recently-proposed method that estimates the sample covariance matrix using a compact model. Simulation results show that using a chest-mounted array reduces noise by an additional 1.3 dB compared to BTE hearing aids. The compact model method is found to yield higher predicted intelligibility than adaptive BMPDR beamforming, regardless of the array geometry.**

***Index Terms*—wearable microphone arrays, binaural beamforming, compact model, hearing aids**

I. Introduction

Classical single-channel speech enhancement aims to exploit time-frequency properties of a signal to reduce its noise level without introducing excessive distortion in the desired speech. Using microphone arrays provides information about the spatial properties of a signal, thus allowing for an additional degree of enhancement through spatial filtering or *beamforming* [1]. Beamforming can be made binaural by processing the signals from the microphone array using two beamformers with outputs for the left and right ears to create a spatialisation of the output [2]. A widely used beamformer is the minimum power distortionless response (MPDR) beamformer which minimises its output power while satisfying a unity gain constraint in a desired look direction [3], [4]. To determine the beamformer filter weights, the MPDR beamformer estimates the spatial coherence of the sound scene at the microphones of the array through the sample covariance matrix (SCM). This matrix is often obtained using theoretical definitions of a spatial coherence function [5]–[7]. When such definitions cannot be used, e.g. in non-stationary noise, a common approach is to adaptively estimate the SCM from the signal samples. The number of estimated matrix elements grows quadratically with the number of microphones, leading to slow convergence for large arrays. Rather than estimating the entire SCM, a model-based estimation was proposed in [8] which constructs the SCM as the weighted sum of plane-wave, isotropic noise, and sensor noise terms. The method showed promising results with behind-the-ear (BTE) hearing aids (HAs) in [8].

The growing availability of connected wearables such as smart watches and glasses has broadened the range of array geometries to consider when designing beamformers. Many devices already incorporate microphones for hands-free communication, voice control, or augmented reality [9], making them natural candidates when creating assisted listening devices [10]. Wearable arrays present the advantage of being potentially spread over the entire body, whereas classical HAs restrict microphone placement usually to an area around the ears. Wearable devices thus benefit from an increased spatial diversity, and it was found in [11] that an MVDR beamformer applied to a body-worn array yielded better results than when applied to an array concentrated at the listener's ears. Moreover, the discomfort of HAs has been reported in [12] to sometimes reduce their adoption; this could be addressed by creating devices that rely on accessories already worn by the user, e.g. glasses, hats, jewellery, etc. Additionally, the design of new forms of assisted listening devices could benefit special needs users. For example, it was suggested in [13] that dementia patients could benefit from personalised devices, and in [14] a speech enhancement system for motor skill impaired patients was developed using an array mounted on glasses.

Only a few studies have so far explored the combination of modern beamformers and the varying geometries of wearable arrays. In [15], a speech acquisition device using speech detection and an MVDR beamformer was implemented on glasses only, while the comparison of array locations for beamforming in [11] was limited to the monaural MVDR beamformer. In this paper, we are interested in studying the performance of binaural beamformers using wearable arrays, as this has not yet been widely examined in the signal processing literature for the binaural case. We thus explore the performance of adaptive and compact-model [8] binaural MPDR beamformers across various wearable array geometries. The paper is structured as follows: Sec. II provides relevant information on SCM estimation and binaural beamforming; Sec. III describes the design and results of simulation experiments; Sec. IV discusses results; and Sec. V draws conclusions.

II. Background

A. Signal model

Given Q source signals arriving at an array of M microphones, the STFT of the signal observed at the mth

This work was supported by the UK Engineering and Physical Sciences Research Council [grant number EP/S035842/1].

