Speech Dereverberation Performance of a Polynomial-EVD Subspace Approach

Vincent W. Neo
Dept. of Elect. and Electron. Eng.
Imperial College London, UK
vincent.neo09@imperial.ac.uk

Christine Evers[†]
Dept. of Electron. and Comput. Sci.
University of Southampton, UK
c.evers@soton.ac.uk

Patrick A. Naylor[*]
Dept. of Elect. and Electron. Eng.
Imperial College London, UK
p.naylor@imperial.ac.uk

Abstract—The degradation of speech arising from additive background noise and reverberation affects the performance of important speech applications such as telecommunications, hearing aids, voice-controlled systems and robot audition. In this work, we focus on dereverberation. It is shown that the parameterized polynomial matrix eigenvalue decomposition (PEVD)-based speech enhancement algorithm exploits the lack of correlation between speech and the late reflections to enhance the speech component associated with the direct path and early reflections. The algorithm's performance is evaluated using simulations involving measured acoustic impulse responses and noise from the ACE corpus. The simulations and informal listening examples have indicated that the PEVD-based algorithm performs dereverberation over a range of SNRs without introducing any noticeable processing artefacts.

Index Terms—Dereverberation, polynomial matrix eigenvalue decomposition, convolutive noise, broadband signal processing, microphone array.

I. Introduction

The processing of degraded speech remains challenging for many applications such as telecommunications, hearing aids, voice-controlled systems and robot audition [1], [2]. The main causes of degradation are additive background noise and reverberation arising from multi-path reflections in an enclosed space [3]. Even without background noise, reverberation results in temporal smearing of the anechoic speech signal. This structural corruption of the original speech signal can degrade speech intelligibility in real-world situations.

To reduce the impact of reverberation and improve the performance in these applications, several dereverberation approaches have been proposed [4]. These can be classified into acoustic channel equalization, homomorphic transformation-based and speech enhancement methods including those based on linear prediction coding (LPC). In the channel equalization approaches, the acoustic channel is first blindly estimated using system identification [5]–[8]. This is followed by the design of inverse filter(s), such as [9] for a single-channel system and [10] for a multi-channel system based on the multi-channel inverse theorem (MINT), or matched filter(s) in [11], which are then used for equalization.

In homomorphic methods, the reverberant signal is transformed into the cepstrum domain where the anechoic speech and reverberant components can be separated by liftering [12], [13]. Performance of this approach has been shown to be limited [14].

Speech enhancement approaches include synthesis-based methods that use the LPC residual to generate the estimated clean speech signal [15], [16]. These methods are usually limited to mildly reverberant signals in order to avoid introducing artefacts [14]. Single-channel filtering-based enhancement methods for dereverberation include spectral subtraction [17] and harmonic filtering [18]. When multiple microphones are available, array processing techniques [3], [19], the generalised weighted prediction error (GWPE) dereverberation filter [20], the multi-channel Wiener filter (MWF) [21], [22] and the multi-channel Karhunen-Loève transform (KLT) [23], [24], may be used in conjunction with post-filtering enhancement techniques [25].

In [26], [27], a polynomial matrix eigenvalue decomposition (PEVD)-based speech enhancement algorithm has been shown capable of achieving noise reduction without introducing artefacts into the enhanced speech signal. This approach uses a polynomial covariance matrix formulation that can simultaneously model the spatial, temporal and spectral correlations of multi-channel signals. PEVD-based algorithms [28]–[31] then perform the enhancement.

In [27], we demonstrated the performance of the PEVD approach focused on noise reduction in a speech enhancement task. In this work, the focus is now on dereverberation. The PEVD-based algorithm [27] is presented for dereverberation and evaluated through simulations using the full set of room impulse responses in the ACE corpus [32], under both noiseless and noisy conditions. The noiseless condition, which was not investigated in [27], provides interesting insights into the dereverberation analysis.

