POLYNOMIAL MATRIX EIGENVALUE DECOMPOSITION OF SPHERICAL HARMONICS FOR SPEECH ENHANCEMENT

Vincent W. Neo[1], 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

Speech enhancement algorithms using polynomial matrix eigenvalue decomposition (PEVD) have been shown to be effective for noisy and reverberant speech. However, these algorithms do not scale well in complexity with the number of channels used in the processing. For a spherical microphone array sampling an order-limited sound field, the spherical harmonics provide a compact representation of the microphone signals in the form of eigenbeams. We propose a PEVD algorithm that uses only the lower dimension eigenbeams for speech enhancement at a significantly lower computation cost. The proposed algorithm is shown to significantly reduce complexity while maintaining full performance. Informal listening examples have also indicated that the processing does not introduce any noticeable artefacts.

Index Terms— Polynomial matrix, speech enhancement, spherical microphone arrays, multichannel signal processing

1. INTRODUCTION

The acquisition of speech is important for many applications such as hearing aids, telecommunications and automatic speech recognition. The performance of these applications is usually reduced because the received signal is often degraded by background noise and reverberation. To overcome these degradations, microphone array processing techniques have been proposed for speech enhancement.

Among many possible array geometries, spherical arrays have gained much interest [1,2]. Spherical microphone arrays can provide a compact representation and efficient processing of the 3D sound field in the spherical harmonic (SH) domain. They are used in sound field decomposition [3], 3D sound reproduction [4, 5] and for applications such as sound zones, spatial audio and virtual reality.

Beamforming using spherical arrays has been proposed for localization and tracking [6, 7], noise reduction [8] and dereverberation [9]. However, the speech enhancement problem considered here requires suppression of both noise and reverberation. This problem has been addressed in, for example, [10, 11] but with the requirement of prior information or with performance dependent on the reliability of direction-of-arrival and/or relative transfer function estimators.

In [12–14], a polynomial matrix eigenvalue decomposition (PEVD)-based speech enhancement approach based on the second-order sequential best rotation (SBR2) [15–17] or sequential matrix diagonalization (SMD) [18] algorithm, has been proposed. This approach can suppress noise and reverberation while improving speech intelligibility and quality, without introducing any audible artefacts and without the requirement to estimate any acoustic parameters. While it has been shown to be robust for linear and arbitrary array geometries in noisy reverberant environments, the algorithm does not scale well in complexity with the number of channels used [19].

In this paper, we propose a novel PEVD-based speech enhancement algorithm that operates on eigenbeams developed in the SH domain, rather than the raw microphone signals in the time-domain. This capitalizes on the compact representation of spherical microphone array signals using spherical harmonics and exploits the capabilities for speech enhancement offered by the PEVD algorithm. The proposed algorithm is compared and evaluated against benchmark approaches using simulated and measured room impulse responses and noise signals. Informal listening examples, available at [20], highlight that speech enhancement is achieved with no significant processing artefacts.

2. PROBLEM FORMULATION

The noisy and reverberant speech signal arriving at the q-th microphone on a spherical array at time sample n, is

$$x_q(n) = \mathbf{h}_q^T \mathbf{s}_0(n) + v_q(n), \quad n = 0, \ldots, N, \tag{1}$$

where $\mathbf{h}_q = \left[h_{q,0}, \ldots, h_{q,J}\right]^T$ is the acoustic channel from the source to the q-th microphone, modelled as a J-th order finite impulse response filter, $\mathbf{s}_0(n) = \left[s_0(n), \ldots, s_0(n-J)\right]^T$ is the anechoic speech signal vector, $v_q(n)$ is the additive noise, that is uncorrelated with the speech component, and $[\cdot]^T$ is the transpose operator. To express the microphone signals in terms of the array geometry explicitly, (1) can be written as $x_q(n) \triangleq x(n, \mathbf{r}_q)$, where $\mathbf{r}_q = (r, \theta_q, \phi_q)$ is the location of the q-th microphone relative to the array centre expressed in spherical polar coordinates, r is the radius, θ_q and ϕ_q are elevation and azimuth angles. The received signals at the array are $\mathbf{x}(n) = \left[x_1(n), \ldots, x_Q(n)\right]^T$ with $\mathbf{v}(n)$ similarly defined, and $\mathbf{x}(n, \mathbf{r}) = \left[x(n, \mathbf{r}_1), \ldots, x(n, \mathbf{r}_Q)\right]^T$ is used when the array geometry is explicitly modelled.