microphone is given by

$$y_m(k,\ell) = \sum_{q=1}^{Q} h_{q,m}(k) s_q(k,\ell) + v_m(k,\ell) \tag{1}$$

where ℓ and k are respectively the time and frequency indices, $s_q(k,\ell)$ is the qth source signal (target or interferer), $h_{q,m}(k)$ is the transfer function, assumed stationary, between the qth source and the mth microphone, and $v_m(k,\ell)$ is sensor noise. The STFT analysis window is assumed to be large relative to the impulse response of the $h_{q,m}$ so that the multiplicative transfer function assumption holds [16], [17]. Stacking microphone signals into a vector form gives (omitting k for clarity)

$$\mathbf{y}(\ell) = \sum_{q=1}^{Q} \mathbf{h}_q s_q(\ell) + \mathbf{v}(\ell) = \mathbf{H}\mathbf{s}(\ell) + \mathbf{v}(\ell) \tag{2}$$

where $\mathbf{y}(\ell) = [y_1(\ell), \ldots, y_M(\ell)]^T$, with $\mathbf{h}_q$ and $\mathbf{v}(\ell)$ defined similarly, $\mathbf{s}(\ell) = [s_1(\ell), \ldots, s_Q(\ell)]^T$, and $\mathbf{H}$ is a $M \times S$ matrix whose columns contain $\mathbf{h}_q$, $q = 1, \ldots, Q$. Separating $\mathbf{H}$ into direct-path, $(\cdot)^{(d)}$, and reverberant, $(\cdot)^{(r)}$, terms gives

$$\mathbf{y}(\ell) = \mathbf{H}^{(d)}\mathbf{s}(\ell) + \mathbf{H}^{(r)}\mathbf{s}(\ell) + \mathbf{v}(\ell). \tag{3}$$

Then, following [8] by assuming sources in the far-field, obeying W-disjoint orthogonality such that only one source is dominant in each time-frequency bin [18], and considering that the sum of all reflections results in a diffuse isotropic noise field, (3) is rewritten in terms of a signal model

$$\dot{\mathbf{y}}(\ell) = \mathbf{a}(\Omega(\ell))\dot{s}_{q(\ell)}(\ell) + \boldsymbol{\gamma}(\ell) + \mathbf{v}(\ell) \tag{4}$$

where $q(\ell)$ is the index of the dominant source, $\dot{s}_{q(\ell)}(\ell)$, in the $(k,\ell)^{th}$ bin as measured at a reference microphone, $\mathbf{a}(\Omega(\ell))$ is the relative transfer function (RTF) relating the reference microphone to the rest of the array for a source direction-of-arrival (DOA) $\Omega(\ell)$, and $\boldsymbol{\gamma}(\ell)$ is the diffuse isotropic noise.

B. MPDR beamforming

For each k, $w_m(\ell)$ is the beamformer coefficient applied to the mth microphone, and the beamformer output is

$$z(\ell) = \mathbf{w}^H(\ell)\mathbf{y}(\ell) \tag{5}$$

where $\mathbf{w}^H(\ell) = [w_1^*(\ell), \ldots, w_M^*(\ell)]$, with $(\cdot)^H$ the Hermitian transpose. In MPDR beamforming, the weights satisfy

$$\underset{\mathbf{w}(\ell)}{\operatorname{argmin}} \ \mathbf{w}^H(\ell)\mathbf{R}_{\mathbf{y}}(l)\mathbf{w}(\ell) \qquad \text{s.t.} \ \ \mathbf{w}^H(\ell)\mathbf{d} = 1 \tag{6}$$

where $\mathbf{d}$ is the steering vector and $\mathbf{R}_{\mathbf{y}}$ is the covariance matrix defined as $\mathbf{R}_{\mathbf{y}} = \mathbb{E}[\mathbf{y}\mathbf{y}^H]$, with $\mathbb{E}$ the expectation operator. This has a well-known solution given in [19] as

$$\mathbf{w}(\ell) = \frac{\mathbf{R}_{\mathbf{y}}^{-1}(\ell)\mathbf{d}}{\mathbf{d}^H\mathbf{R}_{\mathbf{y}}(\ell)\mathbf{d}}. \tag{7}$$

1) Adaptive MPDR beamformer (AMB)

In adaptive MPDR beamforming, the $M \times M$ SCM, $\mathbf{R}_{\mathbf{y}}(\ell)$, can be recursively estimated from samples as

$$\hat{\mathbf{R}}_{\mathbf{y}}(\ell) = \alpha\hat{\mathbf{R}}_{\mathbf{y}}(\ell-1) + (1-\alpha)\mathbf{y}(\ell)\mathbf{y}^H(\ell) \tag{8}$$

where α controls the recursive smoothing.