II. Signal Model and Problem Formulation

In [27], the noisy and reverberant signal, $x_m(n)$, at the m-th microphone for discrete-time sample $n = 0, 1, \ldots, N-1$, is

$$x_m(n) = \mathbf{h}_m^T \mathbf{s}_0(n) + v_m(n) \tag{1}$$
$$= \tilde{\mathbf{h}}_{m,dp}^T \mathbf{s}_0(n) + \tilde{\mathbf{h}}_{m,er}^T \mathbf{s}_0(n) + \tilde{\mathbf{h}}_{m,lr}^T \mathbf{s}_0(n) + v_m(n)$$
$$= \tilde{s}_m(n) + \tilde{v}_m(n), \quad m = 1, 2, \ldots, M, \tag{2}$$

where $\mathbf{h}_m = [h_{m,0}, h_{m,1}, \ldots, h_{m,J}]^T$ is the m-th acoustic channel, which is modelled as a J-th order finite impulse response filter and decomposed into the direct path, $\tilde{\mathbf{h}}_{m,dp}$, early reflections, $\tilde{\mathbf{h}}_{m,er}$, and the late reflections, $\tilde{\mathbf{h}}_{m,lr}$ [3], $\mathbf{s}_0(n) = [s_0(n), s_0(n-1), \ldots, s_0(n-J)]^T$ is the anechoic speech signal, $v_m(n)$ is additive noise and $[\cdot]^T$ denotes the transpose operator. The noise signals are assumed to be zero-mean, not perfectly coherent with each other and uncorrelated with the source signal [23]. Furthermore, exploiting the lack of correlation between the late reflections and anechoic speech signal gives $\tilde{s}_m(n) = \tilde{\mathbf{h}}_{m,dp}^T \mathbf{s}_0(n) + \tilde{\mathbf{h}}_{m,er}^T \mathbf{s}_0(n)$ and $\tilde{v}_m(n) = \tilde{\mathbf{h}}_{m,lr}^T \mathbf{s}_0(n) + v_m(n)$, which are the speech and noise components respectively.

The array output from M microphones can be written as $\mathbf{x}(n) = [x_1(n), \ldots, x_M(n)]^T$, with $\tilde{\mathbf{s}}(n)$, $\mathbf{v}(n)$ and $\tilde{\mathbf{v}}(n)$ similarly defined. In [27], the goal is to retrieve $\tilde{\mathbf{s}}(n)$ from $\mathbf{x}(n)$. Here, the goal is to recover the anechoic speech, $\mathbf{s}_0(n)$, possibly including additive noise, from $\mathbf{x}(n)$.

III. PEVD-BASED SPEECH ENHANCEMENT

A. Formulation of Polynomial Matrices

Broadband speech signals received by microphone arrays exhibit space, time and frequency correlations. Because the different frequency components in speech are affected by different phase shifts and each phase shift requires specific temporal alignment, the correlations across different sensors and at different time lags need to be considered. This is computed using the space-time covariance matrix [27]

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

where the $(p,q)^{\text{th}}$ element, $r_{pq}(\tau) = \mathbb{E}\{x_p(n)x_q(n-\tau)\}$, is the cross-correlation between sensor p and q. Concatenating the covariance matrix, $\mathbf{R}_{\mathbf{xx}}(\tau)$, for all choices of $\tau \in \{-N+1, \ldots, N-1\}$, results in a tensor of dimension $M \times M \times (2N-1)$. Furthermore, because speech signals are typically processed in the short-time Fourier transform (STFT) domain, in order to capture the spectral correlations, the covariance needs to be further expanded to a $M \times M \times (2N-1) \times K$ tensor, where K is the number of frequency bins in the STFT.

A more compact representation of the speech signals, which captures the space, time and frequency correlations, can be obtained by using the z-transform rather than the STFT. The z-transform of (3) is a para-Hermitian polynomial matrix [28]

$$\mathcal{R}_{\mathbf{xx}}(z) = \sum_{\tau=-\infty}^{\infty} \mathbf{R}_{\mathbf{xx}}(\tau)z^{-\tau}. \tag{4}$$

This polynomial matrix is a matrix with polynomial elements or, equivalently, a polynomial with matrix coefficients.

