PEVD-BASED SPEECH ENHANCEMENT IN REVERBERANT ENVIRONMENTS

Vincent W. Neo, Christine Evers and Patrick A. Naylor

Department of Electrical and Electronic Engineering, Imperial College London, UK
Email: {vincent.neo09, c.evers, p.naylor}@imperial.ac.uk

ABSTRACT

The enhancement of noisy speech is important for applications involving human-to-human interactions, such as telecommunications and hearing aids, as well as human-to-machine interactions, such as voice-controlled systems and robot audition. In this work, we focus on reverberant environments. It is shown that, by exploiting the lack of correlation between speech and the late reflections, further noise reduction can be achieved. This is verified using simulations involving actual acoustic impulse responses and noise from the ACE corpus. The simulations show that even without using a noise estimator, our proposed method simultaneously achieves noise reduction, and enhancement of speech quality and intelligibility, in reverberant environments over a wide range of SNRs. Furthermore, informal listening examples highlight that our approach does not introduce any significant processing artefacts such as musical noise.

***Index Terms*—** Speech enhancement, polynomial matrix eigenvalue decomposition, microphone array, noise reduction, broadband signal processing.

1. INTRODUCTION

The degradation of speech by noise is challenging for many applications, such as telecommunications, hearing aids, voice-controlled systems and robot audition. To improve the performance of these systems, both single- and multi-channel speech enhancement methods have been proposed to reduce noise. However, it is well-known that such methods typically also distort the speech signal and introduce processing artefacts [1].

Single-channel enhancement approaches include spectral subtraction [2], statistical-based and subspace methods. In statistical methods, the enhancement filter is typically designed based on the mimimization of the mean square error (MSE) between the clean and estimated speech spectrum [3] or the log-spectrum (log-MMSE) in [4]. Because of the intrinsic coupling between noise reduction and speech distortion in a single-channel system, a parameter can be introduced in the derivation of the optimal Wiener filter to control that trade off [5, 6].

Subspace-based speech enhancement [7] used the Karhunen-Loève transform (KLT) and an optimal solution was derived when the noise is white. The method was extended in [8] to cope with coloured noise by approximating the noise covariance matrix in the KLT domain with a diagonal matrix. This was shown in [9] to be sub-optimal and a generalized eigenvalue decomposition (GEVD) was used to jointly diagonalize the speech signal and noise covariance matrices for coloured noise.

The subspace-based method has also been extended to multi-channel systems. In [10], the temporal signals at different microphones are stacked into a vector before computing the covariance matrix for KLT. Like the multi-channel Wiener filter (MWF), this approach does not fully exploit spatial information to minimize speech distortion [11]. A different approach was adopted in [12, 13], in which KLT is applied to the spatial covariance matrix between different microphones for different frequency bins. This approach, however, processes frequency subbands independently and, therefore, neglects correlations between bands and phase continuities across band boundaries.

Polynomial matrices are capable of capturing the space, time and frequency correlations simultaneously. Polynomial eigenvalue decomposition (PEVD) has been applied to many broadband signal proccessing applications such as blind source separation [14], source identification [15] and adaptive beamforming [16]. Our recent contribution in [17] extended the PEVD to the field of audio signal processing for the enhancement of speech distorted by additive noise. The approach in [17] focuses on direct-path propagation only, but does not incorporate multi-path effects. Direct-path propagation is relevant for telephony applications where the talker is close to the microphones, also considered in [6, 18, 19]. However, with technologies such as voice-controlled systems and robot audition [20] becoming increasingly important and common, multi-path reverberation must also be considered [21].

In this work, we propose a novel PEVD-based speech enhancement method incorporating a reverberant channel model. It is shown that the algorithm is able to achieve speech enhancement for a single-source while suppressing diffused noise without relying on a noise estimator. The robustness of the new speech enhancement method against other baseline approaches is demonstrated through simulations using room impulse responses and noise signals measured in real acoustic environment. Informal listening examples highlight that our method does not introduce any significant artefacts.

