

POLYNOMIAL MATRIX EIGENVALUE DECOMPOSITION-BASED SOURCE SEPARATION USING INFORMED SPHERICAL MICROPHONE ARRAYS

Vincent W. Neo¹, Christine Evers^{2*}, Patrick A. Naylor^{1†}

¹ Department of Electrical and Electronic Engineering, Imperial College London, U.K.

² Electronics and Computer Science, University of Southampton, U.K.

Email: {vincent.neo09, p.naylor}@imperial.ac.uk, c.evers@soton.ac.uk

ABSTRACT

Audio source separation is essential for many applications such as hearing aids, telecommunications, and robot audition. Subspace decomposition approaches using polynomial matrix eigenvalue decomposition (PEVD) algorithms applied to the microphone signals, or lower-dimension eigenbeams for spherical microphone arrays, are effective for speech enhancement. In this work, we extend the work from speech enhancement and propose a PEVD subspace algorithm that uses eigenbeams for source separation. The proposed PEVD-based source separation approach performs comparably with state-of-the-art algorithms, such as those based on independent component analysis (ICA) and multi-channel non-negative matrix factorization (MNMF). Informal listening examples also indicate that our method does not introduce any audible artifacts.

Index Terms— Polynomial matrix eigenvalue decomposition, informed array processing, source separation, microphone arrays.

1. INTRODUCTION

The separation of audio sources is essential for many applications such as speech enhancement in hearing aids and telecommunications, three-dimensional (3D) sound rendering, and robot audition [1]. In real-life scenarios, the received signals often contain a mixture of multiple sources, reverberation, and ambient noise. These degradations severely limit the performance of these applications, motivating the need for source separation algorithms [2–4].

Multi-channel audio source separation algorithms include microphone array processing, independent component analysis (ICA)-based and multi-channel non-negative matrix factorization (MNMF) methods. Through filtering in space, array processing can potentially isolate the spatially separated target and interfering sources [5]. Among many array geometries, spherical microphone arrays have gained much interest due to their compact representation and efficient processing of the 3D sound field in the spherical harmonic (SH) domain [6, 7]. This approach, however, cannot completely remove spatially isotropic noise and reverberation [8].

In [9, 10], the microphone signals are first broken up into frames via the short-time Fourier transform (STFT) before being separated into sources by ICA in the frequency-domain (FDICA). However, inherent to ICA, the generated separating matrix incurs permutation and scaling ambiguities at every frequency. Methods based on the direction-of-arrival (DoA) and inter-frequency signal envelopes [11], and independent vector analysis (IVA) using auxiliary

function (AuxIVA) [12, 13] have been proposed to solve the permutation and scaling problem by grouping components from the same source over different frequencies. The processed frames are then synthesized to form the separated signals using the inverse STFT.

The MNMF approaches [14, 15] jointly estimate the mixing matrix and power spectrograms, which are further decomposed into two non-negative matrices representing the spectral patterns and temporal activations. The spectral and temporal matrices, together with the demixing matrix derived from the mixing matrix, are then used to generate the separated sources. The methods are, however, sensitive to parameter initialization [14]. The benefits of both IVA and non-negative matrix factorization (NMF) are unified under the independent low-rank matrix analysis (ILRMA) algorithm [3, 16].

Polynomial matrix eigenvalue decomposition (PEVD)-based algorithms [17–19] are effective for single-speaker speech enhancement but are computationally expensive for a large number of signals. Instead of directly applying the PEVD to the microphone signals on the spherical array, PEVD can be applied to the lower dimension eigenbeams and still provides good speech enhancement performance without introducing audible artifacts [20].

In this work, we extend the idea of [20] from speech enhancement to source separation. Informed of the source directions, we develop a source separation PEVD approach using only selected eigenbeams and beamformer outputs, to achieve low complexity. The novel contributions are as follows: (i) we incorporate a novel maximum directivity modal beamformer; (ii) we use PEVD to combine eigenbeams and beamformer outputs; and (iii) we develop a PEVD-based approach to source separation. The proposed approach is tested and compared against benchmark approaches. Listening examples are available [21].