2) Compact model MPDR beamformer (CMMB)

The method in [8] uses the signal model in (4) to calculate $\mathbf{R}_{\dot{\mathbf{y}}}(\ell)$ as

$$\mathbf{R}_{\dot{\mathbf{y}}}(\ell) = \sigma_{\mathrm{a}}(\ell)\mathbf{R}_{\mathbf{a}}(\Omega(\ell)) + \sigma_{\gamma}(\ell)\mathbf{R}_{\boldsymbol{\gamma}} + \sigma_{\mathrm{v}}(\ell)\mathbf{R}_{\mathbf{v}} \tag{9}$$

where $\Omega(l)$ is the DOA of the plane-wave component, and $\sigma_{\mathrm{a}}(\ell)$, $\sigma_{\gamma}(l)$, and $\sigma_{\mathrm{v}}(\ell)$ are the power of the plane-wave, diffuse noise, and sensor noise components, respectively. Thus, the compact model estimates 4 parameters, found by solving

$$\underset{\Omega(\ell),\sigma_{\mathrm{a}}(\ell),\sigma_{\gamma}(\ell),\sigma_{\mathrm{v}}(\ell)}{\operatorname{argmin}} ||\hat{\mathbf{R}}_{\mathbf{y}}(\ell) - \mathbf{R}_{\dot{\mathbf{y}}}(\ell)||_F \tag{10}$$

where $||\cdot||_F$ denotes the Frobenius norm.

C. Binaural MPDR beamforming

In binaural MPDR beamforming, two outputs are produced using two sets of filters, $\mathbf{w}_l(\ell)$ and $\mathbf{w}_r(\ell)$, designed for each ear [2], such that (5) and (7) become

$$\boldsymbol{z}(\ell) = [\mathbf{w}_l(\ell), \mathbf{w}_r(\ell)]^H\mathbf{y}(\ell) \tag{11}$$

$$\mathbf{w}_b(\ell) = \frac{\mathbf{R}_{\mathbf{y}}^{-1}(\ell)\mathbf{d}_b}{\mathbf{d}_b^H\mathbf{R}_{\mathbf{y}}(\ell)\mathbf{d}_b} \quad b \in \{l, r\} \tag{12}$$

where l and r indicate the left and right devices respectively, and $\mathbf{d}_l$ and $\mathbf{d}_r$ are the steering vectors defined for the left and right reference microphones. This beamforming can preserve the spatial cues of the target signal [20].

III. Experiments

The performance of binaural MPDR beamforming using SCMs estimated from the adaptive method of (8) and the compact-model method of (10) is evaluated for various microphone array geometries. The steering vectors $\mathbf{d}_{l,r}$ are set to $\mathbf{d}_{l,r} = \mathbf{a}_{l,r}(0)$, where $\mathbf{a}_{l,r}(0)$ are the RTFs, assumed known, from left and right reference microphones to the considered array. The selected reference microphones are located just outside the ear-canals of a mannequin, referred to as left and right 'in-ear' microphones. Diagonal loading is applied to the covariance matrices to limit their condition numbers to ≤ 100 in order to improve the robustness of the beamformers [8], [21]. The time constant in (8) is chosen empirically to be 50 ms, and (10) is solved as the best ordinary least square solution from among the 24 plane-wave DOAs as will be presented in Sec. III-B. The system operates at a sampling rate

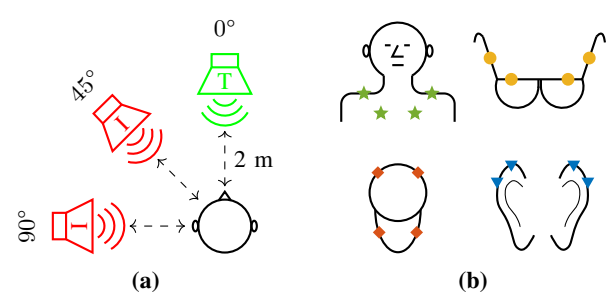

Figure 1: (a) Scenario setup. **(b)**: Array geometries, clockwise from top left: chest, glasses, BTE hearing aids, baseball cap.