B. Polynomial Matrix Eigenvalue Decomposition

The PEVD of a para-Hermitian polynomial matrix [28] is

$$\mathcal{R}_{\mathbf{xx}}(z) \approx \mathcal{U}^P(z)\Lambda(z)\mathcal{U}(z), \tag{5}$$

where the rows of $\mathcal{U}(z)$ are the eigenvectors and the diagonal elements of $\Lambda(z)$ are the eigenvalues, and $[\cdot]^P$ is the para-Hermitian operator such that $\mathcal{U}_\ell^P(z) = \mathcal{U}_\ell^H(z^{-1})$. The PEVD

can be computed using an iterative algorithm [28]–[31] based on similarity transforms involving L polynomial matrices, $\mathcal{U}(z) = \mathcal{U}_L(z) \ldots \mathcal{U}_1(z)$.

At the ℓ-th iteration, the PEVD algorithm first searches for the largest off-diagonal element if it exceeds a predefined threshold, δ. $\mathcal{U}_\ell(z)$ is then constructed using delay polynomial matrices and unitary matrices, which are designed to zero out the off-diagonal elements on the plane of z^0, and applied to the entire polynomial matrix. To keep the polynomial order compact, a fraction of the total Frobenius-norm squared, μ, is truncated as detailed in [28]. After L iterations, $\mathcal{R}_{\mathbf{xx}}(z)$ is approximately diagonalized according to

$$\Lambda(z) \approx \mathcal{U}(z)\mathcal{R}_{\mathbf{xx}}(z)\mathcal{U}^P(z) = \mathcal{U}(z)\mathbb{E}\{\mathbf{x}(z)\mathbf{x}^P(z)\}\mathcal{U}^P(z),$$

where $\mathbf{x}(z)$ is the z-transform of $\mathbf{x}(n)$ based on (4). The zeroing unitary matrix computed at each iteration can take the form of a Givens rotation in second-order sequential best rotation (SBR2) [28], a Householder-like optimisation procedure as in [29], a combination of Householder reflection and Givens rotation matrices in [31] or an eigenvector matrix in the sequential matrix diagonalization (SMD) algorithm [30].

C. PEVD-based Algorithm for Dereverberation

Using (3) and (4), the z-transform of the space-time covariance matrix for the microphone array output is

$$\mathcal{R}_{\mathbf{xx}}(z) = \mathcal{R}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}(z) + \mathcal{R}_{\tilde{\mathbf{v}}\tilde{\mathbf{v}}}(z), \tag{6}$$

where the z-transform of the speech and noise space-time covariance matrices are $\mathcal{R}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}(z)$ and $\mathcal{R}_{\tilde{\mathbf{v}}\tilde{\mathbf{v}}}(z)$ respectively. Assuming stationarity within each processing frame, in practice, (3) is usually estimated using N samples per frame based on

$$\hat{\mathbf{R}}_{\mathbf{xx}}(\tau) \approx \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{x}(n)\mathbf{x}^T(n-\tau). \tag{7}$$

Furthermore, (4) can be approximated using

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

where W is the truncation window which reflects the extent of temporal correlation of the speech signals. Hence, both N and W are parameters for the proposed algorithm. Consequently, the PEVD of (6) gives

$$\mathcal{R}_{\mathbf{xx}}(z) \approx \left[\mathcal{U}_{\tilde{s}}^P(z) \mid \mathcal{U}_{\tilde{v}}^P(z)\right] \left[\begin{array}{c|c} \Lambda_{\tilde{s}}(z) & 0 \\ \hline 0 & \Lambda_{\tilde{v}}(z) \end{array}\right] \left[\begin{array}{c} \mathcal{U}_{\tilde{s}}(z) \\ \hline \mathcal{U}_{\tilde{v}}(z) \end{array}\right],$$

where $\{.\}_{\tilde{s}}$ and $\{.\}_{\tilde{v}}$ represent the orthogonal signal and noise subspace components. The speech subspace comprises anechoic speech convolved with the direct path and early reflections while the noise subspace contains ambient noise and late reflections associated with the reverberant channel.

