FRAME-BASED SPACE-TIME COVARIANCE MATRIX ESTIMATION FOR POLYNOMIAL EIGENVALUE DECOMPOSITION-BASED SPEECH ENHANCEMENT

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

Recent work in speech enhancement has proposed a polynomial eigenvalue decomposition (PEVD) method, yielding significant intelligibility and noise-reduction improvements without introducing distortions in the enhanced signal [1]. The method relies on the estimation of a space-time covariance matrix, performed in batch mode such that a sufficiently long portion of the noisy signal is used to derive an accurate estimate. However, in applications where the scene is non-stationary, this approach is unable to adapt to changes in the acoustic scenario. This paper thus proposes a frame-based procedure for the estimation of space-time covariance matrices and investigates its impact on subsequent PEVD speech enhancement. The method is found to yield spatial filters and speech enhancement improvements comparable to the batch method in [1], showing potential for real-time processing.

***Index Terms*—** polynomial eigenvalue decomposition, speech enhancement, adaptive processing

1. INTRODUCTION

The topic of polynomial eigenvalue decomposition (PEVD) has recently gained traction in the signal processing literature. Applications were found in multichannel enhancement for arbitrarily-shaped arrays [1], spherical microphones [2], or distributed microphone networks [3]; in channel identification [4]; in DOA estimation with polynomial MUSIC [5–7]; in voice activity detection [8]; or in beamforming with a broadband MVDR beamformer [9]. These methods rely on the estimation of a space-time covariance matrix capturing signal correlations in space, time, and frequency, thereby allowing true broadband processing [1]. However, this matrix estimation is performed in batch mode (i.e. a sufficiently long signal duration is used to find an accurate estimate), thus assuming that the recorded acoustic environment is stationary. This assumption is unlikely in acoustic scenarios where sources may be moving, leaving or entering the scene. Moreover, using several snapshots of the signal of interest can resolve magnitude ambiguity for channel identification [4]. If these methods are to be used in real-life scenarios, it is essential to investigate their performance and limitations when moving from batch mode towards frame-based processing.

Two factors are likely to impact the performance of frame-based PEVD methods: the estimation of the space-time covariance matrix using a limited set of samples [10,11], and the algorithm employed to perform its diagonalisation [12–14]. Algorithmic limitations are conditioned by the nature of the matrix to factorise, and by internal order reduction mechanisms [15–17]. On the other hand, the unbiased estimation of sample space-time covariance matrices has closed-form mathematical expressions for the estimator variance, and it is possible to compute the optimal time-lag support over which to estimate the matrix given a limited number of samples and the ground truth support [10, 11]. The estimation error relates to eigenvalue and eigenvector perturbations when performing the matrix PEVD [18].

These limitations suggest that the performance of PEVD-based speech processing methods will be affected by the number of available samples in the frame-based space-time covariance matrix estimation. However, the extent to which performance is impacted has not yet been established, and it is unknown how classical covariance matrix estimation methods [19] would perform for the space-time covariance matrix estimation task. Therefore, this paper proposes an iterative frame-based estimation of the space-time covariance matrix, and aims to quantify its impact on the performance of the PEVD speech enhancement presented in [1, 2]. Performance is first assessed through space-time covariance matrix estimation accuracy. Then, the characteristics of the derived PEVD speech-enhancement filters are investigated. Finally, the impact on noise-reduction and intelligibility improvement are compared between the proposed method and the batch approach in [1].

2. PROBLEM FORMULATION

2.1. Signal model

Given M microphones, the noisy speech recorded at the m^{th} microphone is given by

$$x_m(n) = \mathbf{h}_m^T \mathbf{s}(n) + v_m(n)\,, \tag{1}$$

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

978-1-6654-6867-1/22/$31.00 ©2022 IEEE

where n is the time index, $\mathbf{h}_m$ is the acoustic impulse response (AIR) between the desired source and the m^{th} microphone, assumed stationary and modelled as an FIR filter of order J, $\mathbf{s}(n) = [s(n), \ldots, s(n-J)]^T$ is the anechoic speech signal, $v_m(n)$ 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 gives

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

where $\mathbf{x}(n) = [x_1(n), \ldots, x_M(n)]^T$ with $\mathbf{v}(n)$ defined similarly, and $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_M]$.