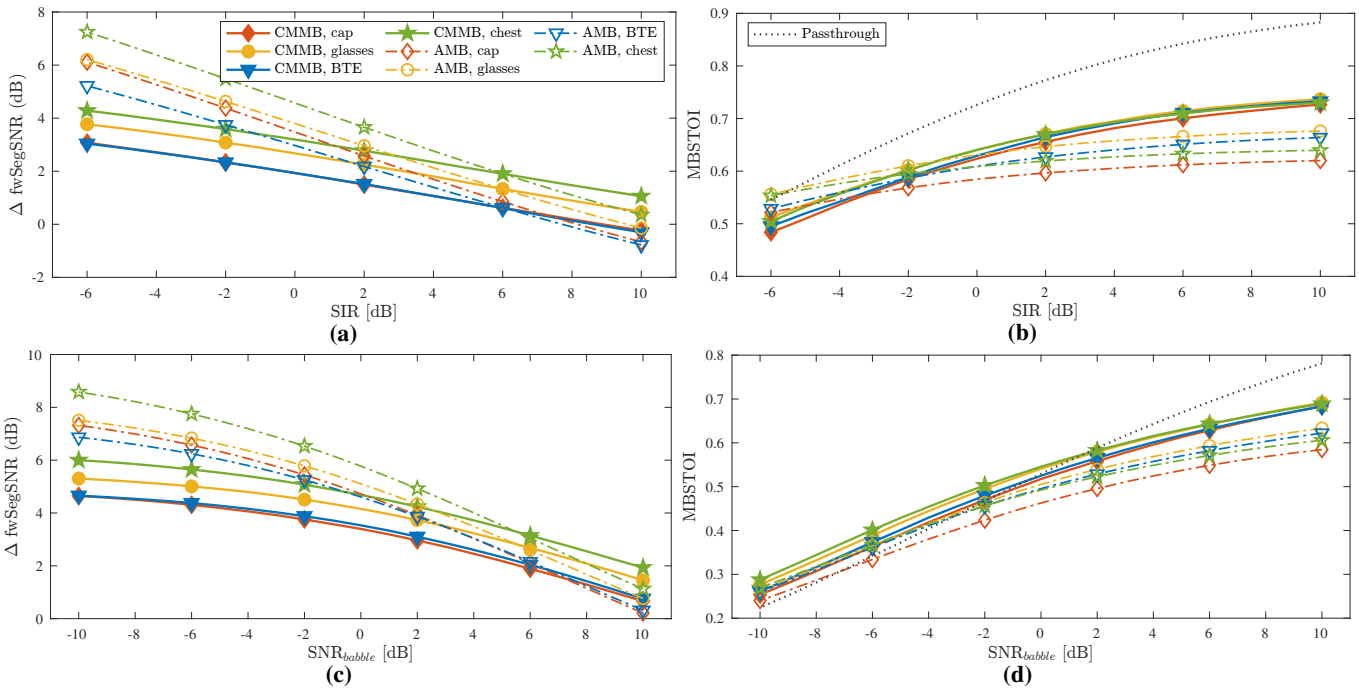

Figure 2: Comparison of MPDR beamformers using adaptive and compact-model SCM estimation for various array geometries. **(a)**, **(b)**: For $\text{SNR}_{\text{babble}}$ = 20 dB and varying SIR. **(c)**, **(d)**: For SIR = 20 dB and varying $\text{SNR}_{\text{babble}}$.

of 20 kHz, and the STFT uses Hamming-windowed frames of 16 ms with 50 % overlap.

