Speech Enhancement in Distributed Microphone Arrays Using Polynomial Eigenvalue Decomposition

Emilie d'Olne, Vincent W. Neo, Patrick A. Naylor
Department of Electrical and Electronic Engineering, Imperial College London, UK
{emilie.dolne16, vincent.neo09, p.naylor}@imperial.ac.uk

***Abstract*—As the number of connected devices equipped with multiple microphones increases, scientific interest in distributed microphone array processing grows. Current beamforming methods heavily rely on estimating quantities related to array geometry, which is extremely challenging in real, non-stationary environments. Recent work on polynomial eigenvalue decomposition (PEVD) has shown promising results for speech enhancement in singular arrays without requiring the estimation of any array-related parameter [1]. This work extends these results to the realm of distributed microphone arrays, and further presents a novel framework for speech enhancement in distributed microphone arrays using PEVD. The proposed approach is shown to almost always outperform optimum beamformers located at arrays closest to the desired speaker. Moreover, the proposed approach exhibits very strong robustness to steering vector errors.**

***Index Terms*—distributed microphone arrays, speech enhancement, polynomial matrix eigenvalue decomposition.**

I. Introduction

Recent years have seen an increase in the number of devices equipped with multiple microphones, from smartphones to wearable devices to home assistants. Consequently, the topic of distributed microphone array processing has grown in popularity with applications ranging from teleconferencing, room geometry estimation or enhanced hearing aids [2]–[5]. For speech enhancement, distributed arrays present several advantages over conventional compact arrays. As individual arrays, also referred to as nodes, are typically positioned arbitrarily, distributed arrays offer a large spatial diversity and better representation of the acoustic environment than single arrays [2]. In some specific use-cases, such as for example involving infirmity and geriatrics, distributed arrays may be less intrusive or uncomfortable for users compared to close-talking microphones or hearing aids [6], [7]. However, distributed array processing also introduces significant technical challenges. The geometry of the array may be unknown and non-stationary, both between nodes and within nodes in the case of wearable devices. Additionally, different devices operate with different sampling rates, or experience clock offset and skew [8]. Microphones in the network cannot be assumed to be calibrated, such that they may experience different frequency-dependent gains. Finally, there is often a limitation on the available transmission bandwidth between devices. These challenges are especially significant for speech enhancement since processing techniques such as beamforming heavily rely on having calibrated arrays with known geometries [9]. Standard array processing cannot generally be employed directly with distributed arrays, and alternative approaches must be found.

Typical distributed array signal processing methods may aim to apply classical beamforming methods by first performing gain equalisation and clock synchronisation between devices, and then estimating the required parameters such as relative transfer functions (RTF) [10] or noise covariance matrices [11]–[13]. The parameter estimation can occur at each node in a distributed manner or at a fusion centre [12]–[15]. Recent results show that, for low-rank multi-channel Wiener filters (MWF), the parameters can be estimated directly from uncalibrated microphone signals without loss of performance [8]. The performance of speech enhancement methods is limited whenever parameter estimation is inaccurate, which itself depends on accurate selection of internal regularisation parameters [13] and the performance of array synchronisation algorithms. Moreover, classical beamforming techniques are typically implemented in the frequency domain, processing separate frequency bands independently with narrowband techniques, without considering inter-band correlations [9].

Instead, multi-channel broadband processing methods based on the polynomial eigenvalue decomposition (PEVD) have been proposed [16]–[18]. Recently, PEVD-based speech enhancement methods were found to be effective for noise reduction and dereverberation in single spherical arrays or of arbitrary shapes [1], [19]. The approach does not rely on array geometry or noise estimation, and instead performs enhancement by strongly decorrelating the microphone signals in space, time, and frequency using PEVD. Moreover, compared to some classical speech enhancement methods such as spectral subtraction, mask-based enhancement and the single-channel Wiener filter, PEVD-based enhancement does not introduce any distortion in the desired speech signal [1].

This paper extends the the work on PEVD-based speech enhancement from single arrays [1] to distributed microphone arrays and proposes a novel processing framework for non-fully-connected networks. Sec. II introduces technical background on classical beamforming and PEVD for speech enhancement. Sec. III highlights the shortcomings of existing methods and presents the approach proposed in this paper. Secs. IV and V describe the design and results of simulations, and Sec. VI draws conclusions on this work.