The eigenvector polynomial matrix, $\mathcal{U}(z)$, can also be interpreted as a filterbank for $\mathbf{x}(z)$ so $\mathbf{y}(z) = \mathcal{U}(z)\mathbf{x}(z)$ performs decorrelation such that [28]

$$\mathbb{E}\{\mathbf{y}(z)\mathbf{y}^P(z)\} = \mathbb{E}\{\mathcal{U}(z)\mathbf{x}(z)\mathbf{x}^P(z)\mathcal{U}^P(z)\} \approx \Lambda(z). \tag{9}$$

Because PEVD algorithms sort $\mathbf{\Lambda}(z)$ in descending order, dereverberation is achieved by combining components in the signal subspace and nulling components in the noise subspace. The PEVD-based algorithm is summarized in Algorithm 1.

Algorithm 1 PEVD-based algorithm [27] for dereverberation.

Inputs: $\mathbf{x}(n) \in \mathbb{R}^M, N, W, \delta, \mu, L.$
$\quad \tilde{\mathbf{R}}_{\mathbf{xx}}(\tau) \leftarrow E\{\mathbf{x}(n)\mathbf{x}^T(n-\tau)\}$ // see (7)
$\quad \tilde{\mathcal{R}}_{\mathbf{xx}}(z) \leftarrow \mathcal{Z}\{\mathbf{R}_{\mathbf{xx}}(\tau)\}$ // see (8)
$\quad \mathcal{U}(z), \mathbf{\Lambda}(z) \leftarrow \text{PEVD}\{\tilde{\mathcal{R}}_{\mathbf{xx}}(z), \delta, \mu, L\}$ // use any PEVD
$\quad algorithm$ [28]–[31]
$\quad \mathbf{x}(z) \leftarrow \mathbf{x}(n)$ // see (4)
$\quad \mathbf{y}(z) \leftarrow \mathcal{U}(z)\mathbf{x}(z)$ // speech dereverberation
$\quad$ **return** $\mathbf{y}(z)$.

IV. SIMULATIONS AND RESULTS

A. Experiment Setup

To evaluate the proposed approach, anechoic speech signals, sampled at 16 kHz, from the TIMIT corpus [33] are used with the noise and room impulse response measurements from the 3-channel mobile recordings in the ACE corpus [32].

To highlight the dereverberation performance, an experiment without background noise for all rooms is conducted for 200 speech signals. For the simulations involving background noise, 50 trials are conducted. In each trial, sentences from a randomly selected speaker are concatenated to a signal length between 8 to 10 s. The anechoic speech signal is convolved with the impulse response at each microphone channel before being corrupted by additive noise using [34]. The noise conditions include fan and babble noise at -5 dB, 0 dB, 5 dB and 20 dB signal to noise ratio (SNR).

The PEVD parameters, chosen following [26]–[28], are $\delta = \sqrt{N_1/3} \times 10^{-2}$ where N_1 is the square of the trace-norm of $\mathbf{R}_{\mathbf{xx}}(0)$, $\mu = 10^{-3}$ and $L = 500$. In the experiments, $N = W = 1600$ samples are used. With this parameter selection, correlations to within 100 ms, which are assumed to include the direct-path and early reflections components, are captured and used by the algorithm.

The proposed PEVD method is compared against the multi-channel subspace (MCSUB) method [23] and the Oracle MWF (OMWF), which uses complete prior knowledge of the clean dry speech and is based on the concatenation of a minimum variance distortionless response (MVDR) followed by a single-channel Wiener filter [25]. We also compared with the diagonal matrix method in GWPE using the published parameters [20].