2.2. Sample space-time covariance matrix

The space-time covariance matrix is given by [12]

$$\mathbf{R}_{\mathbf{xx}}(\tau) = \mathbb{E}[\mathbf{x}(n)\mathbf{x}^H(n-\tau)]\,, \tag{3}$$

with $[\cdot]^H$ defined as the Hermitian transpose operator such that the $(p,q)^{th}$ element of $\mathbf{R}_{\mathbf{xx}}(\tau)$ is given by $r_{p,q}(\tau) = \mathbb{E}[x_p(n)x_q^*(n-\tau)]$, with $[\cdot]^*$ the complex conjugate operator. For a white Gaussian source and in the absence of noise, (3) reduces to [11, 21]

$$\mathbf{R}_{\mathbf{xx}}(\tau) = \sum_J \mathbf{H}^H(j)\mathbf{H}(j-\tau)\,. \tag{4}$$

In practice, when only N snapshots of $\mathbf{x}(n)$ are available such that $n = 0, \ldots, N-1$, (3) is estimated from samples and yields the noisy estimate $\widehat{\mathbf{R}}_{\mathbf{xx}}(\tau)$. Assuming stationarity, the $(p,q)^{th}$ element of $\widehat{\mathbf{R}}_{\mathbf{xx}}(\tau)$ is [11]

$$\widehat{r}_{p,q}(\tau) = \begin{cases} \frac{1}{N-\tau}\sum_{n=0}^{N-\tau-1} x_p(n+\tau)x_q^*(n), & \tau \geq 0 \\ \widehat{r}_{q,p}^*(-\tau), & \tau < 0\,. \end{cases} \tag{5}$$

The variance of this unbiased estimator was derived in [11] for Gaussian signals, and was found to vary with the number of samples, N, and the ground-truth space-time covariance matrix. Given the support, $2T+1$, of $\widehat{\mathbf{R}}_{\mathbf{xx}}(\tau)$, such that $\widehat{\mathbf{R}}_{\mathbf{xx}}(\tau) = 0$ for $|\tau| > T$, truncation errors may arise when $T \leq \tau_{max}$, with $2\tau_{max}+1$ the ground truth support of $\mathbf{R}_{\mathbf{xx}}(\tau)$. The estimation suffers from a tradeoff between truncation errors and errors due to estimator variance: as T increases, truncation errors reach zero, while the estimator variance generally grows [10, 11].

2.3. PEVD-based speech enhancement

The PEVD speech enhancement method in [1] uses the space-time correlation matrix in (3) to capture signal correlations in space, time, and frequency. Concatenating the correlation matrix, $\mathbf{R}_{\mathbf{xx}}(\tau)$, for all values of $\tau \in \{-N+1, \ldots, N-1\}$, results in a 3-dimensional tensor. The z-transform of (3) is given by [12]

$$\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z) = \sum_{\tau=-\infty}^{\infty} \mathbf{R}_{\mathbf{xx}}(\tau)\, z^{-\tau}. \tag{6}$$

The so-obtained polynomial matrix is a matrix with polynomial elements, or equivalently, a polynomial with matrix coefficients. The PEVD of (6) is [12]

$$\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z) \approx \boldsymbol{\mathcal{U}}(z)\boldsymbol{\Lambda}(z)\boldsymbol{\mathcal{U}}^P(z), \tag{7}$$

where $\boldsymbol{\mathcal{U}}(z)$ is the eigenvector polynomial matrix and the diagonal polynomial matrix, $\boldsymbol{\Lambda}(z)$ 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 (7) is due to the use of iterative algorithms to obtain the decomposition [12–14]. Assuming uncorrelated speech and noise signals, the principal eigenvector, $\boldsymbol{u}_1(z)$, is associated with the speech-only subspace, and speech enhancement occurs through [1]

$$y(z) = \boldsymbol{u}_1^P(z)\,\mathbf{x}(z)\,. \tag{8}$$

Details on PEVD speech enhancement are given in [1].