2. SIGNAL MODEL AND PROBLEM FORMULATION

The noisy and reverberant signal, $x_m(n)$, at the m-th microphone for discrete-time sample $n = 0, 1, \ldots, T-1$, is

$$x_m(n) = \mathbf{h}_m^T \mathbf{s}_0(n) + v_m(n) \tag{1}$$

$$= s_m(n) + v_m(n), \quad m = 1, 2, \ldots, M, \tag{2}$$

where $\mathbf{s}_0(n) = [s_0(n), s_0(n-1), \ldots, s_0(n-J)]^T$ is the anechoic speech signal, $\mathbf{h}_m = [h_{m,0}, h_{m,1}, \ldots, h_{m,J}]^T$ represents the m-th channel as a J-th order finite impulse response filter, $s_m(n)$ is the reverberant speech, $v_m(n)$ is the additive noise signal 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 [11]. The channel is also assumed to be time

The research leading to these results has received funding from the UK EPSRC Fellowship grant no. EP/P001017/1.

invariant. The received data vector for M microphones is written as $\mathbf{x}(n) = [x_1(n), \ldots, x_M(n)]^T$ with $\mathbf{v}(n)$ and $\mathbf{s}(n)$, similarly defined.

The goal is to estimate $\mathbf{s}_0(n)$ from $\mathbf{x}(n)$ while suppressing $\mathbf{v}(n)$. Earlier works such as [6, 17–19] aimed to achieve the same goal under a direct-path only propagation assumption.

3. PEVD-BASED SPEECH ENHANCEMENT

3.1. Formulation of Polynomial Matrices

The classical subspace methods for speech enhancement do not fully exploit spatial information, and do not account for correlation between different frequency bins. For broadband signals such as speech, different frequency components are affected by different phase shifts. Each phase shift requires specific temporal alignment to be corrected for. Therefore, the correlations across different sensors and at different time lags need to be considered. To achieve this, we use the space-time covariance matrix

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

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

A more compact representation of the speech signals, that captures the correlations in space, time and frequency, can be obtained by representing the speech signals using z-transform, rather than the STFT. The z-transform of (3) is a para-Hermitian polynomial matrix [22, 23]

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

The polynomial matrix can be interpreted as a matrix with polynomial elements or, equivalently, a polynomial with matrix coefficients.

3.2. Polynomial Matrix Eigenvalue Decomposition

The polynomial eigenvalue decomposition (PEVD) of a para-Hermitian matrix [22] is given by

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

where the rows of $\boldsymbol{\mathcal{U}}(z)$ corresponds to the eigenvectors with their associated eigenvalues on the diagonal polynomial matrix, $\boldsymbol{\Lambda}(z)$. The decomposition is computed using an iterative algorithm [22, 24–26] based on similarity transforms involving L para-unitary polynomial matrices, $\boldsymbol{\mathcal{U}}(z) = \boldsymbol{\mathcal{U}}_L(z) \ldots \boldsymbol{\mathcal{U}}_1(z)$. The polynomial matrix at the ℓ-th iteration, $\boldsymbol{\mathcal{U}}_\ell(z)$, satisfies the para-unitary condition [23],

$$\boldsymbol{\mathcal{U}}_\ell^P(z)\boldsymbol{\mathcal{U}}_\ell(z) = \boldsymbol{\mathcal{U}}_\ell(z)\boldsymbol{\mathcal{U}}_\ell^P(z) = \mathbf{I}, \tag{6}$$

where $\mathbf{I}$ is the identity matrix and $[\cdot]^P$ denotes the para-Hermitian operator such that $\boldsymbol{\mathcal{U}}_\ell^P(z) = \boldsymbol{\mathcal{U}}_\ell^H(z^{-1})$. At each iteration, the PEVD algorithm [22] first searches for the largest off-diagonal element (column norm) before applying a delay matrix to bring the dominant element (column) to the principal plane, the plane of z^0, if it exceeds a predefined threshold, δ. The dominant element (column) is then zeroed out using a unitary matrix computed based on the principal plane but 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 [22]. After L iterations, $\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z)$ is approximately diagonalized according to

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

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) [22], that targets the dominant element, or Householder-like optimization procedure as in [26]. A combination of Householder reflection and Givens rotation matrices is used in [25] and the sequential matrix diagonalization (SMD) algorithm [24], that targets the dominant column, uses the eigenvector matrix.

3.3. Proposed PEVD-based Algorithm

We now consider the case when $\mathbf{h}_m$ in (1) is an arbitrary acoustic impulse response. The m-th channel impulse response, $\mathbf{h}_m$, can be separated into [21]

$$\mathbf{h}_m = \tilde{\mathbf{h}}_{m,dp} + \tilde{\mathbf{h}}_{m,er} + \tilde{\mathbf{h}}_{m,lr}, \tag{8}$$

where $\tilde{\mathbf{h}}_{m,dp}, \tilde{\mathbf{h}}_{m,er}, \tilde{\mathbf{h}}_{m,lr}$ are the direct-path, early reflections and late reflections components respectively. We note that, in contrast to [17], (8) incorporates explicitly the responses corresponding to the early reflections and late reverberation.