B. Performance Evaluation Metrics

The most direct objective measure of dereverberation is direct-to-reverberant ratio (DRR) but the modified impulse responses after dereverberation cannot be precisely determined. Instead, the normalised signal to reverberant ratio (NSRR), which is a signal-based measure and is shown to be equivalent to DRR under certain conditions [35], and the Bark spectral distortion (BSD) are used. Frequency-weighted segmental SNR (FwSegSNR) [36] is also included to measure deviation from

Algorithm	ΔNSRR	ΔBSD	ΔFwSegSNR	ΔPESQ
GWPE	0.68 dB	-0.25 dB	1.46 dB	0.70
MCSUB	-3.20 dB	0.28 dB	1.47 dB	0.01
OMWF	0.10 dB	0.04 dB	1.46 dB	0.16
PEVD	1.01 dB	-0.10 dB	1.47 dB	0.11
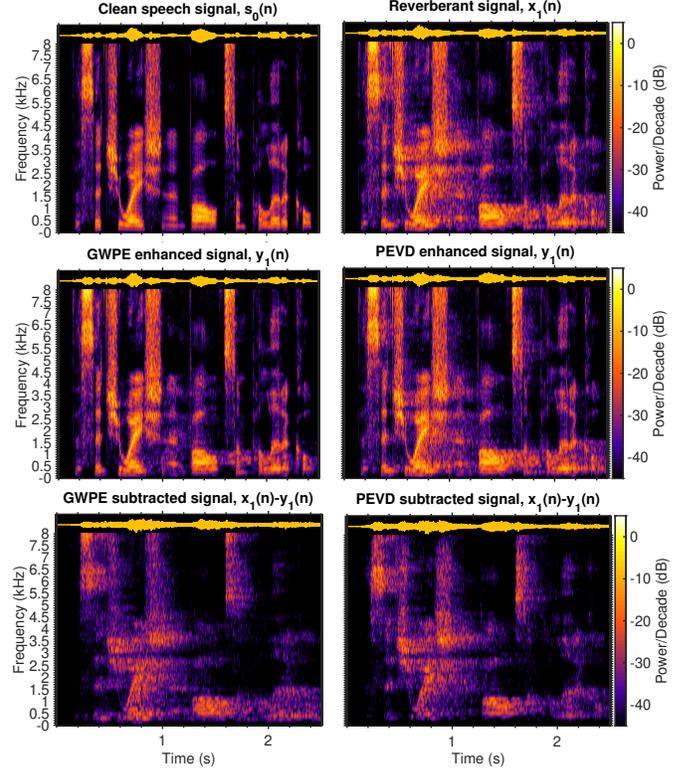
Fig. 1. Zoom-in spectrograms of clean speech, noisy and processed signals and residue using GWPE and PEVD for no noise example in Lecture Room 2.

the dry clean speech, especially under the noiseless condition. Perceptual evaluation of speech quality (PESQ) [37] is also used to account for artefacts introduced by the processing.

The metrics are computed before and after processing using the proposed approach and benchmark algorithms, and the differences are denoted using Δ. The algorithms demonstrate an improvement when (i) ΔNSRR is positive which indicates the strengthening of the direct-path relative to the reverberant component, (ii) ΔBSD is negative which indicates reduction in spectral distortions, (iii) ΔFwSegSNR is positive which indicates noise reduction or reverberation suppression and (iv) ΔPESQ is positive which shows better speech quality.

C. Results and Discussions

Fig. 1 shows the results for a single reverberant speech example in the absence of background noise in Lecture Room 2. The spectrograms in Fig. 1 indicate that both GWPE and PEVD are able to suppress reverberation while retaining the overall speech structure. However, GWPE seems to have applied a more aggressive suppression which results in a cleaner spectrogram plot. This is supported by Table I where

GWPE indicates lower spectral distortion with ΔBSD of -0.25 dB compared to -0.10 dB using PEVD. To understand the effects of the algorithms, the difference between the reverberant signal and the processed signal is computed and the spectrogram of the subtracted signal is also plotted. Fig. 1 shows that the GWPE heavily suppresses most of the earlier but not the later components. PEVD, on the other hand, provided a gentler dereverberation overall as indicated by the greater smearing in the subtracted signal and larger ΔNSRR. Furthermore, listening examples, provided on https://www.commsp.ee.ic.ac.uk/%7esap/pevddrb/, have indicated similar observations as the spectrograms.

