SPEECH ENHANCEMENT USING POLYNOMIAL EIGENVALUE DECOMPOSITION

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

Imperial College London
Department of Electrical and Electronic Engineering, South Kensington, London SW7 2AZ, UK
{vincent.neo09, c.evers, p.naylor}@imperial.ac.uk

ABSTRACT

Speech enhancement is important for applications such as telecommunications, hearing aids, automatic speech recognition and voice-controlled system. The enhancement algorithms aim to reduce interfering noise while minimizing any speech distortion. In this work for speech enhancement, we propose to use polynomial matrices in order to exploit the spatial, spectral as well as temporal correlations between the speech signals received by the microphone array. Polynomial matrices provide the necessary mathematical framework in order to exploit constructively the spatial correlations within and between sensor pairs, as well as the spectral-temporal correlations of broadband signals, such as speech. Specifically, the polynomial eigenvalue decomposition (PEVD) decorrelates simultaneously in space, time and frequency. We then propose a PEVD-based speech enhancement algorithm. Simulations and informal listening examples have shown that our method achieves noise reduction without introducing artefacts into the enhanced signal for white, babble and factory noise conditions between -10 dB to 30 dB SNR.

Index Terms— Polynomial eigenvalue decomposition, broadband multi-channel processing, strong decorrelation, speech enhancement, signal denoising.

1. INTRODUCTION

Speech enhancement of noise corrupted signals remains an important research area due to its relevance in diverse applications ranging from human-to-human communications in telecommunications and hearing aids to human-to-machine interaction in automatic speech recognition, voice-controlled systems and robot audition. Speech enhancement systems aim to reduce interfering noise but may distort the speech signal and introduce processing artefacts such as musical noise, arising from large narrowband fluctuations in the residual noise [1]. To trade off noise reduction against speech distortion, a control parameter is commonly introduced [2, 3]. For instance, aggressive noise reduction might be preferred for human-to-machine applications while mobile phone and hearing aids users might prefer less speech distortion at the expense of higher residual noise.

Existing approaches to speech enhancement can be classified into single- and multi-channel techniques. Methods for single-channel enhancement include spectral subtraction [4, 5], statistical-based and subspace-based approaches. Statistical methods are typically based on minimising the mean-square error (MSE) of the clean and estimated speech spectrum [2], the log-spectrum (log-MMSE) [6] or the single-channel Wiener filter [3, 7]. In subspace methods, noisy signals are decomposed into signal and noise subspaces and enhancement is achieved by recovering the speech signal from the signal subspace [8, 9].

In [3, 7], the multi-channel Wiener filter (MWF) was derived as an optimal filter. Extensions to the work include post-filtering [10], beamformers [11, 12, 13] and the multi-channel Kalman filter [14]. While existing multi-channel approaches may potentially achieve noise reduction without speech distortion under spatially temporally white noise conditions, this remains a practical challenge under other noise conditions in real-world scenarios [15].

A large number of contributions are based on the assumption of narrowband models. Single-channel approaches exploit temporal correlations while multi-channel approaches capture spatio-temporal correlations. However, extensions to broadband signals cannot account for spectral correlations. The modelling of space-time correlations for broadband signal processing can be achieved using polynomial matrices. This has motivated the development of a family of polynomial eigenvalue decomposition (PEVD) algorithms [16, 17, 18], based on the pioneering second-order sequential best rotation (SBR2) algorithm [19]. Since then, PEVD has been widely used in polynomial multiple signal classification (MUSIC) [20], blind source separation [21], source identification [22] and adaptive beamforming [23].

In this paper, we propose to utilize the PEVD for multi-channel speech enhancement in order to explicitly model the spatial, spectral and temporal correlations of broadband signals impinging on the microphone array. The main, novel contributions are (i) the use of polynomial matrix as a broadband, multi-channel signal model for speech, (ii) the proposal of a novel algorithm for speech enhancement using PEVD that does not introduce any audible artefacts, and (iii) the evaluation of the proposed approach for realistic signals under various noise conditions and a comparison against log-MMSE and the MWF as benchmark approaches.