2. PROBLEM FORMULATION

The received signal, $x_q(n, \mathbf{r}_q)$, at the q -th microphone on the spherical array of radius r with Q microphones, is

$$x_q(n, \mathbf{r}_q) = \sum_{p=1}^P \mathbf{h}_{p,q}^T \mathbf{s}_p(n), \quad q = 1, \dots, Q, \quad (1)$$

where $\mathbf{h}_{p,q} = [h_{p,q}(0), \dots, h_{p,q}(J)]^T$ represents the room impulse response (RIR) from the p -th source to the q -th microphone, modelled as a J -th order finite impulse response filter, $\mathbf{s}_p(n) = [s_p(n), \dots, s_p(n-J)]^T$ is the p -th localized source signal, n is the sample index and $[\cdot]^T$ is the transpose operator. The spherical coordinate of the q -th microphone relative to the array center is $\mathbf{r}_q = (r, \theta_q, \phi_q)$, where θ_q and ϕ_q , respectively, are the elevation and azimuth angles measured downwards from the z -axis and from the x -axis towards the y -axis. The received signals at the array are $\mathbf{x}(n, \mathbf{r}) = [x_1(n, \mathbf{r}_1), \dots, x_Q(n, \mathbf{r}_Q)]^T$.

*This work is funded through the U.K. EPSRC Fellowship grant no. EP/P001017/1.

†This work is funded through the U.K. EPSRC grant no. EP/S035842/1.

3. MULTI-CHANNEL ARRAY PROCESSING

3.1. Spherical Array Processing

The real-valued spherical harmonic transform (SHT) of the spatially sampled sound field is approximated by [22]

$$\chi_\ell^m(n) \approx \sum_{q=1}^Q \alpha_q x_q(n, \mathbf{r}_q) R_\ell^m(\mathbf{r}_q), \quad (2)$$

where α_q is the quadrature weight, $\chi_\ell^m(n)$ is the ℓ -th order, m -th degree time-domain eigenbeam signal associated with the real-valued SH basis function, $R_\ell^m(\mathbf{r}_q)$, defined as [7]

$$R_\ell^m(\mathbf{r}_q) = \begin{cases} \sqrt{2}(-1)^m \Im\{Y_\ell^{[m]}(\mathbf{r}_q)\} & m < 0 \\ Y_\ell^0(\mathbf{r}_q) & m = 0 \\ \sqrt{2}(-1)^m \Re\{Y_\ell^{[m]}(\mathbf{r}_q)\} & m > 0 \end{cases}, \quad (3)$$

where $Y_\ell^m(\mathbf{r}_q) = \sqrt{\frac{(2\ell+1)}{4\pi} \frac{(\ell-m)!}{(\ell+m)!}} P_\ell^m(\cos \theta_q) e^{jm\phi_q}$ and $P_\ell^m(\cdot)$ is the associated Legendre function, $|\cdot|$ denotes modulus, \Re and \Im are the real and imaginary parts of a complex number. Any function on the sphere can be expressed using spherical harmonics as [6, 7]

$$x_q(n, \mathbf{r}_q) = \sum_{\ell=1}^L \sum_{m=-\ell}^{\ell} \chi_\ell^m(n) R_\ell^m(\mathbf{r}_q). \quad (4)$$

Alias-free spatial reconstruction of the sound field can be achieved if $Q \geq (L+1)^2$, where L is the maximum SH order of the sound field. Dropping the sample index n for brevity, the vector of eigenbeams is $\chi = [\chi_0^0, \chi_1^{-1}, \chi_1^0, \dots, \chi_L^L]^T$, with elements arranged in ascending SH order and degree. The generation of eigenbeam signals in (2) is also called an eigenbeamformer [23, 24]. Modal beamforming, or the weighted sum of the eigenbeams, produces a beam pattern directed at a desired direction using

$$\psi(n) = \sum_{\ell=1}^L \sum_{m=-\ell}^{\ell} w_\ell^m \chi_\ell^m(n), \quad (5)$$

where w_ℓ^m is the beamformer weight associated with $\chi_\ell^m(n)$ or in vector form, $\psi(n) = \mathbf{w}^T \chi$, where $\mathbf{w} = [w_0^0, w_1^{-1}, w_1^0, w_1^1, \dots, w_L^L]^T$ and $\psi(n) = [\psi_1(n), \dots, \psi_L(n)]^T$, where L denotes the total number of modal beamformer outputs.