3. MULTICHANNEL ARRAY PROCESSING

3.1. Spherical Harmonic Decomposition

The real spherical harmonic transform (SHT) of the sound field, which is spatially sampled using Q microphones on a spherical array, is approximated by [21]

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

†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.

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 SH basis function, $R_\ell^m(\mathbf{r}_q)$, defined as [2]

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

where $Y_\ell^m(\mathbf{r}) = \sqrt{\frac{(2\ell+1)}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}} P_\ell^m(\cos\theta) e^{jm\phi}$, $P_\ell^m(\cdot)$ is the associated Legendre function, $\Re$ and $\Im$ denote the real and imaginary parts of a complex number. Any function on the sphere can be expressed as a weighted combination of the spherical harmonics given by [1, 2]

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

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. The sound field can then be represented by $\mathcal{L} = (L+1)^2$ eigenbeam signals in the time-domain and is compactly written as $\boldsymbol{\chi}(n) = \left[\chi_0^0(n), \chi_{-1}^1(n), \chi_0^1(n), \ldots, \chi_L^L(n)\right]^T$, with elements arranged in ascending SH order and degree.

3.2. Polynomial Matrix Eigenvalue Decomposition

The space-time covariance matrix, which is parameterized by the temporal lag τ, is defined as [15]

$$\mathbf{R}_{\mathbf{xx}}(\tau) = \mathbb{E}\{\mathbf{x}(n)\mathbf{x}^T(n-\tau)\}, \tag{5}$$

where the $(p, q)^{\text{th}}$ element, $r_{pq}(\tau)$, is computed using the correlation between the p-th and q-th microphone signals and $\mathbb{E}\{\cdot\}$ is the expectation operation over n. This produces auto- and cross-correlations on the diagonals and off-diagonals of $\mathbf{R}_{\mathbf{xx}}(\tau)$, respectively.

Classical subspace-based approaches for narrowband signals evaluate (5) only at $\tau = 0$ and then the received signals are decorrelated using an eigenvalue decomposition (EVD). The instantaneous spatial covariance matrix, $\mathbf{R}_{\mathbf{xx}}(0)$, has been shown to be unable to represent fully the correlation structure of broadband signals like speech [12, 15]. Decorrelation of speech signals is more completely achieved by considering a range of time lags. Accordingly, the concatenation of the covariance matrices in (5) for all values of $\tau \in \{-N, \ldots, N\}$ gives a 3D-tensor of dimensions, $Q \times Q \times (2N+1)$.

Speech signals, which are typically processed using the short-time Fourier transform, will require further expansion into a 4D-tensor of dimensions, $Q \times Q \times (2N+1) \times K$, where K is the number of frequency bins. However, this approach divides broadband signals into multiple narrowband signals, ignoring the correlations between frequency bands and neglecting phase coherence across bands [22]. An alternative representation is the z-transform and, because of symmetry, the z-transform of (5) is a para-Hermitian polynomial matrix

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

satisfying $\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z) = \boldsymbol{\mathcal{R}}_{\mathbf{xx}}^P(z) = \boldsymbol{\mathcal{R}}_{\mathbf{xx}}^H(z^{-1})$, where $[\cdot]^H$ and $[\cdot]^P$ are the Hermitian and para-Hermitian operators respectively. The PEVD of (6) is [15]