2. PEVD-BASED PROBLEM FORMULATION

2.1. Signal Model

The output at the m-th microphone at the discrete-time sample n is

$$x_m(n) = \sum_{q=0}^{Q} h_m(n-q)s(q) + v_m(n), \quad m = 1, 2, \ldots, M, \quad (1)$$

where $s(n)$ is the source signal, $h_m(n)$ is the m-th channel modelled as a Qth order finite impulse response filter and $v_m(n)$ is the additive noise signal. The data vector is $\mathbf{x}(n) = [x_1(n), \ldots, x_M(n)]^T \in \mathbb{R}^M$, $n = 0, \ldots, T-1$. Noise signals are assumed to be uncorrelated at different sensors and with the source signal. Furthermore, $\{\cdot\}^*$ denotes the complex conjugate, $\{\cdot\}^T$ is the transpose, $\{\cdot\}^H$ is the Hermitian, $\mathbb{E}\{\cdot\}$ is the ex-

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

pectation operator, $j = \sqrt{-1}$, $\mathbf{I}$ and $\mathbf{0}$ are the identity and zero matrix, $\mathbb{R}$ and $\mathbb{C}$ are the real and complex space, respectively.

2.2. Motivation for Polynomial Matrix Models

Broadband signals received by microphone arrays exhibit spatial, spectral and temporal correlations. The space-time covariance matrix is defined as

$$\mathbf{R}_{\mathbf{xx}}(\tau) = \mathbb{E}\{\mathbf{x}(n)\mathbf{x}^T(n-\tau)\} \in \mathbb{R}^{M\times M}, \tag{2}$$

where the $(p,q)^{\text{th}}$ element, $r_{pq}(\tau) = \mathbb{E}\{x_p(n)x_q(n-\tau)\}$.

If $s(n)$ is a narrowband signal at frequency f_0, $x_m(n)$ is related to the signal, $x_1(n)$, at the reference microphone, $m = 1$, by a constant phase shift, $\phi_m = 2\pi f_0 \tau_m$, assuming only direct-path propagation. Hence, the data vector is equivalent to $\mathbf{x}(n) = [x_1(n), x_1(n)e^{-j\phi_2}, \ldots, x_1(n)e^{-j\phi_M}]^T$, such that

$$\begin{aligned} r_{pq}(\tau) &= \mathbb{E}\{x_1(n)e^{-j\phi_p}[x_1(n-\tau)e^{-j\phi_q}]^*\} \\ &= \mathbb{E}\{x_1(n)x_1(n)\}e^{-j(\phi_p-\phi_q-2\pi f_0\tau)}, \end{aligned} \tag{3}$$

where $r_{pq}(\tau) \in \mathbb{C}$ and $\mathbf{R}_{\mathbf{xx}}(\tau) \in \mathbb{C}^{M\times M}$ for this example. The phase difference between the $(p,q)^{\text{th}}$ sensor pair, $\phi_{pq} = \phi_p - \phi_q$, depends only on the array geometry and is associated with the time-difference of arrival (TDoA). Since the array geometry is fixed and known *a priori*, ϕ_{pq} is constant and can be computed at any τ. Classical subspace-based approaches for the enhancement of narrowband signals approximate (2) by evaluating only the instantaneous spatial covariance matrix at $\tau = 0$ according to

$$\mathbf{R}_{\mathbf{xx}}(0) = \mathbb{E}\{\mathbf{x}(n)\mathbf{x}^T(n)\}. \tag{4}$$

However, for broadband signals such as speech, different frequency components are affected by different phase shifts at the same time lag. Consequently, for speech enhancement, it is crucial to explicitly perform phase corrections for different frequency components at different time lags. Therefore, the correlations across different sensors and temporal lags need to be considered. 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 (2) is a para-Hermitian polynomial matrix [19, 24] such that

$$\mathbf{R}_{\mathbf{xx}}(z) = \sum_{\tau=-\infty}^{\infty} \mathbf{R}_{\mathbf{xx}}(\tau)z^{-\tau} \in \mathbb{C}^{M\times M}, \tag{5}$$

where $z = e^{j2\pi f}$. The polynomial matrix can be interpreted as a matrix with polynomial elements or, equivalently, a polynomial with matrix coefficients.