3.2. Polynomial Matrix Eigenvalue Decomposition

The space-time covariance matrix [25, 26], parameterized by the time lag τ , is computed using the microphone signals in (1) to give

$$\mathbf{R}_{\mathbf{x}\mathbf{x}}(\tau) = \mathbb{E}\{\mathbf{x}(n)\mathbf{x}^T(n-\tau)\}, \quad (6)$$

where $\mathbb{E}\{\cdot\}$ is the expectation operator over n . Each element, $r_{p,q}(\tau)$, calculated using the correlation between the p -th and q -th microphone signals, produces auto- and cross-correlations on the diagonals and off-diagonals, respectively. The concatenation of the covariance matrices for all values of $\tau \in \{-N, \dots, N\}$ gives a 3D-tensor of dimension, $Q \times Q \times (2N+1)$.

The z -transform of (6) is a para-Hermitian polynomial matrix

$$\mathbf{R}_{\mathbf{x}\mathbf{x}}(z) = \sum_{\tau=-\infty}^{\infty} \mathbf{R}_{\mathbf{x}\mathbf{x}}(\tau) z^{-\tau}, \quad (7)$$

satisfying $\mathbf{R}_{\mathbf{x}\mathbf{x}}(z) = \mathbf{R}_{\mathbf{x}\mathbf{x}}^P(z) = \mathbf{R}_{\mathbf{x}\mathbf{x}}^H(1/z^*)$, where $[\cdot]^*$, $[\cdot]^H$, $[\cdot]^P$ are the complex conjugate, Hermitian and para-Hermitian op-

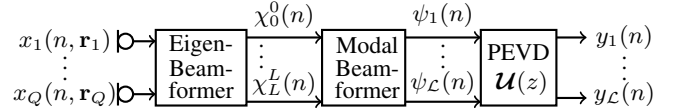


Figure 1: Block diagram of the proposed method using a sequence of eigen, modal and PEVD beamformers for each source.

erators respectively. The PEVD of (7) is [27]

$$\mathbf{R}_{\mathbf{x}\mathbf{x}}(z) \approx \mathbf{U}^P(z) \mathbf{\Lambda}(z) \mathbf{U}(z), \quad (8)$$

where the rows of $\mathbf{U}(z)$ are the polynomial eigenvectors and the elements on the diagonal matrix $\mathbf{\Lambda}(z)$ are the polynomial eigenvalues. Iterative algorithms based on the second-order sequential best rotation (SBR2) [27–29] and the sequential matrix diagonalization (SMD) [30, 31] are used to calculate (8).

At each iteration, the PEVD algorithm will first search for the off-diagonal element with the largest magnitude. If its magnitude exceeds a predefined threshold, a delay polynomial matrix is applied to bring the element to the z^0 plane. A unitary matrix, which is designed to zero out the element, is applied to the entire polynomial matrix. A trimming procedure [27] is also used to keep the polynomial order compact. The algorithm terminates when the magnitudes of all off-diagonal elements are less than the threshold or when the user-defined maximum iteration number is reached.

4. INFORMED SPHERICAL ARRAY USING PEVD

In this paper, we consider the problem of separating multiple directional sources in a reverberant environment. In contrast, our previous work in [20] addressed the enhancement of a single-speaker in diffuse noise. Moreover, in [20], we applied PEVD to all eigenbeams instead of microphone signals to achieve a reduction in computational complexity [31]. However, using the complete set of computed eigenbeams in the PEVD results in undesired reverberant components and interfering signals that leak through the eigenbeam sidelobes and, therefore, reduce separation performance.