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

II. Technical Background

A. Signal model

Given Q microphone arrays containing M_q microphones, the noisy speech signal recorded at the m^{th} microphone of the q^{th} array can be written in the time domain as

$$x_{q,m}(n) = \mathbf{h}_{q,m}^T \mathbf{s}_0(n) + v_{q,m}(n)\,, \tag{1}$$

where $n = 0, \ldots, N$ is the time index, $\mathbf{h}_{q,m}$ is the acoustic impulse response (AIR) between the desired source and the m^{th} microphone of the q^{th} array, assumed stationary and modelled as an FIR filter of order J, $\mathbf{s}_0(n) = [s_0(n), \ldots, s_0(n-J)]^T$ is the anechoic speech signal, $v_{q,m}(n)$ is additive noise, and $[\cdot]^T$ is the transpose operator. The noise signals are assumed to be zero-mean, non-perfectly coherent with each other, and uncorrelated with the source signal [20]. Stacking the microphone signals at the q^{th} array gives

$$\mathbf{x}_q(n) = \mathbf{H}_q^T \mathbf{s}_0(n) + \mathbf{v}_q(n)\,, \tag{2}$$

where $\mathbf{x}_q(n) = [x_{q,1}(n), \ldots, x_{q,M_q}(n)]^T$ with $\mathbf{v}_q$ defined similarly, and $\mathbf{H}_q = [\mathbf{h}_{q,1}, \ldots, \mathbf{h}_{q,M_q}]$. The noisy speech and noise spatial covariance matrices are therefore written as [9]

$$\mathbf{R}_{\mathbf{x}_q\mathbf{x}_q} = \mathbb{E}[\mathbf{x}_q(n)\mathbf{x}_q^T(n)] \qquad \mathbf{R}_{\mathbf{v}_q\mathbf{v}_q} = \mathbb{E}[\mathbf{v}_q(n)\mathbf{v}_q^T(n)]\,. \tag{3}$$

B. Beamforming

Beamforming is often performed in the STFT domain, such that for weights $\check{\mathbf{w}}_q(k,\ell) = [\check{w}_{q,1}(k,\ell), \ldots, \check{w}_{q,M_q}(k,\ell)]^T$ the beamformer output at the q^{th} array is given by

$$\check{y}_q(k,\ell) = \check{\mathbf{w}}_q^H(k,\ell)\check{\mathbf{x}}_q(k,\ell)\,, \tag{4}$$

where $a(n) \multimap \check{a}(k,l)$ represents a STFT pair with k and ℓ the frequency and time-frame indices, and $[\cdot]^H$ the Hermitian transpose. The well known delay-and-sum beamformer (DSB) is then defined as [21]

$$\check{\mathbf{w}}_q^{\mathrm{DSB}}(k,\ell) = \frac{\check{\mathbf{d}}_q(k,\ell)}{\check{\mathbf{d}}_q^H(k,\ell)\check{\mathbf{d}}_q(k,\ell)}\,, \tag{5}$$

where $\check{\mathbf{d}}_q(k,\ell) = [\check{d}_{q,1}(k,\ell), \ldots, \check{d}_{q,M_q}(k,\ell)]^T$ is the steering vector associated with the q^{th} array [9]. Setting $\check{\mathbf{d}}_q(k,\ell) = \check{\mathbf{h}}_q(k)$, where $\check{\mathbf{h}}_q(k)$ is the k^{th} row of the DFT of $\mathbf{H}_q$, results in a filter-and-sum beamformer (FSB). Another widely used beamformer is the minimum variance distortionless response (MVDR) beamformer defined in [22] as

$$\check{\mathbf{w}}_q^{\mathrm{MVDR}}(k,\ell) = \frac{\mathbf{\Phi}_{\check{\mathbf{v}}_q\check{\mathbf{v}}_q}^{-1}(k,\ell)\check{\mathbf{d}}_q(k,\ell)}{\check{\mathbf{d}}_q^H(k,\ell)\mathbf{\Phi}_{\check{\mathbf{v}}_q\check{\mathbf{v}}_q}^{-1}(k,\ell)\check{\mathbf{d}}_q(k,\ell)}\,, \tag{6}$$

where $\mathbf{\Phi}_{\check{\mathbf{v}}_q\check{\mathbf{v}}_q}(k,\ell) = \mathbb{E}[\check{\mathbf{v}}_q(k,\ell)\check{\mathbf{v}}_q^H(k,\ell)]$ corresponds to (3) in the STFT domain. By exploiting the noise covariance matrix, the MVDR beamformer is able to adapt to changes in the acoustic environment and is therefore referred to as an adaptive beamformer. The adaptive estimation of the noise covariance matrix is non-trivial and poor estimates lead to degraded beamforming performance [9].