Since the late reflections comprise randomly distributed small amplitude components which makes the third term in (8) uncorrelated with the first two, it can be treated as an additive, uncorrelated noise component. The early reflections, on the other hand, represents closely spaced echoes which have a direct-path strengthening effect and may improve speech intelligibility in some conditions. Applying (8) in (1), the signal model can be divided into the signal and noise components such that

$$\mathbf{x}(n) = \tilde{\mathbf{s}}(n) + \tilde{\mathbf{v}}(n), \tag{9}$$

where $\tilde{\mathbf{s}}(n) = [s_1(n), \ldots, s_M(n)]^T$ with $s_m(n) = \tilde{\mathbf{h}}_{m,dp}^T\mathbf{s}_0(n) + \tilde{\mathbf{h}}_{m,er}^T\mathbf{s}_0(n)$ and $\tilde{\mathbf{v}}(n) = [v_1(n), \ldots, v_M(n)]^T$ with $v_m(n) = \tilde{\mathbf{h}}_{m,lr}^T\mathbf{s}_0(n) + v_m(n)$. Applying (9) to (3), the space-time covariance matrix of the microphone signals is given by

$$\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z) = \boldsymbol{\mathcal{R}}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}(z) + \boldsymbol{\mathcal{R}}_{\tilde{\mathbf{v}}\tilde{\mathbf{v}}}(z), \tag{10}$$

where the speech and noise space-time covariance matrices are $\boldsymbol{\mathcal{R}}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}(z)$ and $\boldsymbol{\mathcal{R}}_{\tilde{\mathbf{v}}\tilde{\mathbf{v}}}(z)$ respectively, assuming stationarity within each processing frame. By filtering $\mathbf{x}(z)$ through the filterbank $\boldsymbol{\mathcal{U}}(z)$, the channel outputs, $\mathbf{y}(z) = \boldsymbol{\mathcal{U}}(z)\mathbf{x}(z)$, are strongly decorrelated [22] according to

$$\mathbb{E}\{\mathbf{y}(z)\mathbf{y}^P(z)\} = \mathbb{E}\{\boldsymbol{\mathcal{U}}(z)\mathbf{x}(z)\mathbf{x}^P(z)\boldsymbol{\mathcal{U}}^P(z)\} \approx \boldsymbol{\Lambda}(z). \tag{11}$$

Since noise and speech are assumed uncorrelated, the PEVD gives

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

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 reflection components while the noise subspace contains ambient noise and the late reflections in the reverberant channel.

PEVD algorithms sort $\mathbf{\Lambda}(z)$ in descending order which tends to result in the spectrally majorized property [22]. Consequently, noise reduction in the output channels is achieved by combining components in the signal subspace and nulling components in the noise subspace. Unlike previous subspace-based methods and even many non-subspace methods, the proposed method does not rely on a noise estimation since the strong decorrelation property of PEVD implicitly separates the noise and signal components if they are uncorrelated. The PEVD-based speech enhancement algorithm is summarized in Algorithm 1.

Algorithm 1 PEVD-based speech enhancement.

Inputs: $\mathbf{x}(n) \in \mathbb{R}^M, n \in \{0, \ldots, T-1\}, W, \delta, \mu, L.$
$\mathbf{R}_{\mathbf{xx}}(\tau) \leftarrow E\{\mathbf{x}(n)\mathbf{x}^T(n-\tau)\}$ *// see* (3)
$\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z) \leftarrow \mathcal{Z}\{\mathbf{R}_{\mathbf{xx}}(\tau)\}$ *// see* (4)
$\boldsymbol{\mathcal{U}}(z), \mathbf{\Lambda}(z) \leftarrow$ PEVD $\{\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z), \delta, \mu, L\}$ *// use any PEVD algorithm [22, 24–26]*
$\mathbf{x}(z) \leftarrow \mathbf{x}(n)$ *// see* (4)
$\mathbf{y}(z) \leftarrow \boldsymbol{\mathcal{U}}(z)\mathbf{x}(z)$ *// speech enhancement*
return $\mathbf{y}(z)$.