2.4. Batch mode versus frame-based enhancement

When processing the received signal in batch mode as in [1], the acoustic scene is assumed stationary over all $n \in \{0, \ldots, N-1\}$, and N is typically much larger than the ground truth support τ_{max}. The variance of the estimate $\widehat{\mathbf{R}}_{\mathbf{xx}}(\tau)$ can then be reduced using K signal segments of length $L \geq \tau_{max}$, producing the long-term average

$$\widetilde{\mathbf{R}}_{\mathbf{xx}}(\tau) = \frac{1}{K}\sum_{k=1}^{K} \widehat{\mathbf{R}}_{\mathbf{x}^k\mathbf{x}^k}(\tau)\,, \tag{9}$$

where $\widehat{\mathbf{R}}_{\mathbf{x}^k\mathbf{x}^k}(\tau)$ is the space-time covariance matrix obtained following (5) in the k^{th} frame of length L, such that $n \in \{(k-1)L, \ldots, kL-1\}$. An optimum support $T_{opt} \leq \tau_{max}$ can be found to minimise the estimation error [10].

In frame-based processing, however, the scene is only assumed stationary within a frame of length $L << \tau_{max}$. Approaches must therefore be explored that are able to produce a reliable estimate of the space-time covariance matrix when a limited set of signals samples is available in any one frame. Motivated by the approach for recursive spatial covariance matrix estimation [22, 23], this paper proposes the following iterative procedure for the polynomial case

$$\widehat{\mathbf{R}}_{\mathbf{xx}}^{k}(\tau) = \alpha\widehat{\mathbf{R}}_{\mathbf{xx}}^{k-1}(\tau) + (1-\alpha)\widehat{\mathbf{R}}_{\mathbf{x}^k\mathbf{x}^k}(\tau)\,, \tag{10}$$

where $\widehat{\mathbf{R}}_{\mathbf{xx}}^{k}(\tau)$ is the space-time covariance matrix estimate in the k^{th} frame of length L, and α is a recursive smoothing parameter. Thus, the recursive estimate in the k^{th} frame uses

the recursive estimate up to the $(k-1)^{th}$ frame, and the instantaneous estimate obtained at the k^{th} frame. Larger values of α incur longer memory in the system, but limit the amount by which the estimate can adapt to changing scenarios. Truncation errors are constrained by the frame length $L < \tau_{max}$, while errors due to estimator variance can be reduced with increased system memory. Therefore, during stationary periods, larger values of α should lead to lower steady-state errors. Setting $\alpha = \frac{1}{k}$ leads to the form in (9) as $k = K$. The PEVD decomposition in (7) is performed for every frame independently, leading to the filters $\widehat{\boldsymbol{u}}^k(z)$ and the enhanced signal frames $\widehat{y}^k(z)$.

3. EXPERIMENTS

3.1. Metrics

When employing the procedure in (10), speech enhancement performance is likely to be affected by the covariance matrix estimation accuracy and the resulting filters $\widehat{\boldsymbol{u}}_1^k(z)$. The element-wise normalised projection misalignment (NPM) between $\widehat{\mathbf{R}}_{\mathbf{xx}}(\tau)$ and $\mathbf{R}_{\mathbf{xx}}(\tau)$ is used to quantify the space-time covariance matrix estimation error, such that [24]

$$\text{NPM}_{(p,q)} = 10\log\Big(\frac{|\mathbf{r}_{(p,q)} - \beta\widehat{\mathbf{r}}_{(p,q)}|^2}{|\mathbf{r}_{(p,q)}|^2}\Big),\ \beta = \frac{\mathbf{r}_{(p,q)}^H\widehat{\mathbf{r}}_{(p,q)}}{|\widehat{\mathbf{r}}_{(p,q)}|^2},$$

where $\mathbf{r}_{(p,q)} = [r_{(p,q)}(-\tau_{max}), \ldots, r_{(p,q)}(\tau_{max})]^T$ and $\widehat{\mathbf{r}}_{(p,q)}$ defined similarly, and $|\cdot|$ is the Euclidean norm. The aggregated NPM is then obtained by summing over all matrix elements, and it provides a gain invariant distance between the two space-time covariance matrices.