Fig. 2 shows that, without noise, PEVD gives the greatest improvement in ΔNSRR in all rooms. This is followed by GWPE. In terms of ΔBSD and ΔFwSegSNR, GWPE has the best improvement followed by PEVD. Both OMWF and MCSUB, on the other hand, degrade NSRR, BSD and usually FwSegSNR as well. For ΔPESQ, GWPE offers the greatest improvement, followed by PEVD and OMWF with comparable performance and MCSUB which degrades the score.

With background babble noise or fan noise as shown in Fig. 3 and Fig. 4 respectively, PEVD outperforms all algorithms in terms of ΔNSRR over the entire range of SNRs in both rooms. This is followed by GWPE which offers a slight gain, OMWF and MCSUB that usually degrades NSRR.

In terms of ΔBSD, PEVD performs better than the GWPE at -5 dB SNR babble and fan noise and they have comparable performance at the other SNRs. This is different from the noiseless case when GWPE has the best performance. This suggests some superior noise robustness for PEVD. OMWF offers some improvement at -5 dB SNR but does not improve ΔBSD at higher SNRs. MCSUB makes the ΔBSD worse in the presence of background noise in both rooms.

In terms of ΔFwSegSNR, PEVD outperforms all the other algorithms in both noise types and rooms except for OMWF when there is babble noise in Lecture Room 2 as shown in Fig. 3. This is because OMWF uses knowledge of the clean speech signal whereas PEVD is a completely blind method. MCSUB offers improvement in ΔFwSegSNR in most cases except at -5 dB babble noise in Office 1. On the other hand, the GWPE makes ΔFwSegSNR worse in most cases except when the babble noise is at -5 dB SNR.

In terms of ΔPESQ, all algorithms perform competitively with an average improvement of 0.20 when the babble noise in Fig. 3 and fan noise in Fig. 4 are at -5 dB and 5 dB SNRs. At 20 dB SNR, GWPE offers the greatest improvement by up to 0.5 in PESQ scores.

The summary results for comprehensive testing over the complete ACE corpus are given in Table II and, for the example of 0 dB babble noise, indicate that PEVD provides the best dereverberation performance.

V. CONCLUSION

The PEVD-based algorithm designed for dereverberation is introduced. Simulation results have indicated that the PEVD-based approach is effective for dereverberation. PEVD out-
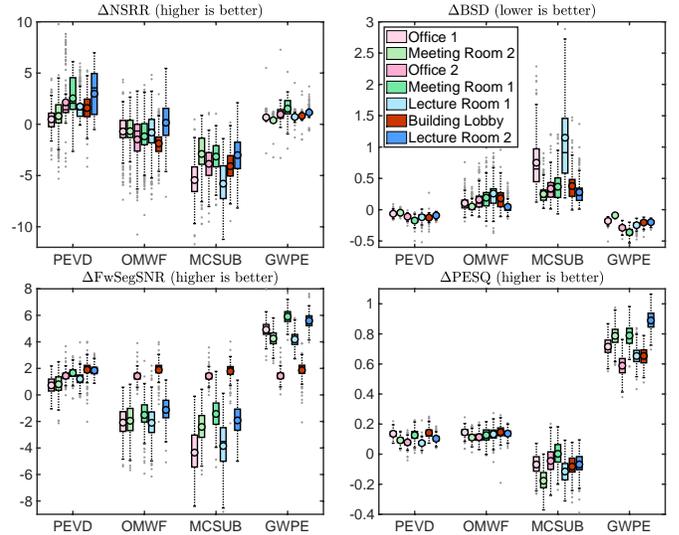
Fig. 2. Box plot for no noise simulation in all ACE corpus rooms.
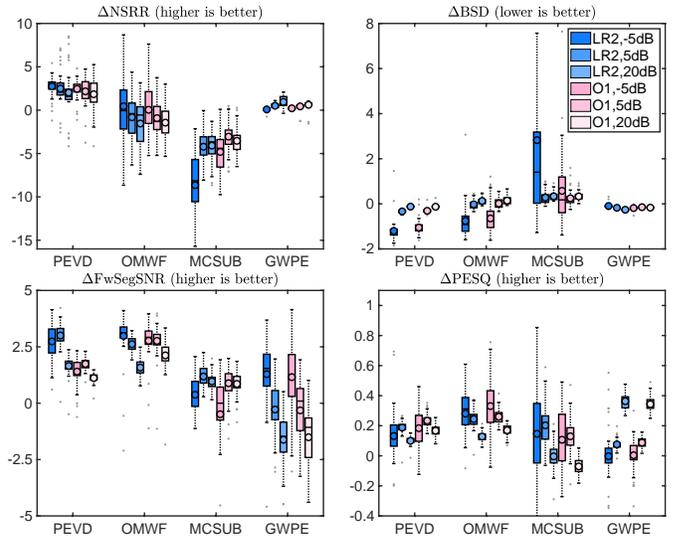
Fig. 3. Box plot for babble noise simulation in Lecture Room 2 and Office 1.
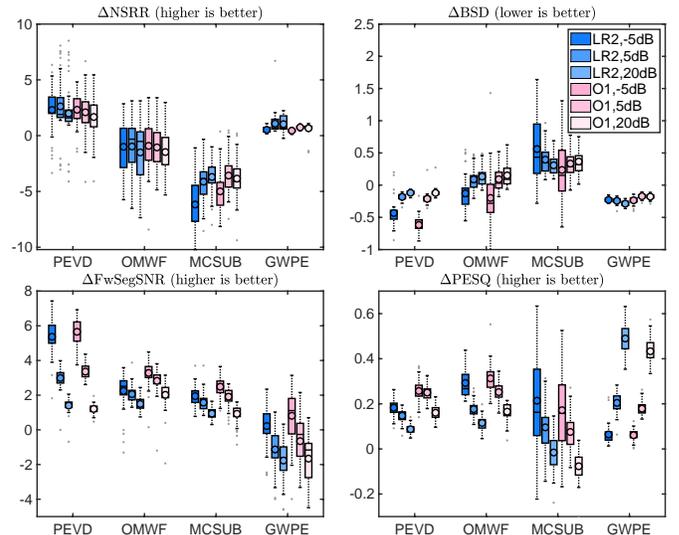
Fig. 4. Box plot for fan noise simulation in Lecture Room 2 and Office 1.