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

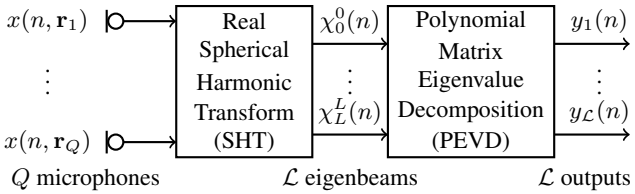

Fig. 1. Block diagram of the proposed algorithm.

where the rows of $\boldsymbol{\mathcal{U}}_{\mathbf{x}}(z)$ are the polynomial eigenvectors and the elements on the diagonal matrix $\boldsymbol{\Lambda}_{\mathbf{x}}(z)$ are the polynomial eigenvalues. To compute (7), iterative algorithms based on the SBR2 [15–17] and the SMD [18] have been proposed.

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 δ, 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 [15] based on μ, a fraction of the total Frobenius-norm squared, is also used to keep the polynomial order compact. The algorithm terminates when the magnitudes of all off-diagonal elements are less than δ or when the user-defined maximum iteration number, ι, is reached.

4. PEVD-BASED ENHANCEMENT USING EIGENBEAMS

The PEVD-based algorithm [12–14] has been shown to be robust and effective for speech enhancement in diverse acoustic environments. $\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z)$, on which the PEVD operates, has $Q^2(2N+1)$ elements and a computational complexity, $\mathcal{O}(Q^3 N)$ due to matrix multiplications applied to every lag [19] and is of much larger complexity than the SHT, which uses matrix-vector products. Hence, the complexity can be reduced by compressing the space-time covariance matrix using the SHT which reduces Q to $\mathcal{L}$, where typically $\mathcal{L} \ll Q$, if the enhancement performance is not significantly compromised.

In our approach, instead of using Q microphone signals to represent the L-th order sound field around the spherical array, a lower dimension $\mathcal{L}$ eigenbeams are used. The SHT uses all microphone signals for processing and this may reduce diffuse sensor noise in each eigenbeam [2]. However, this approach only decomposes the sound field spatially and we aim to improve speech enhancement performance for other noise types and reverberation by also exploiting temporal correlations via PEVD. Accordingly, we aim to achieve further improvement by using the eigenbeams as inputs for the PEVD algorithm, as summarized in Fig. 1 and Algorithm 1.

The space-time covariance matrix of the eigenbeams is

$$\mathbf{R}_{\boldsymbol{\chi}\boldsymbol{\chi}}(\tau) = \mathbb{E}\{\boldsymbol{\chi}(n)\boldsymbol{\chi}^T(n-\tau)\} = \boldsymbol{\Psi}^T \mathbf{R}_{\mathbf{xx}}(\tau) \boldsymbol{\Psi}, \tag{8}$$

where $\mathbf{R}_{\mathbf{xx}}(\tau)$ is the space-time covariance matrix of the raw microphone signals in (5) and $\boldsymbol{\Psi}$, is defined by

$$\boldsymbol{\Psi} = \begin{bmatrix} R_0^0(\mathbf{r}_1), & \ldots, & R_L^L(\mathbf{r}_1) \\ \vdots & \ddots & \vdots \\ R_0^0(\mathbf{r}_Q) & \ldots & R_L^L(\mathbf{r}_Q) \end{bmatrix} \in \mathbb{R}^{Q \times \mathcal{L}}. \tag{9}$$

The PEVD of (8) is

$$\boldsymbol{\mathcal{R}}_{\boldsymbol{\chi}\boldsymbol{\chi}}(z) \approx \boldsymbol{\mathcal{U}}_{\boldsymbol{\chi}}^P(z) \boldsymbol{\Lambda}_{\boldsymbol{\chi}}(z) \boldsymbol{\mathcal{U}}_{\boldsymbol{\chi}}(z) \in \mathbb{R}^{\mathcal{L} \times \mathcal{L}}, \tag{10}$$

and the processed output is

$$\mathbf{y}(z) = \boldsymbol{\mathcal{U}}_{\boldsymbol{\chi}}(z) \boldsymbol{\chi}(z). \tag{11}$$