To improve the quality of the separated signals while reducing complexity, we propose in this paper to use for the PEVD only a subset of eigenbeams and modal beamformer outputs rather than the complete set of eigenbeams or microphone signals. We first apply SHT to the microphone signals to generate eigenbeams and use the resulting eigenbeams to form modal outputs such as the hypercardioid beams. The beamformed signals, including unprocessed or steered eigenbeams, are then used as inputs to the PEVD for polynomial subspace decomposition. Fig. 1 shows a block diagram of the proposed approach.

4.1. Modified Hyper-Cardioid Beamformer (MHCARD)

The work in [23] proposes a hyper-cardioid beamformer that achieves the maximum directivity index by using $(L+1)^2$ eigenbeams. Since the zeroth-order eigenbeam χ_0^0 is similar to an omnidirectional microphone placed at the center of the array [7], it captures also undesired signals arriving from non-target source directions. Each non-zero order eigenbeam, akin to a directional sensor placed at the array center, can be steered or directed at the source signal in a specific direction. Therefore, the beamformer coefficients, used to combine the eigenbeams, are modified to limit undesired signals from non-target directions (see Section 5.3). Accordingly, the modal weight $w_0^0 = 0$ and the other coefficients are scaled by $\frac{L+1}{\sqrt{(L+1)^2 - 1}}$ for unity gain in the source direction (θ_p, ϕ_p) .

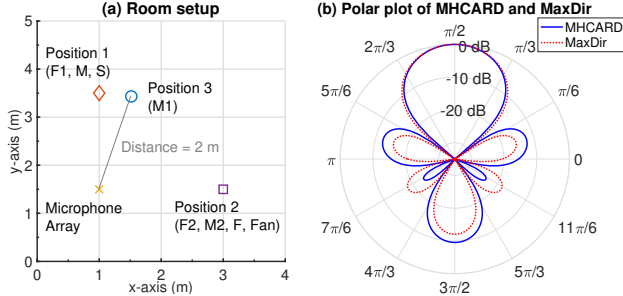


Figure 2: Configuration of room and MHCARD for Source 1.

4.2. PEVD of Modal Beamformer Outputs

Assuming source directions can be estimated [32], the eigenbeams, modified hyper-cardioid beams, or any beamformer output can be conveniently chosen or steered using Wigner-D rotation matrices [6, 7] to generate \mathcal{L} modal outputs, $\psi(n)$. For each source, the space-time covariance matrix of $\psi(n)$ is

$$\mathbf{R}_{\psi\psi}(\tau) = \mathbb{E}\{\psi(n)\psi^T(n-\tau)\} \quad (9)$$

and its z -transform $\mathbf{R}_{\psi\psi}(z)$ is computed using (7). The PEVD is

$$\mathbf{R}_{\psi\psi}(z) \approx \mathbf{U}_{\psi}^P(z) \mathbf{\Lambda}_{\psi}(z) \mathbf{U}_{\psi}(z), \quad (10)$$

where $\mathbf{\Lambda}_{\psi}(z)$ and $\mathbf{U}_{\psi}(z)$ are the eigenvalues and eigenvectors associated with the modal outputs $\psi(z)$ and the PEVD output is

$$\mathbf{y}(z) = \mathbf{U}_{\psi}(z) \psi(z). \quad (11)$$

Consequently, the spectrally majorized outputs generated by the PEVD enable the extraction of the target source in the first channel. This process is repeated for all P sources.

5. SIMULATION AND RESULTS

5.1. Experimental Setup

Anechoic speech signals sampled at 16 kHz are taken from the TIMIT corpus [33]. The SMIRgen tool [34] is used to generate RIRs from 2 sources to 32 microphones on a rigid sphere with $r = 4.2$ cm. The room setup is shown in Fig. 2a, and the T_{60} is varied between 0 s, 0.3 s and 0.7 s. The source signals used in the experiments include male and female speakers as well as fan noise.

In each experiment, 50 trials are conducted. For each speech source, short utterances from a randomly selected speaker are concatenated to generate signals of 8 to 10 s duration. Each source signal is convolved with the RIR and mixed along with 30 dB sensor noise to generate the microphone signals.