4. SIMULATIONS AND RESULTS

4.1. Experiment Setup

To evaluate the proposed approach, speech signals, sampled at 16 kHz, from the TIMIT corpus [27] and room impulse responses and noise measurements from the 3-channel mobile recordings in the ACE corpus [28] are used. The room impulses used are Office 1 and Lecture Room 2 with measured T_{60} of 0.332 s and 1.22 s respectively. To highlight the insights gain from incorporating reverberation, we also include a direct-path only propagation. The propagation delays are drawn from the discrete uniform distribution, $U(1, 1000)$ and ordered such that $\tau_1 < \tau_2 < \tau_3$.

Monte-Carlo simulations involving 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 then convolved with the impulse response at each microphone channel before being corrupted by additive noise using [29]. The noise condition used in the simulations is babble noise ranging from -10 dB to 30 dB signal to noise ratio (SNR).

The PEVD parameters, adapted from [22], 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 $L = 500$. To estimate $\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z)$ in (4), $\mathbf{R}_{\mathbf{xx}}(\tau)$ in (3) is first computed based on the sample mean given by

$$\hat{\mathbf{R}}_{\mathbf{xx}}(\tau) \approx \frac{1}{T} \sum_{n=0}^{T-1} \mathbf{x}(n)\mathbf{x}^T(n-\tau), \tag{12}$$

and $\tau = \pm W$, where W is the truncation window that reflects the temporal correlation of speech signals. In the experiments, we found a good choice to be $T = W = 1600$ samples so that $\hat{\boldsymbol{\mathcal{R}}}_{\mathbf{xx}}(z)$ is recursively estimated every 100 ms.

The proposed PEVD method is compared against the log-MMSE method with published parameters in [3], two subspace methods and two versions of the multi-channel Wiener filter (MWF), which are based on the concatenation of a minimum variance distortionless response (MVDR) followed by a single-channel Wiener filter [30]. The first MWF uses a speech estimator that exploits the relative transfer function and a noise estimator based on the parameters used in [31]. The second is the Oracle-MWF (O-MWF) which will approximate the ideal performance bound since it uses complete prior knowledge of the clean speech signal. The parameters are based on the batch version in [5] where the filter length is 80. The two subspace methods are single-channel method for coloured noise by [9] (COLSUB) and multi-channel subspace (MCSUB) methods [10, 11].

4.2. Performance Measures

For evaluation, the segmental signal to noise ratio (SegSNR), frequency-weighted SegSNR (FwSegSNR) [32], short-time objective intelligibility (STOI) [33] and perceptual evaluation of speech quality (PESQ) [34] scores are used. These measures are computed for the signals before and after enhancement using the proposed and baseline algorithms. We then compute the improvement, Δ, by taking the difference in the scores between the enhanced and noisy signals. Higher ΔSegSNR and ΔFwSegSNR indicate greater noise reduction while higher ΔSTOI and ΔSTOI indicate improvement in speech intelligibility and quality.

4.3. Results and Discussions

Fig. 1 shows the results for reverberant speech corrupted by babble noise in the highly reverberant Lecture Room 2 of the ACE corpus. At the lower range of SNR from -10 dB to 10 dB, in terms of noise reduction, the subspace-based methods, COLSUB and MCSUB outperform the other methods and were able to provide an improvement in SegSNR of up to 7 dB. COLSUB also performs well in FwSegSNR but this is different for MCSUB, which does not seem to improve even though it performs well for SegSNR. At SNRs above 15 dB, log-MMSE outperforms the other methods in both SegSNR and FwSegSNR. At SNRs above -5 dB, PEVD improves both SegSNR and FwSegSNR. OMWF offers noise reduction at lower SNRs up to 5 dB before beginning to introduce more noise into the noisy received signal at SNRs above that.

Even with the knowledge of the clean speech, OMWF does not necessarily improve both SegSNR and FwSegSNR. This is expected because the OMWF attempts to reduce noise while minimizing speech distortion. This is reflected in both STOI and PESQ, where OMWF is shown to outperform the other algorithms. PEVD is always able to improve STOI over the entire range of SNR from -10 dB to 30 dB and, in fact, it performs better than OMWF at SNRs above 10 dB. The subspace methods, COLSUB and MCSUB, on the other hand, worsen STOI at lower SNR values and offer no improvement at higher SNRs. Log-MMSE always worsens the STOI after enhancement.

In terms of PESQ, OMWF outperforms the rest across all SNRs with PEVD being competitively close. At SNRs above 5 dB, the algorithms perform comparatively. Both subspace methods, COLSUB and MCSUB perform the worst at SNRs below 0 dB.