TABLE II
DEREVERBERATION PERFORMANCE, MEASURED USING MEAN (ΔNSRR, ΔBSD), EVALUATED ON THE ACE CORPUS FOR 0 dB BABBLE NOISE.

Room T_{60}	Office 1 0.332 s	Meeting Room 2 0.371 s	Office 2 0.390 s	Meeting Room 1 0.437 s	Lecture Room 1 0.638 s	Building Lobby 0.646 s	Lecture Room 2 1.22 s
GWPE	0.31, -0.15	0.71, -0.21	0.35, -0.14	0.53, -0.19	0.2, -0.07	0.26, -0.12	0.33, -0.15
MCSUB	-3.09, 0.15	-6.01, 1.30	-2.04, 0.05	-2.65, 0.14	-3.28, 0.10	-4.27, 0.57	-4.72, 0.28
OMWF	-0.51, -0.14	-0.41, 0.11	1.53, -0.39	0.51, -0.13	0.32, -0.26	0.88, -0.22	-0.1, -0.25
PEVD	**2.37, -0.56**	**1.68, -0.25**	**3.94, -0.68**	**2.85, -0.73**	**2.41, -0.48**	**2.02, -0.54**	**2.85, -0.65**

performs other algorithms in ΔNSRR and has comparable performance with GWPE in ΔBSD over a wide range of SNRs. The analysis on the noiseless reverberant example and informal listening examples including noisy cases have also indicated that the proposed PEVD algorithm provides consistent suppression of reverberation without introducing any noticeable artefacts.

REFERENCES

[1] C. Evers, Y. Dorfan, S. Gannot, and P. A. Naylor, "Source tracking using moving microphone arrays for robot audition," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Mar. 2017.

[2] C. Evers and P. A. Naylor, "Acoustic SLAM," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 9, pp. 1484–1498, Sep. 2018.

[3] P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation*. Springer-Verlag, 2010.