Filter characteristics are described by the beampattern $\text{B}(\varphi, e^{j\Omega})$ and the directivity index (DI) defined herein as [25]

$$\text{B}(\varphi, e^{j\Omega}) = \boldsymbol{u}_1{}^H(e^{j\Omega})\boldsymbol{d}(\varphi, e^{j\Omega}) \tag{11}$$

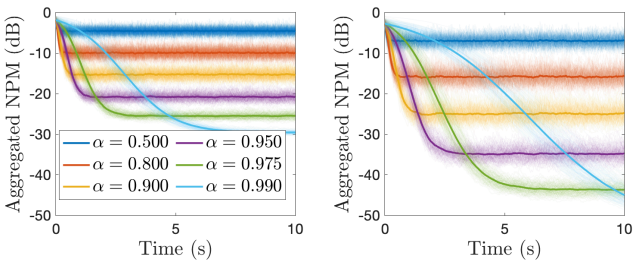
$$\text{DI} = \frac{1}{4\pi}\int_{\Omega=0}^{2\pi}\Big(\frac{B(\varphi_0, e^{j\Omega})}{\int_{\varphi=0}^{2\pi}B(\varphi, e^{j\Omega})d\varphi}\Big)d\Omega\,, \tag{12}$$

with Ω the discrete angular frequency, $\boldsymbol{d}(\varphi, e^{j\Omega})$ the direct-path plane-wave array manifold for a source originating at an azimuth φ, and φ_0 is the target source azimuth direction.

Finally, the frequency-weighted segmental SNR [26, 27] is computed to measure denoising performance, while STOI [28] is used to predict the speech intelligibility.

3.2. Setup

Simulations consider a linear array of $M = 3$ microphones spaced by 5 cm, receiving a source placed 1 m away along the array's axis and corresponding to an azimuth direction of 90°. The sampling rate is 16 kHz. Room impulse responses (RIRs) are simulated using the generator in [29] for a $4 \times 4 \times 3$ room with reverberation time $T_{60} = 400$ ms. For the recursive analysis, non-overlapping frames of the signal are taken using a rectangular window of length L.



(a) $L = 20$ ms **(b)** $L = 40$ ms

Fig. 1: Aggregated NPM between $\widehat{\mathbf{R}}_{\mathbf{xx}}^k(\tau)$ and $\mathbf{R}_{\mathbf{xx}}(\tau)$ as a function of time, for various α and $L = \{20, 40\}$ ms.

3.3. Experiment 1: Estimation accuracy

This experiment uses a 10 s segment of white Gaussian noise as the recorded source, such that the ground truth $\mathbf{R}_{\mathbf{xx}}(\tau)$ can be computed according to (4). The performance of the space-time covariance matrix estimation in (10) is measured using the aggregated NPM as a function of time, for various frame lengths L and smoothing factors α. Results are plotted in Fig. 1 for 200 realisations of the source and for $L = \{20, 40\}$ ms.

Results show that lower values of α lead to faster convergence of the estimate, but also yield a higher steady-state error than large values of α. This is expected, as large values of α incur longer memory in the system such that more frames are used in the estimate, thus reducing its error. Additionally, Fig. 1(b) shows that using longer frame lengths L also lead to a lower steady-state error. This is due to reduced truncation errors in the estimate, as explained in Sec. 2.2. Doubling the frame length L also doubles the convergence time of estimates, such that convergence in frame number is constant. The mean aggregated NPM between the long-term estimate $\widetilde{\mathbf{R}}_{\mathbf{xx}}(\tau)$ in (5) and the ground truth was -79.09 dB.