Assuming stationarity, (8) is estimated in practice using

$$\mathbf{R}_{\boldsymbol{\chi}\boldsymbol{\chi}}(\tau) \approx \frac{1}{\lambda+1}\sum_{n=0}^{\lambda}\boldsymbol{\chi}(n)\boldsymbol{\chi}^T(n-\tau) \tag{12}$$

where λ is the frame size. The z-transform of (12) is

$$\boldsymbol{\mathcal{R}}_{\boldsymbol{\chi}\boldsymbol{\chi}}(z) \approx \sum_{\tau=-W}^{W}\mathbf{R}_{\boldsymbol{\chi}\boldsymbol{\chi}}(\tau)z^{-\tau}, \tag{13}$$

where W is the truncation window that captures the temporal correlations of the eigenbeam signals, outside which $\mathbf{R}_{\boldsymbol{\chi}\boldsymbol{\chi}}(\tau) \approx 0$.

We consider a single speech source and therefore the rank of the signal subspace in $\boldsymbol{\Lambda}_{\mathbf{x}}(z)$ or $\boldsymbol{\Lambda}_{\boldsymbol{\chi}}(z)$ is 1, thereby, suggesting an opportunity for compression. By construction, $\boldsymbol{\Psi}$ is rank deficient when $\mathcal{L} < Q$. For an order-limited sound field with Q being necessarily large to avoid spatial aliasing, the proper selection of the order, L, can sufficiently represent $\mathbf{R}_{\mathbf{x}\mathbf{x}}(\tau)$ of the microphone signals using a smaller dimension $\mathbf{R}_{\boldsymbol{\chi}\boldsymbol{\chi}}(\tau)$ of the eigenbeams. This effective compression allows the PEVD to be computed at a lower cost.

Algorithm 1 PEVD-based enhancement using eigenbeams.

Inputs: $\mathbf{x}(n) \in \mathbb{R}^Q, \alpha, L, \lambda, W, \delta, \mu, \iota$.
$\boldsymbol{\chi}(n) \leftarrow \mathbf{x}(n)$ *// SHT, see* (2)
$\mathbf{R}_{\boldsymbol{\chi}\boldsymbol{\chi}}(\tau) \leftarrow E\{\boldsymbol{\chi}(n)\boldsymbol{\chi}^T(n-\tau)\}$ *// see* (12)
$\boldsymbol{\mathcal{R}}_{\boldsymbol{\chi}\boldsymbol{\chi}}(z) \leftarrow \mathcal{Z}\{\mathbf{R}_{\boldsymbol{\chi}\boldsymbol{\chi}}(\tau)\}$ *// see* (13)
$\boldsymbol{\mathcal{U}}_{\boldsymbol{\chi}}(z), \boldsymbol{\Lambda}_{\boldsymbol{\chi}}(z) \leftarrow$ PEVD $\{\boldsymbol{\mathcal{R}}_{\boldsymbol{\chi}\boldsymbol{\chi}}(z), \delta, \mu, \iota\}$ *// use any [15–18]*
$\boldsymbol{\chi}(z) \leftarrow \mathcal{Z}\{\boldsymbol{\chi}(n)\}$ *// see* (13)
$\mathbf{y}(z) \leftarrow \boldsymbol{\mathcal{U}}_{\boldsymbol{\chi}}(z)\boldsymbol{\chi}(z)$ *// speech enhancement*
return $\mathbf{y}(z)$.