C. PEVD-based speech enhancement

By processing the signal in frequency bands independently as in (4), classical beamforming methods ignore spectro-temporal correlation across bands [1]. To overcome this issue, the PEVD speech enhancement method in [1] exploits the space-time correlation matrix defined as [16]

$$\mathbf{R}_{\mathbf{xx}}(\tau) = \mathbb{E}[\mathbf{x}(n)\mathbf{x}^H(n-\tau)], \tag{7}$$

where $\mathbf{x}(n) = [\mathbf{x}_1^T(n), \ldots, \mathbf{x}_Q^T(n)]^T$ is the concatenation of microphone signals from all Q arrays and τ is a temporal lag. Note that the spatial correlation matrix in (3) is a special case of (7) for $\tau = 0$. Concatenating the correlation matrix, $\mathbf{R}_{\mathbf{xx}}(\tau)$, for all values of $\tau \in \{-N, \ldots, N\}$, results in a 3-dimensional tensor. Instead of processing signals in the STFT domain as in (4), the z-transform is used which captures and preserves spatial, temporal, and spectral correlations of the received signals. The z-transform of (7) is [16]

$$\mathcal{R}_{\mathbf{xx}}(z) = \sum_{\tau=-\infty}^{\infty} \mathbf{R}_{\mathbf{xx}}(\tau)\, z^{-\tau}. \tag{8}$$

The so-obtained polynomial matrix is a matrix with polynomial elements, or equivalently, a polynomial with matrix coefficients. The PEVD of (8) is [16]

$$\mathcal{R}_{\mathbf{xx}}(z) \approx \boldsymbol{\mathcal{U}}(z)\mathbf{\Lambda}(z)\boldsymbol{\mathcal{U}}^P(z), \tag{9}$$

where $\boldsymbol{\mathcal{U}}(z)$ is the eigenvector polynomial matrix and the diagonal polynomial matrix, $\mathbf{\Lambda}(z)$ contains the eigenvalues, and $[\cdot]^P$ is the para-Hermitian operator such that $\boldsymbol{\mathcal{U}}^P(z) = \boldsymbol{\mathcal{U}}^H(1/z^*)$. The approximation in (9) is due to the use iterative algorithms [16], [17], [23], [24]. The enhanced speech signal can be obtained as [1]

$$y^{\mathrm{PEVD}}(z) = \boldsymbol{u}_1^P(z)\,\mathbf{x}(z) \tag{10}$$

where $\boldsymbol{u}_1(z)$ is the eigenvector associated with the first eigenvalue. More details on PEVD are given in [1].

III. Proposed Method

A principal shortcoming of classical beamforming methods for speech enhancement is their inability to account fully for spectro-temporal correlations in microphone signals [9]. Speech signals are known to exhibit periodic patterns in time and frequency (e.g. during periods of voiced speech) and it is therefore reasonable to assume that algorithms capable of capturing these patterns would demonstrate better enhancement capabilities. One such method is the PEVD-based speech enhancement method in [1], [19] which at its core exploits the space-time correlation matrix in (7) to estimate the desired signal subspace. However, PEVD methods are computationally heavy, scaling as $O(M^3)$ where M is the number of microphone signals [25]. In this work, we propose a framework for PEVD-based speech enhancement in distributed microphone arrays, illustrated in Fig. 1. Rather than processing all microphone signals in all arrays as in (10), every array is first processed by a beamformer defined in (4), thereby producing Q beamformed signals. These are then processed

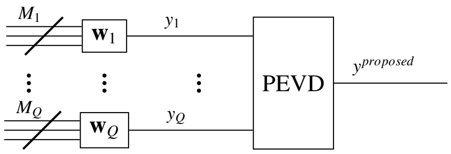

Figure 1: Proposed method block diagram.

by the PEVD-enhancement algorithm to produce a single-channel output. The computational complexity of the proposed approach is therefore limited to $\mathcal{O}(Q^3)$, where Q is smaller than the total number of microphones. By beamforming in local nodes rather than across arrays, there is no need for the proposed method to synchronise microphones located in different devices, and to estimate the acoustic transfer function between devices.