3.4. Experiment 2: Impact on speech enhancement

This experiment evaluates the performance of the frame-based space-time covariance matrix for use in PEVD speech enhancement. The considered source is an IEEE sentence [30] recorded by a male native British English speaker, corrupted by spherical isotropic speech-shaped noise simulated using [31] at 0 dB SNR [27, 32]. Performance in this scenario can no longer be compared against a ground-truth matrix, as the recorded signal is no longer white Gaussian. Instead, this section compares the estimation accuracy of $\widehat{\mathbf{R}}_{\mathbf{xx}}^k(\tau)$ in (10) against the long-term estimate $\widetilde{\mathbf{R}}_{\mathbf{xx}}(\tau)$ in (5), and subsequently, the beampattern accuracy of $\widehat{\boldsymbol{u}}_1^k(z)$ against $\widetilde{\boldsymbol{u}}_1(z)$.

Fig. 2 shows the aggregated NPM between $\widehat{\mathbf{R}}_{\mathbf{xx}}^k(\tau)$ and $\widetilde{\mathbf{R}}_{\mathbf{xx}}(\tau)$, and the directivity index error between $\widehat{\boldsymbol{u}}_1^k(z)$ and $\widetilde{\boldsymbol{u}}_1(z)$ as a function of time, for various values of α and for $L = 40$ ms. The target speech signal $s[n]$ and the diffuse noise recorded at one microphone $v[n]$ are also shown to provide context to the analysis.

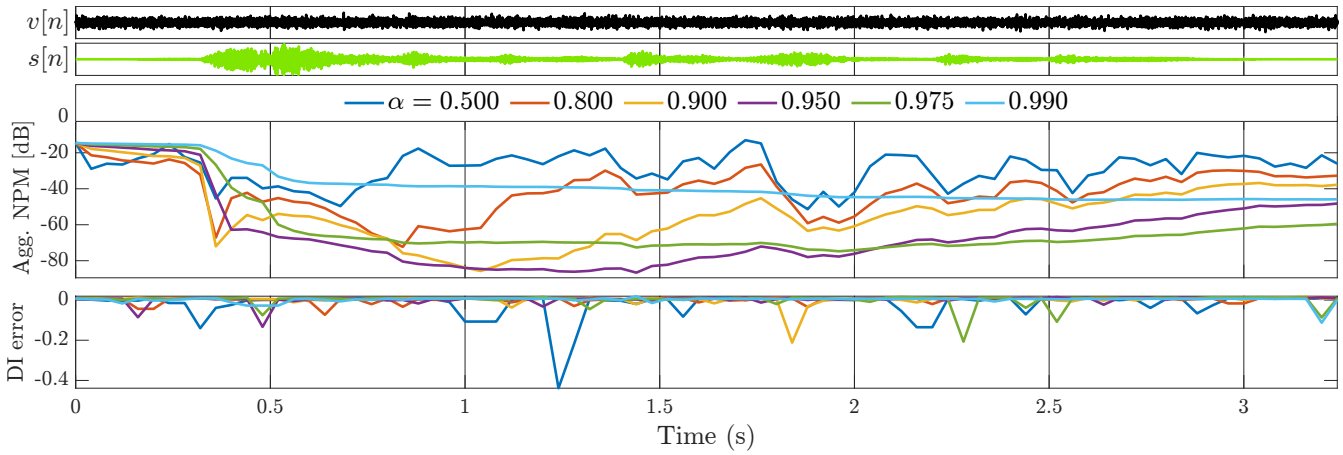


Fig. 2: Aggregated NPM and directivity index error between frame-based ($L = 40$ ms) and long-term ($L = 100$ ms) estimates of the space-time covariance matrix and resulting speech-enhancement filter, for various values of α.