From these examples, we observe a trade-off between noise reduction and speech intelligibility, as indicated by the STOI measure. The noise reduction offered by MMSE and subspace-based methods, COLSUB and MCSUB, is achieved at the expense of speech intelligibility. At the other extreme, OMWF, which has knowledge of the clean speech, performs the best in terms of speech intelligibility and quality. This also reflects the fact that speech intelligibility may not necessarily be affected by noise levels, up to some limit, compared to speech. PEVD, on the other hand, is able to simultaneously suppress the noise while improving intelligibility.

Similar results are observed in the mildly reverberant Office 1 environment as shown in Fig 2. PEVD was able to approach the performance of OMWF in STOI while reducing the noise level as demonstrated by both SegSNR and fwSegSNR. It is able to approach the performance of subspace-based methods, which are capable of suppressing noise to a greater level, but at the expense of distorting the speech signal.

For anechoic environments, where only the direct-path response is simulated, OMWF offers the highest SegSNR and FwSegSNR improvement. This, however, is not observed when the environment is reverberant. Hence, the direct-path only model is insufficient for noise reduction using the OMWF method when the environment is reverberant. The performance of PEVD approaches that of OMWF in terms of STOI. While OMWF gives the best performance in all metrics, it relies on prior knowledge of the clean speech signal compared to PEVD, that is a completely blind method and does not rely on noise estimation.

Furthermore, the listening examples, which can be found on `https://www.commsp.ee.ic.ac.uk/%7esap/pevdr/`, have indicated that our approach does not introduce any processing artefacts such as musical noise into the enhanced signal.

5. CONCLUSION

A PEVD-based speech enhancement algorithm designed for a reverberant signal model has been proposed. By decomposing the impulse response of the reverberant channel into the direct path, early and late reflections, the new method is shown to be able to exploit the lack of correlation between the source and the late reflections to provide further noise reduction. It has been demonstrated that even without a noise estimator, the proposed method is robust to reverberation and is able to simultaneously reduce noise and improve speech intelligibility and quality. This is in strong contrast to the other methods that require a trade-off between both quantities. Furthermore, informal listening examples highlighted that the approach does not introduce any processing artefacts such as musical noise.

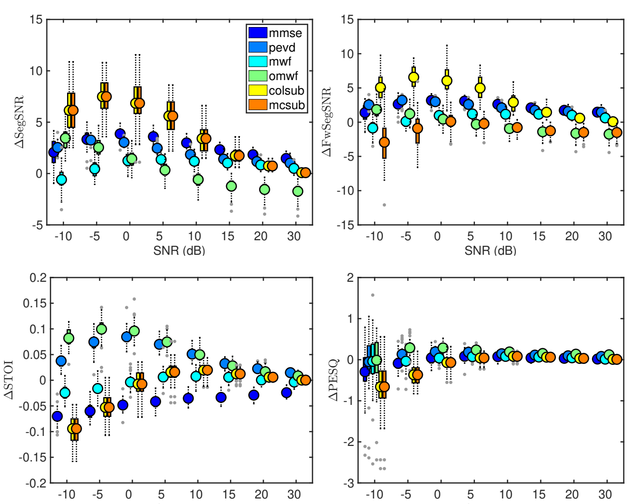

Fig. 1. Simulations for babble noise in Lecture Room 2.

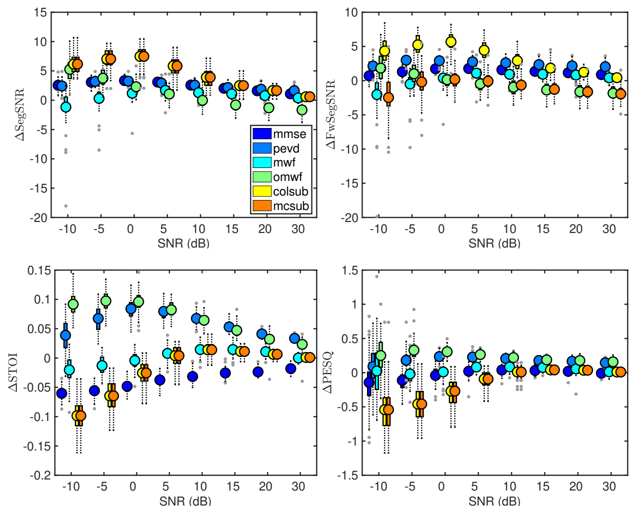

Fig. 2. Simulations for babble noise in Office 1.