A. Formulation

For the q^{th} array, the M_q microphone signals are beamformed using (4). For example, using the DSB in (5) with $\check{\mathbf{d}}_q(k,\ell) = \check{\mathbf{h}}_q(k)$, produces $y_q^{DSB}(n)$ for $q = 1, \ldots, Q$. In this method, the fixed DSB is preferred over the MVDR beamformer as it does not rely on the estimation of the instantaneous noise correlation matrix. The DSB output can be stacked in vector form such that $\mathbf{y}^{DSB}(n) = [y_1^{DSB}(n), \ldots, y_Q^{DSB}(n)]^T$. The space-time covariance matrix $\mathbf{R}_{\mathbf{y}^{DSB}\mathbf{y}^{DSB}}$ is computed using (7). The enhanced signal is then obtained through the PEVD decomposition of $\mathbf{R}_{\mathbf{y}^{DSB}\mathbf{y}^{DSB}}(z)$ following (9) and (10), and yielding the single-channel signal $y^{DSB+PEVD}(n)$.

IV. Experimental Setup

A. Setup

Anechoic speech signals are taken from IEEE sentences [26] recorded by a male native British English speaker, sampled at 16 kHz. The considered scenario is depicted in Fig. 2; three arrays of two microphones each are arranged in a circle of radius $r = 0.5$ m in the centre of a $5 \times 5 \times 3$ m room, as if placed around a table. The simulated target speaker is 1.25 m away from the closest microphone array. Speech signals are convolved with simulated room impulse responses (RIR) obtained using the generator in [27] to obtain the microphone array signals. White sensor noise is added at a 30 dB signal-to-noise ratio (SNR). Spherical isotropic noise is generated at the microphones using [28], and added to the speech components using [29] in [30] such that the noise level is kept constant within the room. This leads to different SNRs at different microphones, and the input SNR is defined in the remainder of this article at a reference microphone (in red in Fig. 2). Babble noise is taken from the NATO RSG-10 noise database [31], while speech-shaped noise [32] is generated using [30].

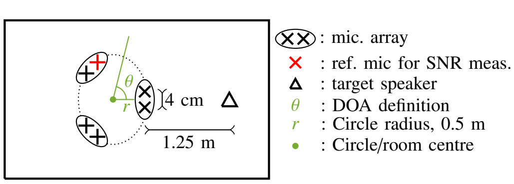

Figure 2: Simulated room configuration (not to scale).

B. Performance measures

To measure the denoising performance, the frequency-weighted segmental SNR (fwSegSNR) [30], [33] is computed. The STOI [34] measure is used to predict the speech intelligibility.

C. Algorithm parameter selection

The speech enhancement performance of the beamformers in (5) and (6), the PEVD in (10) applied to all 6 raw microphone signals, and the proposed method are compared. The steering vectors in (5) and (6) are selected using the known reverberant RIRs between the target and individual arrays. The noise covariance matrix in (6) is computed based on the oracle noise signal, and diagonal loading is applied to limit its condition number to ≤ 100 [35]. The STFT is implemented with Hamming windows of 16 ms overlapping by 50%. The PEVD parameters were chosen following [1]. With this parameter selection, correlations within 100 ms, which were assumed to include the direct-path and early reflection components, were captured and used by the algorithm.

V. Results and Discussion

A. Experiment 1: Baseline performance

The speech enhancement performance of the DSB, MVDR, PEVD and proposed methods are evaluated for the configuration in Fig. 2, with results averaged over 50 speech signals. Other array positions were considered which yielded similar results, and are therefore not presented in this work. The STOI and fwSegSNR results for corruptive speech-shaped noise in an anechoic environment and a reverberant room ($T_{60} = 400$ ms) are plotted in Fig. 3. Since the DSB and MVDR methods yield 3 outputs (1 per array), the minimum and maximum performance values are plotted. The raw microphone signals are similarly plotted for comparison purposes.