2.3. Family of PEVD Algorithms

The PEVD of a para-Hermitian matrix [19] is given by

$$\mathbf{R}_{\mathbf{xx}}(z) \approx \mathbf{U}^P(z)\mathbf{\Lambda}(z)\mathbf{U}(z), \tag{6}$$

where the rows of $\mathbf{U}(z)$ are the eigenvectors with corresponding eigenvalues on the diagonal polynomial matrix, $\mathbf{\Lambda}(z)$. The decomposition is computed using an iterative algorithm [19, 16, 17, 18] based on similarity transforms involving L para-unitary polynomial matrices, $\mathbf{U}(z) = \mathbf{U}_L(z)\ldots\mathbf{U}_1(z)$. The polynomial matrix at the ℓ-th iteration, $\mathbf{U}_\ell(z)$, satisfies the para-unitary condition [24],

$$\mathbf{U}_\ell^P(z)\mathbf{U}_\ell(z) = \mathbf{U}_\ell(z)\mathbf{U}_\ell^P(z) = \mathbf{I}, \tag{7}$$

where $\{\cdot\}^P$ denotes the para-Hermitian operator such that $\mathbf{U}_\ell^P(z) = \mathbf{U}_\ell^H(z^{-1})$. At each iteration, the PEVD algorithm [19] 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 [19].

Due to the para-Hermitian symmetry of $\mathbf{R}_{\mathbf{xx}}(z)$, the search space is confined to half the number of off-diagonal elements and similarity transforms ensure that operations act on dominant element (column-row) pairs. After L iterations, $\mathbf{R}_{\mathbf{xx}}(z)$ is approximately diagonalized according to

$$\mathbf{\Lambda}(z) \approx \mathbf{U}(z)\mathbf{R}_{\mathbf{xx}}(z)\mathbf{U}^P(z) = \mathbf{U}(z)\mathbb{E}\{\mathbf{x}(z)\mathbf{x}^P(z)\}\mathbf{U}^P(z), \tag{8}$$

where $\mathbf{x}(z)$ is the z-transform of $\mathbf{x}(n)$ based on (5). The zeroing unitary matrix computed at iteration ℓ can take the form of a Givens rotation in SBR2 [19], that targets the dominant element, or Householder-like optimization procedure as in [18]. A combination of Householder reflection and Givens rotation matrices is used in [17] and the sequential matrix diagonalization (SMD) algorithm [16], that targets the dominant column, uses the eigenvector matrix.

2.4. PEVD-based Speech Enhancement

By filtering $\mathbf{x}(z)$ through the filterbank $\mathbf{U}(z)$, the channel outputs, $\mathbf{y}(z) = \mathbf{U}(z)\mathbf{x}(z)$, are strongly decorrelated [19] according to

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

Since noise and speech are assumed uncorrelated, the PEVD gives

$$\begin{aligned} \mathbf{R}_{\mathbf{xx}}(z) \approx &\left[\begin{array}{c|c} \mathbf{U}_S^P(z) & \mathbf{0} \end{array}\right] \left[\begin{array}{c|c} \mathbf{\Lambda}_S(z) & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \left[\begin{array}{c} \mathbf{U}_S(z) \\ \hline \mathbf{0} \end{array}\right] \\ &+ \left[\begin{array}{c|c} \mathbf{0} & \mathbf{U}_V^P(z) \end{array}\right] \left[\begin{array}{c|c} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{\Lambda}_V(z) \end{array}\right] \left[\begin{array}{c} \mathbf{0} \\ \hline \mathbf{U}_V(z) \end{array}\right], \end{aligned} \tag{10}$$

where $\{.\}_S$ and $\{.\}_V$ represent the orthogonal signal and noise subspace components. PEVD algorithms sort $\mathbf{\Lambda}(z)$ in descending order which tends to result in the spectrally majorized property [19]. Consequently, noise reduction in the output channels is achieved by combining components in the signal subspace and nulling components in the noise subspace. The PEVD-based speech enhancement is summarized in Algorithm 1.