A. Experimental setup

The simulation experiments follow a competing dialogue scenario pictured in Fig. 1(a): a target male speaker is presented in front of the listener, while female and male interfering speakers are located on their left, at 45° and 90° respectively. The interfering speakers have equal signal level and are not simultaneously active. Target and interfering speakers can be, however, simultaneously active. White Gaussian sensor noise is added with reverberant target signal-to-sensor-noise ratio ($\text{SNR}_{\text{sensor}}$) fixed at 20 dB. A diffuse babble noise field is simulated by generating babble noise originating from 24 equally spaced azimuth angles, ensuring equal direct-path power from all directions. The speech and noise levels are computed respectively as the mean active level [22] and mean power at the two reference microphones.

B. Data and body-related transfer functions (BRTF)

The source speech signals are obtained through concatenation of anechoic speech from the TIMIT database [23]. The transfer functions used to generate microphone signals and to obtain steering vectors are the body-related transfer functions (BRTF) presented in [11]. The database contains BRTFs for 80 omnidirectional microphones mounted on a mannequin, as well as for BTE hearing aids and various accessories. The BRTFs are available for 24 source azimuth angles, with the sources placed 2 m away from the microphones. The measurements were obtained using a linear sweep in an acoustically treated room with reverberation time of approximately 200 ms, and are sampled with 24 bits at a rate of 48 kHz.

C. Performance measures

To compare the performance of the beamformers for the noise reduction task, the frequency-weighted segmental SNR (fwSNRseg) is computed in frames of length L, such that in a given frame

$$\text{fwSNRseg}_b \triangleq 10\log_{10}\Big(\frac{\sum_{i=1}^{L}\tilde{s}_b^*(i)^2}{\sum_{i=1}^{L}(\tilde{z}_b(i)-\tilde{s}_b^*(i))^2}\Big)\ \text{dB} \qquad (13)$$

where $b \in \{l, r\}$, i represents the sample index, $\tilde{s}_b^*(i)$ is the frequency-weighted direct-path target speech signal at a reference microphone, and $\tilde{z}_b(i)$ is the frequency-weighted beamformer output for the corresponding reference microphone. The fwSNRseg is then computed as the mean over all frames using the A-weighting provided in Voicebox [24].

The MBSTOI measure [25] is used to quantify the performance in terms of predicted intelligibility, as it has been shown to predict perceived binaural intelligibility. In the MBSTOI computation, the reference signals are the direct-path target speech signals recorded at the in-ear microphones, and the noisy signals are the outputs of the beamformers.

D. Experiment 1: Array geometries

Four array geometries, represented schematically in Fig. 1(b), are considered in this example, each containing M=4 microphones mounted respectively on the chest of a mannequin, on glasses, on a baseball cap, and on BTE HAs. In Fig. 2, the input reverberant target signal-to-babble noise ratio ($\text{SNR}_{\text{babble}}$) is varied while the signal-to-interference ratio (SIR) is fixed at 20 dB, and vice-versa. Listening examples are given in [26].

For each array geometry, Fig. 2(a) shows the fwSNRseg improvement (ΔfwSNRseg) of the left ear beamformers with respect to a passthrough signal simulated at the left in-ear microphone as a function of SIR. At negative SIR, the AMB performs better than the CMMB for all arrays. The relative advantage of the AMB decreases with increasing SIR, ultimately dropping below the CMMB at approximately 6 dB SIR. The rate of change of ΔfwSNRseg is always smaller for the CMMB than for the AMB. The chest array gives the highest ΔfwSNRseg for both beamformers and across all SIRs, with a mean improvement of 1.6 dB (AMB) or 1.3 dB (CMMB) over a BTE HA. Figure 2(b) plots the MBSTOI measure for all beamformers and array geometries as a function of SIR. The results for the passthrough signal simulated at the in-ear microphones are also shown for reference. For SIRs higher than -2 dB, the CMMB outperforms the AMB for all geometries. For the AMB, the glasses yield the highest MBSTOI scores, while the CMMB shows little variations in scores across array geometries. Overall, the passthrough signals show the highest MBSTOI scores across all SIRs. Investigation of this result using the monaural STOI [27] at the left in-ear microphone does not show a similar advantage. The high performance of the passthrough signal is therefore assumed to be associated with binaural cues in the signal, and is further discussed in Sec. IV. Figure 2(c) shows the ΔfwSNRseg for varying $\mathrm{SNR}_{\mathrm{babble}}$. The plots follow similar patterns, with the chest array giving the highest ΔfwSNRseg and the CMMB outperforming the AMB for SNRs above 6 dB. Figure 2(d) shows the MBSTOI scores for varying $\mathrm{SNR}_{\mathrm{babble}}$. The CMMB slightly outperforms the AMB for all SNRs and geometries, and the passthrough model for negative SNR. For positive SNR, the passthrough system yields the best performance, as expected and discussed in Sec. IV.