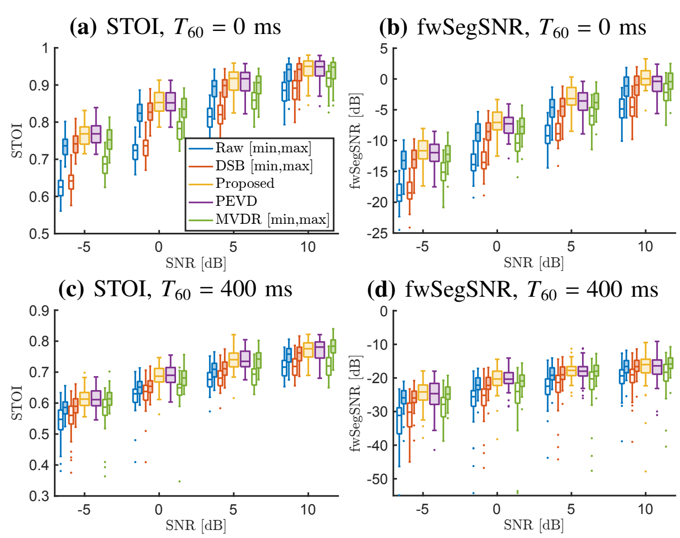

Figure 3: STOI and fwSegSNR improvements for diffuse speech-shaped noise.

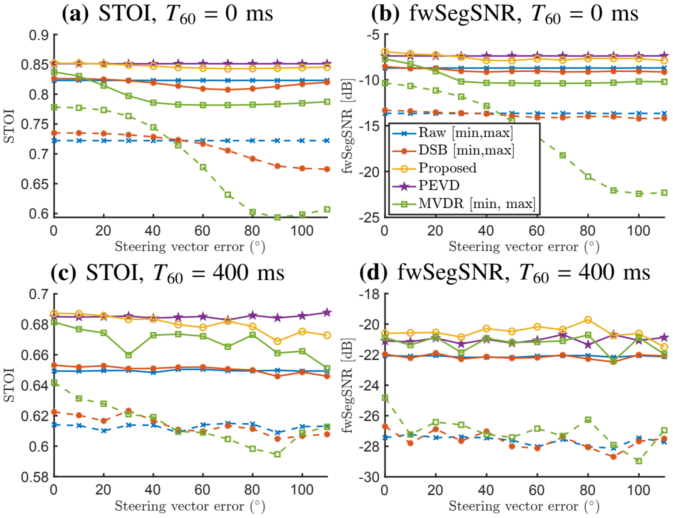

Figure 4: Steering error robustness for diffuse speech-shaped noise at 0 dB SNR, averaged over 50 sentences.

For the anechoic scenario in Figs. 3(a) and 3(b), the PEVD and proposed methods yield very similar STOI and fwSegSNR scores and slightly outperform the best performing DSB and MVDR beamformers. The proposed method gives an increase in STOI of up to 0.15 over the minimum scoring raw microphone signal, and of 7.1 dB in fwSegSNR. The improvements over the maximum scoring microphone and beamformers are less significant (< 0.05 in STOI and < 2 dB in fwSegSNR). A potential explanation for this marginal improvement is that the maximum scoring microphone and beamformers are located at the array closest to the desired source, therefore already benefiting from a higher SNR. This effect is assumed to be a reasonable modelling of real environments where individual nodes, e.g. phones, are likely to be situated near talkers.

Similar results are observed for the reverberant scenario in Figs. 3(c) and 3(d), with an improvement in STOI of 0.08 and in fwSegSNR of 8.2 dB of the proposed method over the minimum scoring microphone.

Additional scenarios considering higher levels of reverberation (T_{60} = 700 and 1000 ms) and babble and white noise were examined which yielded similar results and are therefore not included here for conciseness. The full results alongside listening examples can be found in [36].

These results show two important trends. Firstly, the performance of the PEVD enhancement applied to all 6 raw microphones and that of the proposed method are comparable across all SNRs and levels of reverberation. This indicates that the reduced computational complexity of the proposed method over the classical PEVD does not necessarily lead to a decrease in denoising performance. Secondly, the PEVD and proposed method always perform as well, or better than a beamformer applied to the array closest to the desired talker. This result is significant as PEVD-based methods achieve this near-optimal performance without prior knowledge of the source location or of the array geometry.

B. Experiment 2: Robustness to steering errors

In this experiment, the robustness of the proposed method to steering vector errors is examined. When considering distributed arrays, there is ambiguity in the definition of direction-of-arrival and thus in the quantification of steering vector errors [2], [37]. In this experiment, the steering angle θ is defined with respect to the centre of the distributed array network and the line connecting it to the right-most array, so that the target is considered to be at 0°and angles increase in an anti-clockwise manner (see Fig. 2). This is such that when errors are introduced, all arrays are steered in the same, erroneous direction. Results for diffuse speech-shaped noise at 0 dB SNR in anechoic and reverberant rooms (T_{60} = 400 ms) are plotted in Fig. 4. Similarly to Sec. V-A, minimum and maximum scoring beamformers are plotted when applicable.