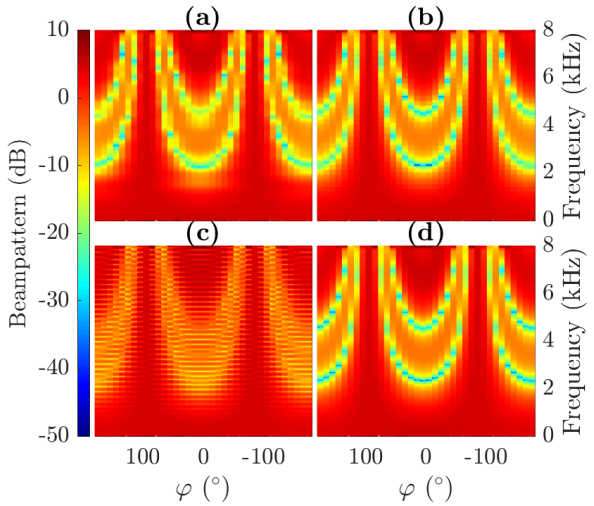


Fig. 3: Beampattern examples for the scenario in Sec. 3.4. **(a):** Long-term averaged $\widetilde{\mathbf{u}}_1(\tau)$. **(b), (c), (d):** $\widehat{\mathbf{u}}_1^k(\tau)$ with $\alpha = 0.9$ and at $t = \{1, 1.84, 3\}$ s.

For all α, the aggregated NPM drops with speech onset at 0.4 s, meaning the iterative estimates are close to the long-term average. For low values of α, the aggregated NPM fluctuates with speech activity. For example for $\alpha = 0.5$, the NPM increases at $t = 1.5$ s simultaneously with speech decay, but at $t = 1.7$ s the NPM drops as speech resumes. As the value of α increases, the NPM is less sensitive to changes in speech and varies slowly over time. For $\alpha = 0.99$, this leads to a higher steady-state aggregated NPM than $\alpha = 0.975$, as the NPM is less affected by the initial speech onset.

The directivity index error between $\widehat{\boldsymbol{u}}_1^k(z)$ and $\widetilde{\boldsymbol{u}}_1(z)$ is close to 0 dB for all α and at every time index. However, spurious error peaks occur at various time indices and for a few values of α. To investigate this behaviour, Fig. 3 shows the beampatterns of $\widetilde{\boldsymbol{u}}_1(z)$ and snapshots of $\widehat{\boldsymbol{u}}_1^k(z)$ with $\alpha = 0.9$ and for time indices $t = [1, 1.84, 3]$ s. Full animations can be found in [33]. All four beampatterns show beams steered to the target source direction at 90°. The snapshots at $t = 1$ s and $t = 3$ s exhibit very similar behaviours regardless of the fact that their corresponding aggregated NPMs are around -80 dB and -40 dB. Therefore, a large difference in space-time covariance matrix estimation does not necessarily lead to a large difference in PEVD filters for speech enhancement. Moreover, the snapshot at $t = 1.84$ s corresponds to a peak in DI error – it can be seen that the beam is steered in the correct direction but there is larger leakage to other azimuths. This implies that this suboptimal filter is still capable of retaining the source coming from the target direction, but might have reduced noise reduction capabilities.

Table 1: SNR and STOI improvements of the proposed method compared to the batch method in [1].

α	0.50	0.80	0.90	0.95	0.975	0.99
ΔSNR [dB]	1.29	1.28	1.18	1.08	0.80	0.49
ΔSTOI	0.01	0.03	0.02	0.01	0.02	0.02

Finally, the impact of the frame-based PEVD enhancement on speech enhancement metrics is investigated in Table 1, where ΔSNR and ΔSTOI denote the difference in SNR and STOI scores between $\widehat{y}(z)$ and the long-term average signal $\widetilde{y}(z)$, for various values of α. The table shows around 1 dB SNR improvement of the frame-based method over the classical PEVD enhancement, while STOI scores are broadly unaffected. The improvement in SNR is likely due to better noise reduction in noise-only frames.

4. CONCLUSION

This paper introduced a frame-based space-time covariance matrix estimation method for application to PEVD-based speech enhancement. The estimation procedure was found to converge to ground truth matrices, within at least -20 dB aggregated NPM. An experiment with speech enhancement showed that the frame-based method not only could in principle adapt to non-stationary acoustic scenarios, but it also yielded filters and enhancement measures similar to or better than the batch estimation in [1], thus opening the possibility to investigate real-time PEVD-based speech enhancement.