The parameters for the PEVD are based on [20]. For each target source, $\mathcal{L} = 4$ instead of 32 microphone signals are used as inputs to the PEVD. The modal signals for source position 1 are $\chi_1^{-1}, \chi_3^{-3}, \chi_3^{-1}$ and the modified hyper-cardioid (MHCARD) beam directed at $(\frac{\pi}{2}, \frac{\pi}{2})$ as shown in Fig. 2b. For source 2, the modal signals are $\chi_1^1, \chi_3^1, \chi_3^3$ and the MHCARD beam directed at $(\frac{\pi}{2}, 0)$. The PEVD input for source position 3 is the set of signals for position 1 but steered towards $(\frac{\pi}{2}, \frac{5\pi}{12})$ using Wigner-D matrices [6].

The proposed approach is compared against the third-order hyper-cardioid with maximum directivity index (MaxDir) [23], auxiliary function-based IVA (AuxIVA) [12], independent low-rank matrix analysis (ILRMA) [3] and fast MNMF (FastMNMF) [15], implemented in [35]. During experimentation, we found that the ICA and MNMF-based methods do not perform well when all 32 microphones are used, and signals from two microphones closest to each source are chosen instead.

Table 1: A source separation example involving 2 female speakers in anechoic environment and results for 1 female speaker.

Algorithm	Δ SDR	Δ SIR	Δ SAR	Δ STOI	Δ PESQ
AuxIVA	17.7 dB	25.3 dB	11.4 dB	0.21	1.05
AuxIVA32	-1.1 dB	4.5 dB	-6.1 dB	-0.30	-0.07
FastMNMF	20.6 dB	35.2 dB	13.8 dB	0.21	1.28
ILRMA	19.5 dB	31.3 dB	12.8 dB	0.21	1.21
MaxDir	3.9 dB	3.4 dB	4.7 dB	0.07	0.22
MHCARD	16.9 dB	17.8 dB	13.4 dB	0.21	0.93
PEVD	21.8 dB	25.3 dB	16.4 dB	0.24	1.39

5.2. Evaluation Measures

For evaluation, source-to-distortion ratio (SDR), source-to-interferences ratio (SIR), and source-to-artifacts ratio (SAR) are used to measure the overall source separation ability, interference rejection and processing artifacts, respectively [36]. Short-time objective intelligibility (STOI) [37] and perceptual evaluation of speech quality (PESQ) [38] are also used to measure speech intelligibility and quality. The metrics are computed for the reference microphone and processed signals and their difference Δ is reported. Positive Δ values indicate improvements in all measures.

5.3. Experiments and Discussion

Table 1 summarizes the results for a scenario involving the separation of 2 female speakers in an anechoic room. ICA-based methods such as AuxIVA works better for 2 microphones than 32 microphones (AuxIVA32) by 18.8 dB in Δ SDR, and signals from two microphones closest to each source are chosen. Although more data is available with 32 microphones, a greater number of permutations needs to be considered and may have resulted in a poorer demixing matrix, accounting for the artifacts and reduction in STOI.

Across all metrics, the larger improvement offered by the modified hyper-cardioid (MHCARD) over the original hyper-cardioid beam (MaxDir) highlights the advantages of excluding the zeroth-order eigenbeam χ_0^0 in the modal beamformer. MHCARD provided an overall greater reduction in sidelobe levels as shown in Fig. 2b. FastMNMF leads Δ SIR by 35.2 dB and is closely followed by ILRMA, AuxIVA and PEVD. PEVD outperforms other algorithms in Δ SDR, Δ SAR, Δ STOI and Δ PESQ by 21.75 dB, 16.38 dB, 0.237 and 1.39, respectively. Listening examples in [21] also indicate that PEVD does not introduce processing artifacts, and non-target speech signals are attenuated but remain intelligible.