3. SIMULATIONS AND RESULTS

3.1. Experiment Setup

To evaluate the proposed approach, noisy speech signals are generated at 16 kHz sampling frequency using anechoic speech from

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* (2)
$\mathbf{R}_{\mathbf{xx}}(z) \leftarrow \mathcal{Z}\{\mathbf{R}_{\mathbf{xx}}(\tau)\}$ *// see* (5)
$\mathbf{U}(z), \mathbf{\Lambda}(z) \leftarrow$ PEVD $\{\mathbf{R}_{\mathbf{xx}}(z), \delta, \mu, L\}$ *// use any PEVD algorithm [16, 17, 18, 19]*
$\mathbf{x}(z) \leftarrow \mathbf{x}(n)$ *// see* (5)
$\mathbf{y}(z) \leftarrow \mathbf{U}(z)\mathbf{x}(z)$ *// speech enhancement*
return $\mathbf{y}(z)$.

TIMIT corpus [25] and babble as well as factory noise signals from the Noisex database [26]. Monte-Carlo simulations involving 150 trials are conducted. In each trial, sentences from a randomly selected speaker are concatenated so that each lasted for 8 to 10 s and are then corrupted by additive noise using [27]. The noise conditions used in the simulations include white, babble and factory noise ranging from -10 dB to 30 dB signal to noise ratio (SNR). For the multi-channel algorithms, the propagation delays for the 3 channels are drawn from the discrete uniform distribution, $U(1, 1000)$ and are ordered such that $\tau_1 > \tau_2 > \tau_3$.

The PEVD parameters, adapted from [19], 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 $\mathbf{R}_{\mathbf{xx}}(z)$ in (5), $\mathbf{R}_{\mathbf{xx}}(\tau)$ in (2) is first computed based on the sample mean given by

$$\hat{\mathbf{R}}_{\mathbf{xx}}(\tau) \approx \frac{1}{T} \sum_{n=0}^{T-1} \mathbf{x}(n)\mathbf{x}^T(n-\tau), \tag{11}$$

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

The proposed PEVD method is compared against the log-MMSE method in [2] and two versions of the MWF, which are based on the concatenation of a minimum variance distortionless response (MVDR) followed by a single-channel Wiener filter [28]. The first MWF uses a speech estimator that exploits the relative transfer function and a noise estimator based on the parameters used in [14]. 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 [3] where the filter length is 80.

3.2. Performance Measures

For performance evaluation, the segmental signal to noise ratio (SegSNR), frequency-weighted SegSNR (FwSegSNR) [29], short-time objective intelligibility (STOI) [30] and perceptual evaluation of speech quality (PESQ) [31] scores are averaged over all 150 trials for the proposed approach, benchmark algorithms and noisy signals.

3.3. Results and Discussions

Fig. 1 shows the results for a single trial based on a clean speech corrupted by 5 dB babble noise. During the silence period from 2.2 to 2.4 s, the energy of the babble noise has been significantly reduced after the PEVD-based enhancement. For the log-MMSE-based enhancement, the structures of the babble noise and speech signal are lost and the remaining large narrowband fluctuations result in musical noise. For this example, Table 1 shows that O-MWF performed the best in all measures because it uses prior knowledge of the clean