As is expected, the PEVD algorithm is unaffected by steering vector errors since its implementation does not rely on steering vectors. For the anechoic scenario in Figs. 4(a) and 4(b), the proposed method shows very little variations across all considered angular errors, with a maximum deviation from the ideal steering of 0.01 in STOI and 1.0 dB in fwSegSNR. Conversely, the MVDR beamformer is strongly affected by mis-steering, with maximum variations of 0.18 in STOI and 12.1 dB in fwSegSNR over the ideal steering. Similar observations can be made for the reverberant case in Figs. 4(c) and 4(d), with maximum deviations for the proposed method of 0.02 in STOI and 1.8 dB in fwSegSNR. Additionally, the proposed method slightly outperforms the PEVD method in fwSegSNR score by an average of 0.5 dB.

The proposed method shows very high robustness to steering vector errors in comparison to classical beamforming methods. It also exhibits a very similar performance to the PEVD algorithm applied to the 6 raw microphone signals, notwithstanding the fact that it is based on 3 wrongly steered beamformers. This suggests that the PEVD processing in the proposed method is indeed capable of extracting spectro-temporal correlation in signals even when the spatial cues are sub-optimal. This result is highly significant for arrays with unknown or varying geometries, where steering vectors can only be estimated to a certain degree of accuracy.

VI. Conclusion

This work extended the use of PEVD from singular microphone arrays to distributed microphone networks. It introduced a novel framework for speech enhancement using a combination of local beamformers and PEVD enhancement. The proposed method was shown to improve predicted intelligibility by up to 0.15 in STOI and to increase the frequency-weighted segmental SNR by up to 7.1 dB compared to raw microphone signals. The proposed method also showed strong robustness to steering vector error, with a maximum deviation in STOI scores of 0.02 compared to the ideal steering case. These results were observed in anechoic and reverberant conditions. Overall, the results show the potential of PEVD-based enhancement methods in distributed arrays, where the estimation of array geometry and steering vectors is challenging.

References

[1] V. W. Neo, C. Evers, and P. A. Naylor, "Enhancement of noisy reverberant speech using polynomial matrix eigenvalue decomposition," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 3255–3266, Oct. 2021.

[2] S. Pasha, J. Lundgren, C. Ritz, and Y. Zou, "Distributed microphone arrays, emerging speech and audio signal processing platforms: A review," *Advances in Science, Technol. and Eng. Systems*, vol. 5, no. 4, pp. 331–343, 2020.

[3] J. Amini, R. C. Hendriks, R. Heusdens, M. Guo, and J. Jensen, "Spatially correct rate-constrained noise reduction for binaural hearing aids in wireless acoustic sensor networks," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 28, pp. 2731–2742, 2020.

[4] S. Araki, N. Ono, K. Kinoshita, and M. Delcroix, "Comparison of reference microphone selection algorithms for distributed microphone array based speech enhancement in meeting recognition scenarios," in *Proc. Int. Workshop on Acoust. Signal Enhancement (IWAENC)*, 2018, pp. 316–320.

[5] I. Dokmanić, L. Daudet, and M. Vetterli, "From acoustic room reconstruction to slam," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2016, pp. 6345–6349.

[6] 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.

[7] 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.

[8] S.-E. Kotti, R. Heusdens, and R. C. Hendriks, "Clock-offset and microphone gain mismatch invariant beamforming," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*. Amsterdam, Netherlands: IEEE, Jan. 2021, pp. 176–180.

[9] M. S. Brandstein and D. B. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin, Germany: Springer-Verlag, 2001.

[10] 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.

[11] A. Hassani, A. Bertrand, and M. Moonen, "LCMV beamforming with subspace projection for multi-speaker speech enhancement," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2016, pp. 91–95.

[12] A. Bertrand and M. Moonen, "Distributed adaptive generalized eigenvector estimation of a sensor signal covariance matrix pair in a fully connected sensor network," *Signal Process.*, vol. 106, pp. 209–214, 2015.

[13] V. M. Tavakoli, J. R. Jensen, R. Heusdens, J. Benesty, and M. G. Christensen, "Distributed max-SINR speech enhancement with ad hoc microphone arrays," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2017, pp. 151–155.