These findings are also observed in Fig. 3 for 50 trials involving two female speakers, F1 and F2 at source positions 1 and 2, respectively. The experiment is extended to scenarios involving different source types, e.g., male and female speakers under M and F, a single speaker and localized fan under S and FAN, and different source positions using two male speakers, M1 and M2 at source positions 3 and 2, as shown in Fig. 2a. When different source types are used in an anechoic room, the source separation of the PEVD is comparable with FastMNMF and ILRMA and is usually better than AuxIVA. At different source positions M1 and M2, MHCARD is better than MaxDir by up to 10 dB in Δ SDR. PEVD, which uses the MHCARD, achieves an average of 8 dB improvement in Δ SDR.

When the room has a reverberation time of 0.3 s, Fig. 4 shows the source separation results. FastMNMF outperforms all algorithms in Δ SDR and Δ SIR but is closely followed by ICA-based approaches when the scenario involves only speech signals. PEVD observe an average of 5 dB improvement in SDR and SIR, but a more considerable increase by up to 20 dB in the single speech

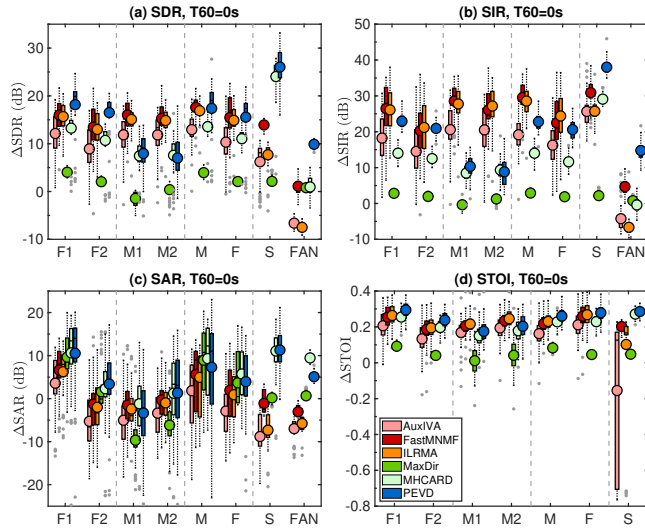
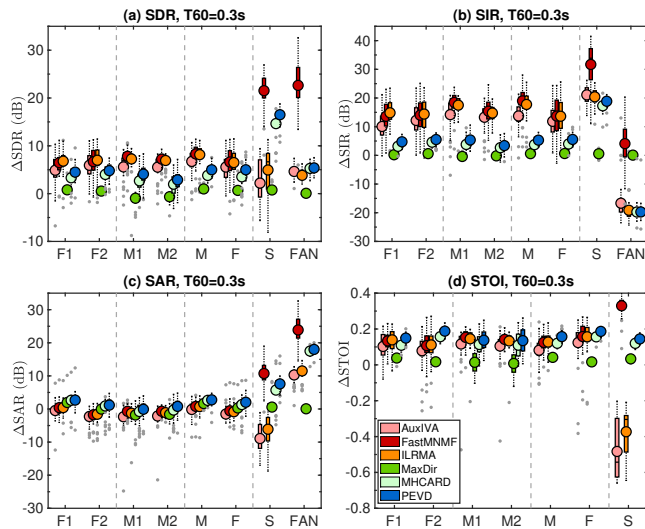
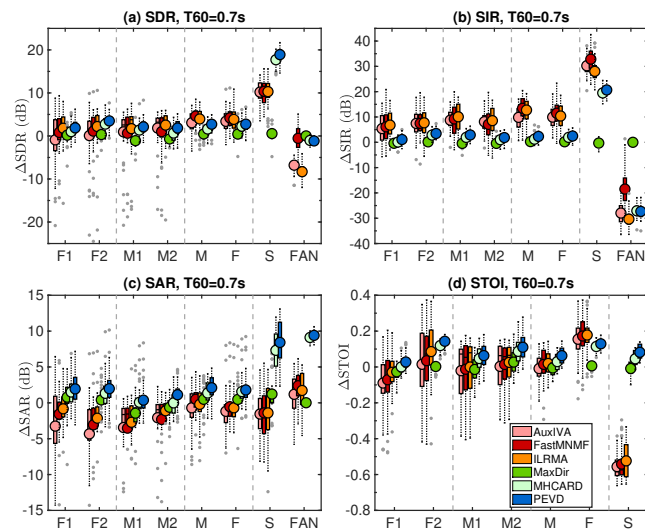


Figure 3: Box plot for source separation in anechoic room.