5. SIMULATION AND RESULTS

5.1. Experimental Setup

Anechoic speech signals sampled at 16 kHz are taken from the TIMIT corpus [23]. In Experiment 1, the SMIRgen tool in [24] is used to generate room impulse responses for a 32 microphone spherical array with radius 4.2 cm. The room has dimensions 5 m$\times$ 6 m $\times$ 4 m and $T_{60} = 0.3$ s. The positions of the source and the array centre in metres are $(3.37, 4.0, 1.7)$ and $(1.67, 4.0, 1.7)$. White Gaussian noise is used to simulate sensor noise. In Experiment 2, the Lecture Room 2 impulse response with $T_{60} = 1.22$ s and the babble noise signals, which are recorded using the 32-channel Eigenmike spherical array [25], are taken from the ACE corpus [26].

In each experiment, 50 trials were conducted. For each trial, sentences from a randomly selected speaker were concatenated so that each speech signal lasted for 8 to 10 s. The anechoic speech signal was convolved with the room impulse response before adding noise to obtain the microphone signals at a specified input SNR. The SNR was varied from -10 dB to 20 dB.

The parameters for the PEVD algorithms were $\mu = 10^{-3}$, $\iota = 500$, $\delta = \sqrt{N_1/3} \times 10^{-2}$, where N_1 is the square of the trace-norm of $\mathbf{R}_{\mathbf{x}\mathbf{x}}(0)$ and $\lambda = W = 1600$ samples. To study the effects of using SHT, two cases $L = 1$ (PEVD L1) and $L = 2$ (PEVD L2) were computed. Correspondingly, 4 and 9 eigenbeam signals were used as inputs to the PEVD algorithm. The quadrature weights $\boldsymbol{\alpha}$ were computed based on the non-uniform method in [27].

The proposed approach is compared against the PEVD-based method which uses the raw microphone signals (RAW PEVD) [13].

Table 1. Spherical harmonic order, L, number of eigenbeam signals $\mathcal{L}$, approximation error, $\varepsilon(\%)$ and PEVD computational complexity factor β relative to $Q = 32$, for a single speech example.

L	0	1	2	3	4
$\mathcal{L}$	1	4	9	16	25
$\varepsilon(\%)$	3.82	3.77	3.45	2.74	1.38
β	-	0.002	0.022	0.125	0.477

Eigenbeams $\chi_0^0(n)$ and $\chi_1^1(n)$ are also evaluated since the former can provide some noise reduction while the latter represents a dipole directed at the source for Experiment 1. To demonstrate the decorrelation ability of the PEVD, Karhunen-Loève transform (KLT) [28] is also applied to the eigenbeam $\chi_0^0(n)$ (KLT$\{\chi_0^0(n)\}$) since space and time decoupling is achieved by using SHT and KLT, respectively.

5.2. Evaluation Metrics

The evaluation uses measures of the segmental signal to noise ratio (SegSNR) and frequency-weighted SegSNR (FwSegSNR) for noise reduction [29], normalized signal-to-reverberant ratio (NSRR) and Bark spectral distortion (BSD) for dereverberation [30], short-time objective intelligibility (STOI) for speech intelligibility [31] and perceptual evaluation of speech quality (PESQ) [32] for speech quality. The metrics are computed for the microphone and enhanced signals and the improvement Δ is reported for each case. Positive Δ values indicate improvements in all measures, except ΔBSD, for which a negative value implies a reduction in spectral distortions.

The approximation loss, quantified as a percentage, is

$$\varepsilon = \frac{\sum_{q=1}^{Q}(\hat{x}(n,\mathbf{r}_q) - x(n,\mathbf{r}_q))^2}{\sum_{q=1}^{Q}(x(n,\mathbf{r}_q))^2} \times 100\%, \tag{14}$$

that is the total squared error between the received signals x and reconstructed microphone signals using the L-th order SH $\hat{x}$, normalized by the total energy of the received signals in Q microphones.

The PEVD complexity factor relative to $Q = 32$, which is estimated using $\beta = (\frac{\mathcal{L}}{Q})^3$, quantifies the computational savings.