5. 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] 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.

[3] E. d'Olne, V. W. Neo, and P. A. Naylor, "Speech enhancement in distributed microphone arrays using polynomial eigenvalue decomposition," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2022.

[4] S. Weiss, N. J. Goddard, S. Somasundaram, I. K. Proudler, and P. A. Naylor, "Identification of broadband source-array responses from sensor second order statistics," in *Sensor Signal Process. for Defence Conf. (SSPD)*, 2017.

[5] M. A. Alrmah, S. Weiss, and S. Lambotharan, "An extension of the MUSIC algorithm to broadband scenarios using a polynomial eigenvalue decomposition," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2011, pp. 629–633.

[6] M. Alrmah, "Broadband angle of arrival estimation using polynomial matrix decompositions," Ph.D. dissertation, University of Strathclyde, Scotland, Oct. 2015.

[7] A. Hogg, V. Neo, C. Evers, S. Weiss, and P. Naylor, "A polynomial eigenvalue decomposition MUSIC approach for broadband sound source localization," in *Proc. IEEE Workshop on Appl. of Signal Process. to Audio and Acoust. (WASPAA)*, New Paltz, NY, Oct. 2021.

[8] V. W. Neo, S. Weiss, S. W. McKnight, A. O. T. Hogg, and P. A. Naylor, "Polynomial eigenvalue decomposition-based speaker activity detection in the presence of competing talkers," in *Proc. IEEE Workshop on Acoust. Signal Enhancement (IWAENC)*, Sep. 2022.

[9] A. Alzin, F. Coutts, J. Corr, S. Weiss, I. Proudler, and J. Chambers, "Adaptive broadband beamforming with arbitrary array geometry," in *Proc. IET Int. Conf. on Intelligent Signal Process.*, 2015, pp. 1–6.

[10] C. Delaosa, J. Pestana, N. J. Goddard, S. Somasundaram, and S. Weiss, "Support estimation of a sample space-time covariance matrix," in *Sensor Signal Process. for Defence Conf. (SSPD)*, 2019.

[11] C. Delaosa, J. Pestana, N. J. Goddard, S. Somasundaram, and S. Weiss, "Sample space-time covariance matrix estimation," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, May 2019, pp. 8033–8037.

[12] 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.

[13] 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.

[14] 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.

[15] J. Corr, K. Thompson, S. Weiss, I. K. Proudler, and J. G. McWhirter, "Impact of source model matrix conditioning on PEVD algorithms," in *Proc. IET Int. Conf. on Intelligent Signal Process.*, 2015, pp. 1–6.

[16] J. Corr, K. Thompson, S. Weiss, I. Proudler, and J. G. McWhirter, "Shortening of paraunitary matrices obtained by polynomial eigenvalue decomposition algorithms," in *Sensor Signal Process. for Defence Conf. (SSPD)*, 2015.

[17] J. Foster, J. G. McWhirter, and J. Chambers, "Limiting the order of polynomial matrices within the SBR2 algorithm," in *IMA Int. Conf. on Math. in Signal Process.*, 2006.

[18] C. Delaosa, F. K. Coutts, J. Pestana, and S. Weiss, "Impact of space-time covariance estimation errors on a parahermitian matrix EVD," in *Proc. IEEE Sensor Array and Multichannel Signal Process. Workshop (SAM)*, 2018, pp. 164–168.

[19] S. Haykin, *Adaptive Filter Theory*, 4th ed. Prentice Hall, 2002.

[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] 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.

[22] 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.

[23] 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.

[24] D. R. Morgan, J. Benesty, and M. Sondhi, "On the evaluation of estimated impulse responses," *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 174–176, Jul. 1998.

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

[26] 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.

[27] 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

[28] 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.

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

[30] 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.

[31] 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.

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

[33] E. d'Olne, "Iterative space-time covariance matrix estimation for PEVD-based speech enhancement," 2022. [Online]. Available: https://ed1016.github.io/adaptive_PEVD/