Figure 4: Box plot for source separation in room with $T_{60} = 0.3s$.Figure 5: Box plot for source separation in room with $T_{60} = 0.7s$.

and fan scenario. PEVD also usually performs better in ΔSAR and $\Delta STOI$ than the ICA-based approaches in all cases.

In the room with a $T_{60} = 0.7s$, FastMNMF performs less well than the mildly reverberant $0.3s$ case. As shown in Fig. 5, PEVD is comparable to the ICA-based methods in ΔSDR , worse in ΔSIR but best in ΔSAR and $\Delta STOI$ in most cases including different source positions. For the single speaker and fan scenario, PEVD performs better in SDR and SAR for both sources as it does not rely on source density functions which may not model fan noise well.

We also computed results for the PEVD algorithm using the original hyper-cardioid, MaxDir, instead of MHCARD. Results across various scenarios indicate that using MHCARD offers an improvement of up to 12 dB in ΔSDR over using MaxDir for PEVD.

6. CONCLUSION

We have proposed a PEVD-based source separation method based on informed spherical arrays. With knowledge of the source directions, standard modal beamformers, eigenbeam signals, and the proposed modified hyper-cardioid beam can be combined by the PEVD algorithm. The proposed PEVD-based approach is among the best performing algorithm and seems more robust to reverberation than the other algorithms, but does not always outperform them. Listening examples [21] also support the SAR and STOI scores, indicating that PEVD achieves separation without introducing any audible artifacts. In comparison, other algorithms considerably distort the signal as part of performing the separation.

7. REFERENCES

- [1] E. Vincent, N. Bertin, R. Gribonval, and F. Bimbot, "From blind to guided audio source separation: How models and side information can improve the separation of sound," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 107–115, Apr. 2014.
- [2] S. Makino, T.-W. Lee, and H. Sawada, Eds., *Blind Speech Separation*, ser. Signals and Communication Technology. Dordrecht: Springer-Verlag, 2007.
- [3] S. Makino, *Audio Source Separation*, ser. Signals and Communication Technology. Springer-Verlag, 2018.
- [4] E. Vincent, T. Virtanen, and S. Gannot, *Audio Source Separation and Speech Enhancement*. John Wiley & Sons, Inc., 2018.
- [5] M. S. Brandstein and D. B. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin, Germany: Springer-Verlag, 2001.
- [6] B. Rafaely, *Fundamentals of Spherical Array Processing*, ser. Springer Topics in Signal Processing. Springer, 2015.
- [7] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, *Theory and Applications of Spherical Microphone Array Processing*, ser. Springer Topics in Signal Processing, 2017.
- [8] P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation*. Springer-Verlag, 2010.
- [9] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 2, pp. 109–116, Mar. 2003.
- [10] S. Araki, S. Makino, R. Mukai, and H. Saruwatari, "Equivalence between frequency domain blind source separation and frequency domain adaptive null beamformers," in *Proc. Eur.*