5.3. Experiments and Discussion

For a single example received by the spherical microphone array in a SMIRgen simulated room and corrupted by 0 dB white noise, Table 1 shows the approximation error, ε, associated with different SH order L. As expected, ε decreases with increasing L. The relatively small values of ε obtained even using just the low-order eigenbeams suggests that the raw microphone signals may be highly redundant and may be compactly represented in the SH domain. β is significantly lower for a small number of eigenbeam signals $\mathcal{L}$, and is omitted for $\mathcal{L} = 1$ since PEVD is a multichannel algorithm.

Table 2 compares the speech enhancement performance of the algorithms for a single example from Experiment 1. In this setup, the PEVD-based algorithms perform comparably with PEVD L2 performing best in all metrics. PEVD L1 gives a bigger improvement in ΔFwSegSNR and ΔBSD compared to RAW PEVD, which uses all microphone signals directly. After PEVD L2, the $\chi_1^1(n)$ eigenbeam directed at the source gives the best ΔSTOI. Applying the KLT to $\chi_0^0(n)$ improves all measures except for ΔSTOI.

Speech enhancement performance for 50 Monte-Carlo trials in Experiment 1 is shown in Fig. 2. Across all measures and SNRs, the PEVD-based algorithms including RAW PEVD, PEVD L1 and PEVD L2 are ranked first or second after KLT$\{\chi_0^0(n)\}$, which is

Table 2. Speech enhancement performance for a single speech example in a SMIRgen simulated room, corrupted by 0 dB white noise.

Algorithm	ΔFwSegSNR	ΔSTOI	ΔPESQ	ΔBSD
$\chi_0^0(n)$	4.86 dB	0.055	0.42	-1.53 dB
KLT$\{\chi_0^0(n)\}$	5.56 dB	0.054	**0.51**	-1.65 dB
$\chi_1^1(n)$	0.89 dB	0.122	0.44	-0.65 dB
PEVD L1	5.72 dB	0.110	0.47	-1.68 dB
PEVD L2	**5.92 dB**	**0.125**	**0.51**	**-1.71 dB**
RAW PEVD	5.59 dB	0.119	0.49	-1.62 dB

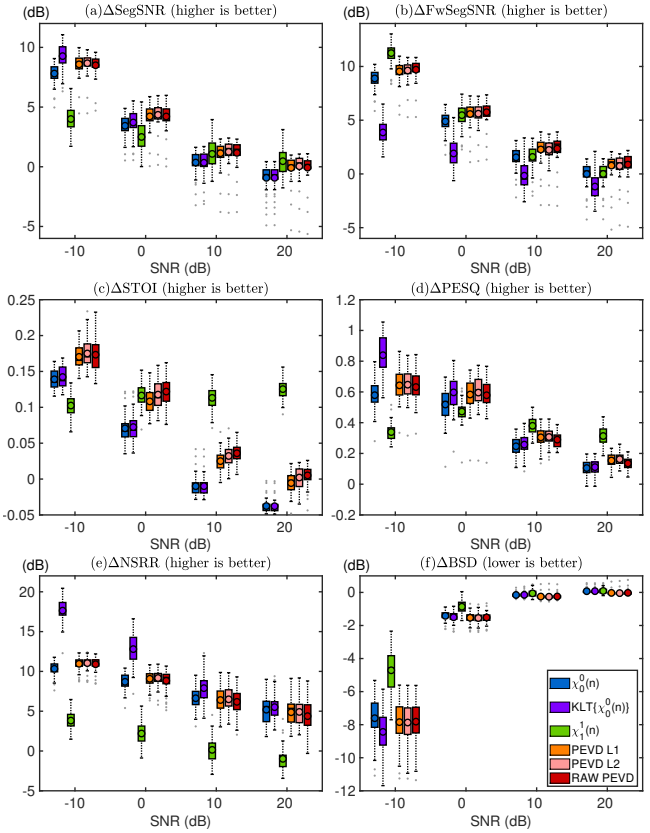

Fig. 2. Speech enhancement results for white noise in a SMIRgen simulated room with 0.3 s reverberation time.

optimal for white noise, or the $\chi_1^1(n)$ eigenbeam, which is pointing directly at the source. The PEVD-based algorithms perform comparably well, even though different numbers of channels are used for processing. This highlights that processing the eigenbeam signals instead of the raw microphone signals for speech enhancement can be effective and computationally advantageous.