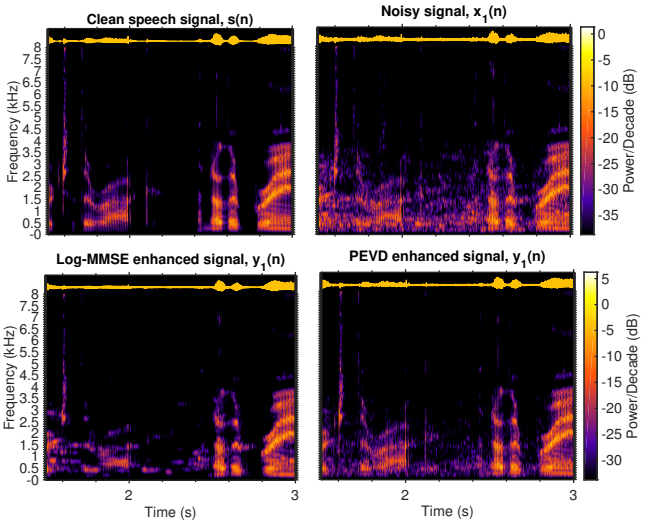

Figure 1: Spectrograms of clean speech, noisy and enhanced signals using log-MMSE and PEVD for a 5 dB babble noise example.

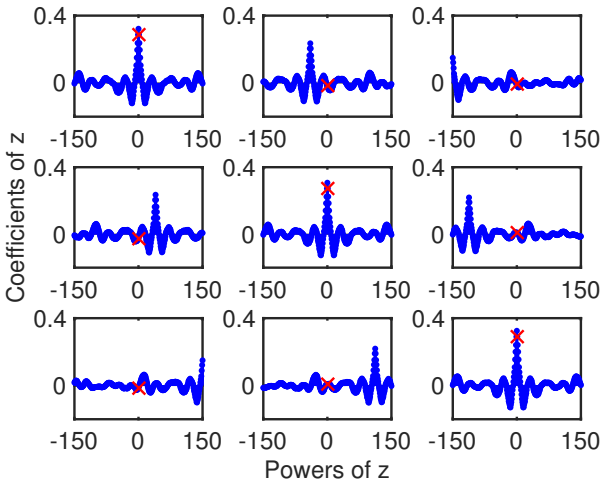

Figure 2: $\hat{\mathbf{R}}_{\mathbf{xx}}(z)$ for the example in Fig. 1 represented by blue dots and $\hat{\mathbf{R}}_{\mathbf{xx}}(z^0)$ is represented by red cross signs.

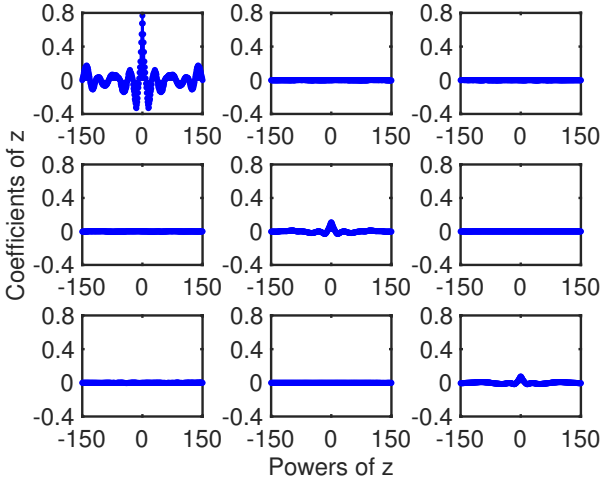

Figure 3: $\mathbf{\Lambda}(z)$ produced by using PEVD for the example in Fig. 2.

Algorithm	ΔSegSNR	ΔFwSegSNR	ΔSTOI	ΔPESQ
log-MMSE	3.69 dB	2.46 dB	-0.007	0.08
MWF	1.07 dB	1.54 dB	0.002	0.15
O-MWF	4.67 dB	4.04 dB	0.084	0.31
PEVD	4.30 dB	4.00 dB	0.080	0.29

Table 1: Algorithm performance compared to noisy signal before enhancement for speech corrupted by 5 dB babble noise example.

speech signal. PEVD outperforms both log-MMSE and MWF in all aspects and is slightly worse than O-MWF by 0.37 dB in SegSNR, 0.04 dB in FwSegSNR, 0.004 in STOI and 0.02 in PESQ. In fact, enhancement using log-MMSE results in a poorer STOI value for this example. This example shows that the performance of PEVD can be comparable to a supervised algorithm that uses prior knowledge of the clean speech signal.