E. Experiment 2: Hybrid arrays

In this experiment, the effect of combining BTE HAs with the arrays presented in Fig. 1(b) is investigated. Assuming clock synchronisation, the BTE HAs are complemented by the chest, glasses, and cap arrays respectively, giving a total of $M = 8$ microphones per array. Figure 3 investigates the advantage of using hybrid arrays (chest_+, $\mathrm{glasses}_+$, and cap_+) over the pair of BTE hearing aids. Figure 3(a) shows the improvement in ΔfwSNRseg at the left ear, and Fig. 3(b) shows the improvement in MBSTOI, both as a function of SIR with $\mathrm{SNR}_{\mathrm{babble}}$ fixed at 20 dB.

Figure 3(a) shows that hybrid arrays lead to an increase in ΔfwSNRseg for all geometries and beamformers. For both beamforming methods, the chest_+ array shows the highest improvements, followed by the cap_+ and the $\mathrm{glasses}_+$. For the cap_+ and chest_+ arrays, the CMMB yields higher ΔfwSNRseg improvements than the AMB for SIR>0 dB. For the $\mathrm{glasses}_+$ array, the same behaviour occurs when SIR>6 dB. On average, the largest ΔfwSNRseg improvement is obtained by the CMMB using the chest_+ array, with mean improvement of 2.9 dB over the considered SIR range. For ΔMBSTOI scores, Fig. 3(b) shows that MBSTOI improvements are monotonically decreasing with increasing SIR. For all hybrid arrays, the CMMB yields higher improvements compared to the AMB. Moreover, the AMB leads to decreases in MBSTOI for positive SIR. Overall, the highest improvements are obtained by the CMMB using the cap_+ and chest_+ arrays, giving mean improvements of 6.2 % and 6.7 %, respectively, over the CMMB using BTE HAs.

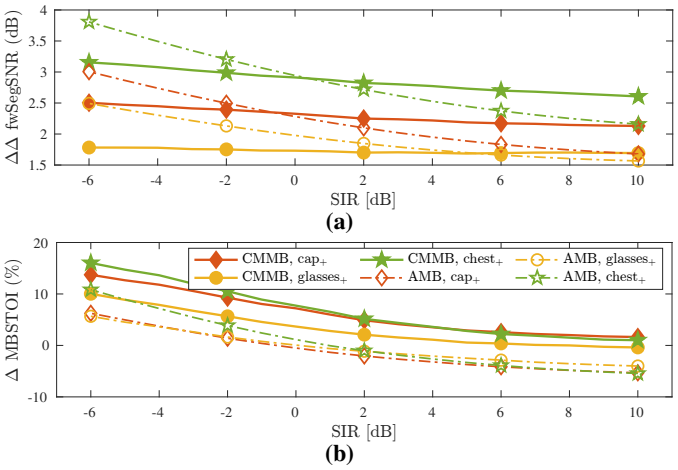

Figure 3: Improvement, Δ, in **(a)**: ΔfwSNRseg (dB), **(b)**: MBSTOI (%), when combining BTE HAs with other arrays, relative to the BTE HAs, for $\mathrm{SNR}_{\mathrm{babble}}$ = 20 dB.