[14] A. Bertrand and M. Moonen, "Distributed LCMV beamforming in a wireless sensor network with single-channel per-node signal transmission," *IEEE Trans. Signal Process.*, vol. 61, no. 13, pp. 3447–3459, 2013.

[15] E. d'Olne, A. H. Moore, and P. A. Naylor, "Model-based beamforming for wearable microphone arrays," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2021.

[16] J. G. McWhirter, P. D. Baxter, T. Cooper, S. Redif, and J. Foster, "An EVD algorithm for para-hermitian polynomial matrices," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2158–2169, May 2007.

[17] S. Redif, S. Weiss, and J. G. McWhirter, "Sequential matrix diagonalisation algorithms for polynomial EVD of parahermitian matrices," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 81–89, Jan. 2015.

[18] S. Weiss, S. Bendoukha, A. Alzin, F. K. Coutts, I. K. Proudler, and J. Chambers, "MVDR broadband beamforming using polynomial matrix techniques," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2015, pp. 839–843.

[19] V. W. Neo, C. Evers, and P. A. Naylor, "Polynomial matrix eigenvalue decomposition of spherical harmonics for speech enhancement," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Jun. 2021.

[20] Y. Huang, J. Benesty, and J. Chen, "Analysis and comparison of multichannel noise reduction methods in a common framework," *IEEE Trans. Audio, Speech, Language Process.*, vol. 16, no. 5, pp. 957–968, Jul. 2008.

[21] H. L. Van Trees, *Optimal Array Processing. Part IV of Detection, Estimation, and Modulation Theory*. John Wiley & Sons, 2002.

[22] 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.

[23] S. Redif, J. G. McWhirter, and S. Weiss, "Design of FIR paraunitary filter banks for subband coding using a polynomial eigenvalue decomposition," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5253–5264, Nov. 2011.

[24] V. W. Neo and P. A. Naylor, "Second order sequential best rotation algorithm with Householder transformation for polynomial matrix eigenvalue decomposition," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2019, pp. 8043–8047.

[25] F. K. Coutts, J. Corr, K. Thompson, S. Weiss, I. K. Proudler, and J. G. McWhirter, "Memory and complexity reduction in parahermitian matrix manipulations of PEVD algorithms," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2016, pp. 1633–1637.

[26] E. H. Rothauser, W. D. Chapman, N. Guttman, M. H. L. Hecker, K. S. Nordby, H. R. Silbiger, G. E. Urbanek, and M. Weinstock, "IEEE recommended practice for speech quality measurements," *IEEE Trans. Audio and Electroacoust.*, vol. 17, no. 3, pp. 225–246, 1969.

[27] E. Habets, "Room impulse response generator," May 2021. [Online]. Available: https://github.com/ehabets/RIR-Generator/blob/3cf914df697f2bc2d235708644a05bbff410df47/rir_generator.pdf

[28] E. A. P. Habets, I. Cohen, and S. Gannot, "Generating nonstationary multisensor signals under a spatial coherence constraint," *J. Acoust. Soc. Am.*, vol. 124, no. 5, pp. 2911–2917, Nov. 2008.

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

[30] 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

[31] H. J. M. Steeneken and F. W. M. Geurtsen, "Description of the RSG.10 noise data-base," TNO Institute for Perception, Tech. Rep., 1988.

[32] "Artificial voices," Int. Telecommun. Union (ITU-T), Standard, Sep. 1999.

[33] Y. Hu and P. C. Loizou, "Evaluation of objective measures for speech enhancement," in *Proc. Conf. of Int. Speech Commun. Assoc. (INTERSPEECH)*, 2006, pp. 1447–1450.

[34] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "An algorithm for intelligibility prediction of time-frequency weighted noisy speech," *IEEE J ASLP*, vol. 19, no. 7, pp. 2125–2136, Sep. 2011.

[35] 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.

[36] E. d'Olne, "Speech enhancement in distributed microphone arrays using polynomial eigenvalue decomposition," 2022. [Online]. Available: https://ed1016.github.io/se_PEVD_da/

[37] S. Araki, H. Sawada, R. Mukai, and S. Makino, "DOA estimation for multiple sparse sources with arbitrarily arranged multiple sensors," *J. of Signal Process. Systems*, vol. 63, no. 3, pp. 265–275, 2011.