Fig. 2 shows $\hat{\mathbf{R}}_{\mathbf{xx}}(z)$ for the above example used in PEVD. In contrast to the PEVD, the eigenvalue decomposition (EVD) corresponds to the polynomial z^0 in (4). Therefore, the EVD decorrelates the signal only for one specific time lag. This only corrects the phase at a particular frequency and is therefore, not fully adequate for broadband signals which requires phase correction at different lags for different frequencies. Instead, PEVD decorrelates the signal across all time lags by accounting for different phase shifts for different frequency components as demonstrated in Fig. 3, where every off-diagonal element has a magnitude less than $\delta = 3.2 \times 10^{-3}$ for this example.

Comparative results for the Monte-Carlo simulations comprising 150 trials involving white noise, babble noise and factory noise are shown in Fig. 4, 6 and 7 respectively. As expected, O-MWF performs the best in all metrics for almost all cases because it uses prior knowledge of the clean speech signal. In terms of STOI, PEVD outperforms log-MMSE and MWF under all noise conditions and approaches O-MWF performance bound as SNR increases. In terms of SegSNR and FwSegSNR, log-MMSE performs better than PEVD by up to 5 dB at lower SNR but PEVD performs better than log-MMSE at higher SNR. Both log-MMSE and PEVD perform better than MWF even though the MWF enhanced signal has a slight improvement in SegSNR and FwSegSNR at lower SNR. In terms of PESQ, the performance of PEVD approaches that of O-MWF at higher SNRs. Fig. 5 shows the standard deviation plots for SegSNR and STOI for the white noise simulations which reflect the same trends for other noise conditions. In addition, listening examples, provided on `https://www.commsp.ee.ic.ac.uk/\%7esap/pevd/`, have indicated that unlike log-MMSE and MWF, PEVD does not introduce audible artefacts like musical noise or speech distortions into the enhanced signal while reducing noise substantially.

4. CONCLUSION

We have introduced polynomial matrices as a multi-channel broadband signal model for speech which can capture the spatial, spectral as well as temporal correlation between microphones and PEVD as an approach to decorrelate simultaneously in space, time and frequency. We then outlined a method of PEVD-based speech enhancement and have shown that noise reduction can be achieved using the weighted combination of the signal subspace. Comparative simulations and informal listening examples indicate that the proposed method improves both objective, SegSNR and FwSegSNR, and subjective scores, STOI and PESQ, under the diverse range of noise conditions without introducing any artefacts like musical noise into the enhanced signal.

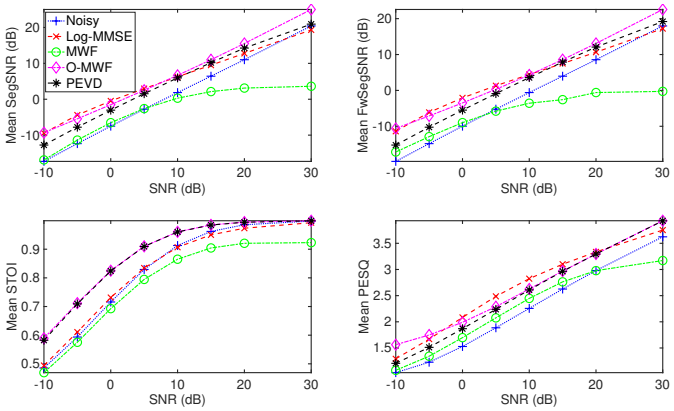

Figure 4: Mean of the results for white noise involving 150 trials.