[4] E. A. P. Habets and P. A. Naylor, "Dereverberation," in *Audio Source Separation and Speech Enhancement*, E. Vincent, T. Virtanen, and S. Gannot, Eds. John Wiley & Sons, Inc., 2018, pp. 317–343.

[5] Y. Huang and J. Benesty, "Adaptive multi-channel least mean square and Newton algorithms for blind channel identification," *Signal Process.*, vol. 82, pp. 1127–1138, Aug. 2002.

[6] ——, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 11–24, Jan. 2003.

[7] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951–1968, 1998.

[8] A. W. Khong, P. A. Naylor, and J. Benesty, "A low delay and fast converging improved proportionate algorithm for sparse system identification," *EURASIP J. on Audio, Speech, and Music Process.*, vol. 2007, no. 1, p. 084376, Apr. 2007.

[9] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *J. Acoust. Soc. Am.*, vol. 66, no. 1, pp. 165–169, Jul. 1979.

[10] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 2, pp. 145–152, Feb. 1988.

[11] J. L. Flanagan, A. C. Surendran, and E. E. Jan, "Spatially selective sound capture for speech and audio processing," *Speech Commun.*, vol. 13, no. 1-2, pp. 207–222, Oct. 1993.

[12] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Prentice Hall, 1975.

[13] D. Bees, M. Blostein, and P. Kabal, "Reverberant speech enhancement using cepstral processing," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, vol. 2, 1991, pp. 977–980.

[14] Y. Huang, J. Benesty, and J. Chen, "Dereverberation," in *Springer Handbook of Speech Processing*. Springer-Verlag, 2008, pp. 929–943.

[15] J. B. Allen, "Synthesis of pure speech from a reverberant signal," U.S. Patent 3 786 188, 1974.

[16] S. M. Griebel and M. S. Brandstein, "Wavelet transform extrema clustering for multi-channel speech dereverberation," in *Proc. Int. on Workshop Acoust. Echo and Noise Control (IWAENC)*, Pocono Manor, Pennsylvania, Sep. 1999, pp. 52–55.

[17] K. Lebart, J. M. Boucher, and P. N. Denbigh, "A new method based on spectral subtraction for speech de-reverberation," *Acta Acoustica*, vol. 87, pp. 359–366, 2001.

[18] T. Nakatani, M. Miyoshi, and K. Kinoshita, "Single-microphone blind dereverberation," in *Speech Enhancement*, J. Benesty, S. Makino, and J. Chen, Eds. Springer-Verlag, Apr. 2005.

[19] H. L. Van Trees, *Optimal Array Processing. Part IV of Detection, Estimation, and Modulation Theory*, 2002.

[20] T. Yoshioka and T. Nakatani, "Generalization of multi-channel linear prediction methods for blind MIMO impulse response shortening," *IEEE Trans. Audio, Speech, Language Process.*, vol. 20, no. 10, pp. 2707–2720, Dec. 2012.

[21] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Signal Process.*, vol. 50, no. 9, pp. 2230–2244, Sep. 2002.

[22] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 4, pp. 1218–1234, Jul. 2006.

[23] 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.

[24] F. Jabloun and B. Champagne, "A multi-microphone signal subspace approach for speech enhancement," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2001, pp. 205–208.

[25] K. U. Simmer, J. Bitzer, and C. Marro, "Post-filtering techniques," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. S. Brandstein and D. B. Ward, Eds. Berlin, Germany: Springer-Verlag, 2001, pp. 39–60.

[26] 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.

[27] ——, "PEVD-based speech enhancement in reverberant environments," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2020.

[28] 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.

[29] 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)*, Bilbao, Spain, 2011, pp. 421–425.

[30] ——, "Sequential matrix diagonalisation algorithms for polynomial EVD of para-Hermitian matrices," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 81–89, Jan. 2015.

[31] 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)*, May 2019.

[32] 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.

[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, Corpus, 1993.

[34] "Objective measurement of active speech level," Int. Telecommun. Union (ITU-T), Recommendation, Mar. 1993.

[35] 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.

[36] 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.

[37] "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," Int. Telecommun. Union (ITU-T), Recommendation P.862, Nov. 2003.