IV. Discussion

Experiment 1: Array geometries. Results show that the MPDR using adaptive SCM estimation performs better for noise reduction at low input SNR (and SIR) than when using the compact-model SCM estimation, but the opposite is true at high SNR. This is because, at low SNR, the adaptive SCM is dominated by noise and, as the SNR increases, it begins to contain more desired speech. This may lead to target signal cancellation, especially if there is a steering vector mismatch [19], [28]. MBSTOI scores show that, while the adaptive method outperforms the compact model for noise reduction in some specific cases, the opposite is generally true for predicted intelligibility. These results are indeed consistent with the study of BTE arrays in [8], and suggest that the adaptive method introduces target signal degradation.

Results also show that the chest array presents a clear advantage over other geometries for the noise reduction task. This is likely due to the superior spatial diversity and wider aperture of the array with respect to the other geometries, and extends findings from [11], considering arrays of M=18 microphones, to arrays of M=4 microphones. This result further motivates the need to consider several geometries when designing beamformers: using a chest array can lead to an additional 1.6 dB mean attenuation compared to BTE HAs.

The relatively small variability of MBSTOI scores across array geometries when using compact-model SCM estimation shows an interesting application for special needs users. Indeed, a wearable array could be substituted for another, e.g. to

accommodate for the listener's disability, without substantial loss of intelligibility at the beamformer output.

Finally, the MBSTOI scores highlight one shortcoming of classical binaural beamforming: at positive SNR (or SIR), the best MBSTOI scores are obtained by a passthrough system at the listener's ears. Unless otherwise designed, the beamformer output changes spatial cues such that the target signal and residual noise appear as coming from the same direction, thus suppressing noise-related spatial cues. As a result, spatial release from masking cannot occur and the predicted intelligibility score is limited. This effect is less noticeable at low SNR as there is a large potential for noise reduction by the beamformers. This phenomenon could be addressed by designing beamformers recreating true spatialisation of all the recorded sources and interferences [20], [29], [30].

Experiment 2: Hybrid arrays. Results show that the compact model SCM estimation method often finds more benefits in the merger of BTE HAs with wearable arrays than the adaptive method does, both in terms of noise reduction and predicted intelligibility. Additionally, while the adaptive method leads to fwSNRseg improvements when using hybrid arrays, it is also shown to lead to decreases in MBSTOI scores. This, together with informal listening tests provided in [26], further confirm that the adaptive method is likely to create target signal degradation.

V. Conclusion

The use of a model-based [8] SCM estimation method for MPDR beamforming has been explored for wearable arrays. When compared to a classical adaptive SCM estimation method, the model-based method resulted in higher intelligibility scores, with little variability across geometries. An array placed on a user's chest exhibited the highest noise reducing capabilities, with a mean 1.3 dB improvement over BTE hearing aids. Combining BTE hearing aids with various wearable arrays showed that the compact model estimation method can yield mean MBSTOI improvements up to 6.7 % higher than when using the BTE hearing aids alone.

References

[1] B. D. van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[2] J. G. Desloge, W. M. Rabinowitz, and P. M. Zurek, "Microphone-array hearing aids with binaural output .I. Fixed-processing systems," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 6, pp. 529–542, 1997.

[3] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[4] H. L. van Trees, *Detection, Estimation, and Modulation Theory*. New York, USA: John Wiley & Sons, Inc., Apr. 2002, vol. IV, Optimum Array Processing.

[5] A. H. Moore, J. M. de Haan, M. S. Pedersen, P. A. Naylor, M. Brookes, and J. Jensen, "Personalized signal-independent beamforming for binaural hearing aids," *J. Acoust. Soc. Am.*, vol. 145, no. 5, pp. 2971–2981, 2019.

[6] I. McCowan and H. Bourlard, "Microphone array post-filter based on noise field coherence," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 709–716, Nov. 2003.

[7] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. Thompson, "Measurement of Correlation Coefficients in Reverberant Sound Fields," *J. Acoust. Soc. Am.*, vol. 27, no. 6, pp. 1072–1077, 1955.

[8] A. Moore, P. Naylor, and M. Brookes, "Improving robustness of adaptive beamforming for hearing devices," in *Proc. Int. Symp. on Auditory and Audiological Research. (ISAAR)*, vol. 7, Nyborg, Denmark, Jul. 2019, pp. 305–316.