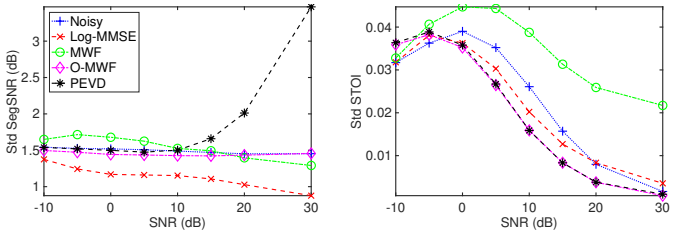

Figure 5: Standard deviation of the results for white noise in Fig. 4.

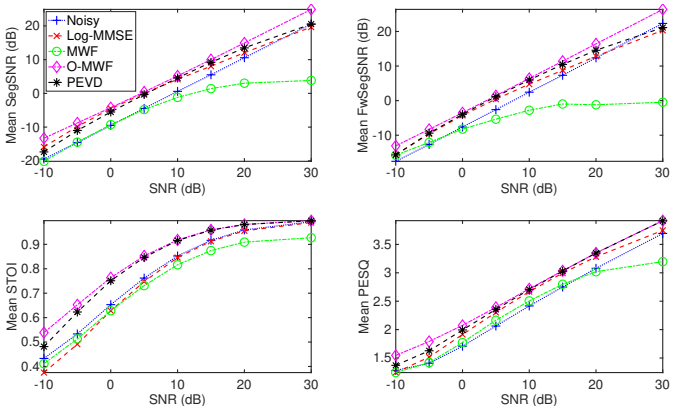

Figure 6: Mean of the results for babble noise involving 150 trials.

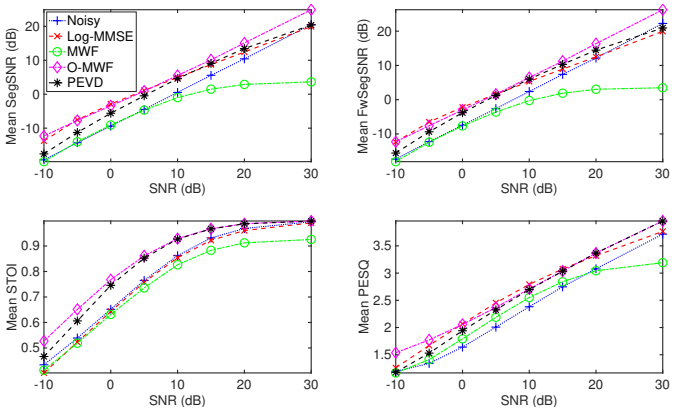

Figure 7: Mean of the results for factory noise involving 150 trials.

5. 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] 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.

[3] 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.

[4] 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.

[5] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol. 67, no. 12, pp. 1586–1604, Dec. 1979.

[6] 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.

[7] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filters," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, pp. 1218–1234, Jul. 2006.

[8] 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.

[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, Jul. 2002.

[10] I. Cohen and B. Berdugo, "Microphone array post-filtering for non-stationary noise suppression," in *Proc. IEEE Intl. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, May 2002, pp. 901–904.

[11] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[12] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, Germany: PUB-SV, 2008.

[13] 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. PUB-SIP, 2017.

[14] W. Xue, A. H. Moore, M. Brookes, and P. A. Naylor, "Modulation-Domain Multichannel Kalman Filtering for Speech Enhancement," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 10, pp. 1833–1847, Oct. 2018.

[15] 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, Jul. 2008.

[16] 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.

[17] 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)*, 2019.

[18] 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)*, 2011, pp. 421–425.

[19] 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.

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

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

[22] 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.

[23] 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.

[24] P. P. Vaidyanathan, *Multirate Systems and Filters Banks*. New Jersey, USA: PUB-PH, 1993.

[25] 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.

[26] A. Varga and H. J. M. Steeneken, "Assessment for automatic speech recognition II: NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech Commun.*, vol. 3, no. 3, pp. 247–251, Jul. 1993.

[27] "Objective measurement of active speech level," Intl. Telecommun. Union (ITU-T), Recommendation P.56, Mar. 1993.

[28] 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: PUB-SV, 2001, pp. 39–60.

[29] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 1, Jan. 2008.

[30] 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, Sep. 2011.

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