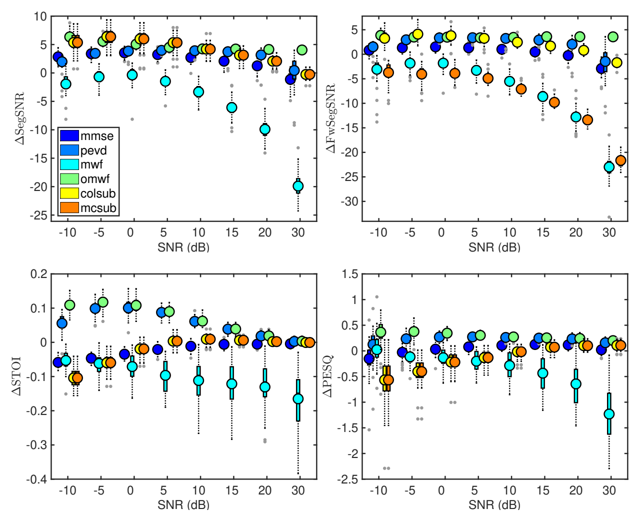

Fig. 3. Simulations for babble noise in direct-path only propagation.

6. REFERENCES

[1] T. Esch and P. Vary, "Efficient musical noise suppression for speech enhancement systems," in *Proc. IEEE Intl. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, 2009, pp. 4409–4412.

[2] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, no. 2, pp. 113–120, Apr. 1979.

[3] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[4] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 443–445, 1985.

[5] 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, Sept. 2002.

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

[7] 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, July 1995.

[8] A. Rezayee and S. Gazor, "An adaptive KLT approach for speech enhancement," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 2, pp. 87–95, 2001.

[9] Y. Hu and P. C. Loizou, "A subspace approach for enhancing speech corrupted by colored noise," *IEEE Signal Process. Lett.*, vol. 9, no. 7, pp. 204–206, July 2002.

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

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

[12] F. Asano, S. Hayamizu, T. Yamada, and S. Nakamura, "Speech enhancement based on the subspace method," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 5, pp. 497–507, 2000.

[13] F. Asano, S. Ikeda, M. Ogawa, H. Asoh, and N. Kitawaki, "Combined approach of array processing and independent component analysis for blind separation of acoustic signals," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 3, pp. 204–215, May 2003.

[14] S. Redif, S. Weiss, and J. G. McWhirter, "Relevance of polynomial matrix decompositions to broadband blind signal separation," *Signal Processing*, vol. 134, pp. 76–86, May 2017.

[15] 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.

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

[17] V. W. Neo, C. Evers, and P. A. Naylor, "Speech enhancement using polynomial eigenvalue decomposition," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2019.

[18] K. Hermus, P. Wambacq, and H. Van hamme, "A review of signal subspace speech enhancement and its application to noise robust speech recognition," *EURASIP J. on Advances in Signal Processing*, vol. 2007, no. 45821, pp. 15, 2006.

[19] J. Benesty, M. G. Christensen, J. R. Jensen, and J. Chen, "A brief overview of speech enhancement with linear filtering," *EURASIP J. on Advances in Signal Processing*, vol. 2014, no. 1, pp. 162, 2014.

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

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

[22] 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.

[23] P. P. Vaidyanathan, *Multirate Systems and Filters Banks*, Prentice Hall, New Jersey, USA, 1993.

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

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

[26] S. Redif, S. Weiss, and J. G. McWhirter, "An approximate polynomial matrix eigenvalue decomposition algorithm for para-Hermitian matrices," in *Proc. Intl. Symp. on Signal Process. and Inform. Technology (ISSPIT)*, Bilbao, Spain, 2011, pp. 421–425.

[27] 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," Corpus, Linguistic Data Consortium (LDC), Philadelphia, 1993.

[28] 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, Lang. Process.*, vol. 24, no. 10, pp. 1681–1693, Oct. 2016.

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

[30] 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., pp. 39–60. Springer-Verlag, Berlin, Germany, 2001.

[31] A. Kuklasiński, S. Doclo, S. H. Jensen, and J. Jensen, "Maximum Likelihood PSD Estimation for Speech Enhancement in Reverberation and Noise," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 9, pp. 1599–1612, Sept. 2016.

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

[33] 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, Lang. Process.*, vol. 19, no. 7, pp. 2125–2136, Sept. 2011.

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