- Conf. on Speech Commun. and Technol.*, Aalborg Denmark, Scandinavia, Sept. 2001, pp. 2595–2598.
- [11] H. Sawada, S. Araki, and S. Makino, “A robust and precise method for solving the permutation problem of frequency-domain blind source separation,” *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 530–538, Sept. 2004.
 - [12] N. Ono, “Stable and fast update rules for independent vector analysis based on auxiliary function technique,” in *Proc. IEEE Workshop on Appl. of Signal Process. to Audio and Acoust. (WASPAA)*, Oct. 2011, pp. 189–192.
 - [13] J. Janský, Z. Koldovský, and N. Ono, “A computationally cheaper method for blind speech separation based on AuxIVA and incomplete demixing transform,” in *Proc. Int. Workshop on Acoust. Signal Enhancement (IWAENC)*, Sept. 2016, pp. 1–5.
 - [14] A. Ozerov and C. Févotte, “Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, No. 3, pp. 550–563, 2010.
 - [15] K. Sekiguchi, A. A. Nugraha, Y. Bando, and K. Yoshii, “Fast multichannel source separation based on jointly diagonalizable spatial covariance matrices,” in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2019.
 - [16] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, “Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 24, no. 9, pp. 1626–1641, Sept. 2016.
 - [17] V. W. Neo, C. Evers, and P. A. Naylor, “Speech enhancement using polynomial eigenvalue decomposition,” in *Proc. IEEE Workshop on Appl. of Signal Process. to Audio and Acoust. (WASPAA)*, 2019, pp. 125–129.
 - [18] V. W. Neo, C. Evers, and P. A. Naylor, “PEVD-based speech enhancement in reverberant environments,” in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2020, pp. 186–190.
 - [19] V. W. Neo, C. Evers, and P. A. Naylor, “Speech dereverberation performance of a polynomial-EVD subspace approach,” in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2020, pp. 221–225.
 - [20] 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)*, June 2021.
 - [21] V. W. Neo, C. Evers, and P. A. Naylor, “PEVD-based Source Separation Using Informed Spherical Microphone Arrays,” Apr. 2021. [Online]. Available: <https://vwn09.github.io/pevd-separate/>
 - [22] B. Rafaely, “Analysis and design of spherical microphone arrays,” *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 135–143, Jan. 2005.
 - [23] J. Meyer and G. Elko, “A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield,” in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, vol. 2, May 2002, pp. 1781–1784.
 - [24] J. Meyer and G. W. Elko, “Spherical microphone arrays for 3D sound recording,” in *Audio Signal Processing For Next-Generation Multimedia Communication Systems*, Y. Huang and J. Benesty, Eds. Kluwer Academic Publisher, 2004, pp. 67–89.
 - [25] S. Weiss, J. Pestana, and I. K. Proudler, “On the existence and uniqueness of the eigenvalue decomposition of a parahermitian matrix,” *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2659–2672, May 2018.
 - [26] S. Weiss, “Mathematical tools for processing broadband multi-sensor signals,” in *Proc. Int. Conf. on Multimedia and Human-Comput. Interaction*, Aug. 2020.
 - [27] 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.
 - [28] S. Redif, S. Weiss, and J. G. McWhirter, “An approximate polynomial matrix eigenvalue decomposition algorithm for para-hermitian matrices,” in *Proc. Int. Symp. on Signal Process. and Inform. Technol. (ISSPIT)*, 2011, pp. 421–425.
 - [29] 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.
 - [30] 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.
 - [31] F. K. Coutts, I. K. Proudler, and S. Weiss, “Efficient implementation of iterative polynomial matrix EVD algorithms exploiting structural redundancy and parallelisation,” *IEEE Trans. Circuits Syst. I*, vol. 66, no. 12, pp. 4753–4766, Dec. 2019.
 - [32] C. Evers, H. W. Löllmann, H. Mellmann, A. Schmidt, H. Barfuss, P. A. Naylor, and W. Kellermann, “The LOCATA challenge: Acoustic source localization and tracking,” *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 28, pp. 1620–1643, Apr. 2020.
 - [33] 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.
 - [34] D. P. Jarrett, E. A. P. Habets, M. R. P. Thomas, and P. A. Naylor, “Rigid sphere room impulse response simulation: Algorithm and applications,” *J. Acoust. Soc. Am.*, vol. 132, no. 3, pp. 1462–1472, Sept. 2012.
 - [35] R. Scheibler, E. Bezzam, and I. Dokmanić, “Pyroomacoustics: A python package for audio room simulation and array processing algorithms,” in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Apr. 2018, pp. 351–355.
 - [36] E. Vincent, R. Gribonval, and C. Févotte, “Performance measurement in blind audio source separation,” *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 4, pp. 1462–1469, July 2006.
 - [37] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, “A short-time objective intelligibility measure for time-frequency weighted noisy speech,” in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Dallas, Texas, USA, Mar. 2010, pp. 4214–4217.
 - [38] “Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,” Int. Telecommun. Union (ITU-T), Recommendation P.862, Nov. 2003.