When recorded signals from the ACE corpus are used, Fig. 3 shows that RAW PEVD provides the greatest improvement across all metrics for SNR$\geq$ 0 dB and is closely followed by PEVD L2 and PEVD L1. The $\chi_0^0(n)$ eigenbeam shows some improvement and the use of KLT does not offer further improvement and may even introduce processing artefacts when the noise is not white. This can be seen from a slight reduction in ΔSTOI and ΔPESQ in Fig. 3c and Fig. 3d for babble noise. The $\chi_1^1(n)$ eigenbeam is not expected to perform well because it is no longer pointing at the source directly and may only pick up weaker reverberant components along with noise.

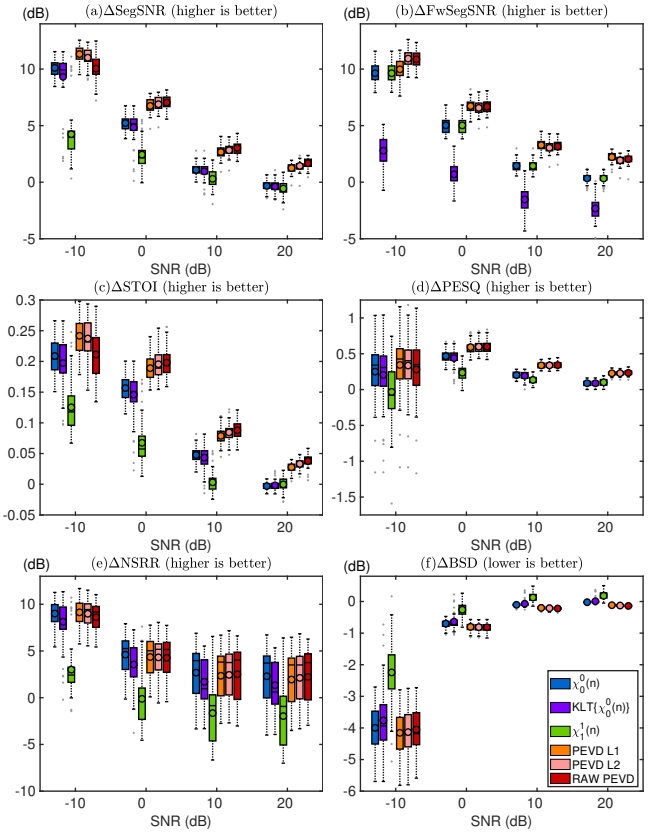

Fig. 3. Speech enhancement results for babble noise in ACE Lecture Room 2 with 1.22 s reverberation time.

At -10 dB SNR, the PEVD algorithms which use the eigenbeams, PEVD L1 and PEVD L2, perform better than RAW PEVD. In very noisy environments, the generated eigenbeams take advantage of the noise reduction offered by the signal independent SH domain processing but this is not available to RAW PEVD. At other SNRs, they perform comparably even though PEVD L1 and PEVD L2 use 4 and 9 channels respectively compared to the 32 channels used in the RAW PEVD approach.