[9] D. Van Krevelen and R. Poelman, "A survey of augmented reality technologies, applications and limitations," *Int. J. of Virtual Reality*, vol. 9, no. 2, pp. 1–20, 2010.

[10] V. Valimaki, A. Franck, J. Ramo, H. Gamper, and L. Savioja, "Assisted listening using a headset: Enhancing audio perception in real, augmented, and virtual environments," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 92–99, 2015.

[11] R. M. Corey, N. Tsuda, and A. C. Singer, "Acoustic impulse responses for wearable audio devices," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2019, pp. 216–220.

[12] A. McCormack and H. Fortnum, "Why do people fitted with hearing aids not wear them?" *Int. J. of Audiology*, vol. 52, no. 5, pp. 360–368, 2013.

[13] C. J. D. Hardy, C. R. Marshall, H. L. Golden, C. N. Clark, C. J. Mummery, T. D. Griffiths, D.-E. Bamiou, and J. D. Warren, "Hearing and dementia," *J. of Neurology*, vol. 263, no. 11, pp. 2339–2354, 2016.

[14] A. Palla, L. Fanucci, R. Sannino, and M. Settin, "Wearable speech enhancement system based on MEMS microphone array for disabled people," in *Proc. IEEE Int. Conf. on Des. & Tech. of Integr. Syst. in Nanoscale Era (DTIS)*, 2015, pp. 1–5.

[15] D. Y. Levin, E. A. Habets, and S. Gannot, "Near-field signal acquisition for smartglasses using two acoustic vector-sensors," *Speech Commun.*, vol. 83, pp. 42–53, 2016.

[16] Y. Avargel and I. Cohen, "On multiplicative transfer function approximation in the short-time Fourier transform domain," *IEEE Signal Process. Lett.*, vol. 14, no. 5, pp. 337–340, 2007.

[17] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[18] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1830–1847, Jul. 2004.

[19] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Trans. Signal Process.*, vol. 35, no. 10, pp. 1365–1376, Oct. 1987.

[20] B. Cornelis, S. Doclo, T. V. dan Bogaert, M. Moonen, and J. Wouters, "Theoretical analysis of binaural multimicrophone noise reduction techniques," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 2, pp. 342–355, Feb. 2010.

[21] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 397–401, 1988.

[22] "Objective measurement of active speech level," Int. Telecommun. Union (ITU-T), Geneva, Switzerland, Recommendation P.56, Mar. 1993.

[23] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, and V. Zue, "TIMIT acoustic-phonetic continuous speech corpus," Linguistic Data Consortium (LDC), Philadelphia, USA, Corpus, 1993.

[24] D. M. Brookes, "VOICEBOX: A speech processing toolbox for MATLAB," 1997. [Online]. Available: http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

[25] A. H. Andersen, J. M. d. Haan, Z. H. Tan, and J. Jensen, "Refinement and validation of the binaural short time objective intelligibility measure for spatially diverse conditions," *Speech Commun.*, vol. 102, pp. 1–13, Sep. 2018.

[26] E. d'Olne. (2021) Model-based beamforming for wearable microphone arrays. [Online]. Available: https://ed1016.github.io/CMMB-wearable/

[27] 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, Language Process.*, vol. 19, no. 7, pp. 2125–2136, Sep. 2011.

[28] L. Ehrenberg, S. Gannot, A. Leshem, and E. Zehavi, "Sensitivity analysis of MVDR and MPDR beamformers," in *IEEE Conv. Electrical and Electronics Engineers*, Eilat, Israel, Nov. 2010, pp. 416–420.

[29] E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, "Binaural multichannel Wiener filter with directional interference rejection," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Apr. 2015.

[30] N. Gößling, D. Marquardt, and S. Doclo, "Perceptual evaluation of binaural MVDR-based algorithms to preserve the interaural coherence of diffuse noise fields," *Trends Hear.*, vol. 24, 2020.