6. CONCLUSION

The proposed approach, suitable for a spherical microphone array sampling an order-limited sound field, uses spherical harmonics to generate eigenbeam signals. Instead of processing the microphone signals directly, the PEVD algorithm is applied to the lower dimension eigenbeam signals. The proposed PEVD processing of the eigenbeams for speech enhancement has been shown to perform almost identically, and sometimes even better, than applying PEVD to all the raw microphone signals. However, importantly, the complexity factor of the proposed algorithm was shown to be 0.002 to 0.477 compared to PEVD processing of the raw microphone signals. Listening examples, available at [20], also indicate that our approach does not introduce any noticeable artefacts.

7. REFERENCES

[1] B. Rafaely, *Fundamentals of Spherical Array Processing*, ser. Springer Topics in Signal Processing. Springer-Verlag, 2015.

[2] 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.

[3] M. Park and B. Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array," *J. Acoust. Soc. Am.*, vol. 118, no. 5, pp. 3094–3103, Nov. 2005.

[4] E. Fernandez-Grande, "Sound field reconstruction using a spherical microphone array," *J. Acoust. Soc. Am.*, vol. 139, no. 3, pp. 1168–1178, Mar. 2016.

[5] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 81–91, 2015.

[6] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, "3D source localization in the spherical harmonic domain using a pseudointensity vector," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2010, pp. 442–446.

[7] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, "Eigenbeam-based acoustic source tracking in noisy reverberant environments," in *Proc. Asilomar Conf. on Signals, Syst. & Comput.*, Nov. 2010, pp. 576–580.

[8] D. P. Jarrett, E. A. P. Habets, J. Benesty, and P. A. Naylor, "A tradeoff beamformer for noise reduction in the spherical harmonic domain," in *Proc. Int. Workshop on Acoustic Signal Enhancement (IWAENC)*, Sep. 2012, pp. 1–4.

[9] D. P. Jarrett, E. A. P. Habets, M. R. P. Thomas, N. D. Gaubitch, and P. A. Naylor, "Dereverberation performance of rigid and open spherical microphone arrays: Theory & simulation," in *Proc. Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, 2011, pp. 145–150.

[10] Y. Peled and B. Rafaely, "Linearly-constrained minimum-variance method for spherical microphone arrays based on plane-wave decomposition of the sound field," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 12, pp. 2532–2540, Dec. 2013.

[11] S. Braun, D. P. Jarrett, J. Fischer, and E. A. P. Habets, "An informed spatial filter for dereverberation in the spherical harmonic domain," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, May 2013, pp. 669–673.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

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

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

[20] V. W. Neo, C. Evers, and P. A. Naylor, "PEVD using spherical harmonics for speech enhancement," Oct. 2020. [Online]. Available: https://vwn09.github.io/shd-pevd/

[21] B. Rafaely, "Analysis and design of spherical microphone arrays," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 135–143, Jan. 2005.

[22] A. Rao and R. Kumaresan, "On decomposing speech into modulated components," *IEEE Trans. on Speech and Audio Process.*, vol. 8, no. 3, pp. 240–254, May 2000.

[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, Corpus, 1993.

[24] 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, Sep. 2012.

[25] "EM32 Eigenmike microphone array release notes (v17.0)," M. H. Acoust., NJ USA, Hardware, Oct. 2013. [Online]. Available: http://www.mhacoustics.com/sites/default/files/ReleaseNotes.pdf

[26] J. Eaton, N. D. Gaubitch, A. H. Moore, and P. A. Naylor, "Estimation of room acoustic parameters: The ACE challenge," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 24, no. 10, pp. 1681–1693, Oct. 2016.

[27] S. Brown and D. Sen, "Error analysis of spherical harmonic soundfield representations in terms of truncation and aliasing errors," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2013, pp. 360–364.

[28] Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech Audio Process.*, vol. 3, no. 4, pp. 251–266, Jul. 1995.

[29] 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.

[30] P. A. Naylor, N. D. Gaubitch, and E. A. P. Habets, "Signal-based performance evaluation of dereverberation algorithms," *J. of Elect. and Comput. Eng.*, vol. 2010, pp. 1–5, 2010.

[31] 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.

[32] "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.