Speech Enhancement using Polynomial Eigenvalue Decomposition

Imperial College London

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

21 October 2019

Outline

Introduction

Motivation

Imperial College London

- Single-channel subspace speech enhancement [Ephraim1995; Hu2002]
  - Use an EVD to decorrelate spectrally
- Multi-channel subspace speech enhancement [Asano2000]
  - Use an EVD to decorrelate spatially

$\Rightarrow$ Limitation: Only decorrelates instantaneously

- Other methods typically use STFT to process [Cohen2002; Ephraim1984; Gannot2001; Markovich2009]
  - Use DFT to divide broadband into multiple narrowband signals
  - Require a 4D tensor to model the space, time, spectral correlations

$\Rightarrow$ Limitations: Lacks phase coherence across bands
: Ignores correlation between bands

Motivation for PEVD

- Polynomial Eigenvalue Decomposition (PEVD)
  - Simultaneously captures correlation across space, time and frequency using a 3D tensor
  - Impose spatial decorrelation over a range of time shifts
  - No phase discontinuity
- PEVD-based broadband applications:
  - blind source separation [Redif2017]
  - adaptive beamforming [Weiss2015]
  - source identification [Weiss2017]

This Talk: PEVD for Speech Enhancement

Background

Multichannel Signal Model

The received signal at the q-th sensor with time index n is

$$x_q(n) = \sum_{j=0}^{J} h_q(n-j)s(j) + v_q(n),$$

where

- $s(n)$ is the source signal,
- $h_q(n)$ is the channel modelled as an order J FIR filter,
- $v_q(n)$ is the noise signal at the q-th sensor.

The data vector collected from Q sensors is

$$\mathbf{x}(n) = [x_1(n), x_2(n), \ldots, x_Q(n)]^T.$$

Space-time Covariance Polynomial Matrix

Assuming stationarity, space-time covariance matrix is

$$\mathbf{R}_{\mathbf{xx}}(\tau) = \mathbb{E}[\mathbf{x}(n)\mathbf{x}^H(n-\tau)],$$

where $(i,j)^{\text{th}}$ element is the correlation function $r_{ij}(\tau) = \mathbb{E}[x_i(n)x_j^*(n-\tau)]$ and τ is the time-shift.

Z-transform of $\mathbf{R}_{\mathbf{xx}}(\tau)$ is a para-Hermitian polynomial matrix

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

where $\mathbf{R}_{\boldsymbol{xx}}(\tau) \approx 0$ for $|\tau| > W$, calligraphy $\boldsymbol{\mathcal{R}}$ for tensor and regular $\boldsymbol{R}$ for matrix.

The PEVD of $\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z)$ is defined as [McWhirter2007]

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

where $\boldsymbol{\Lambda}(z), \boldsymbol{\mathcal{U}}(z)$ are the eigenvalue and eigenvector polynomial matrices and $\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z) = \boldsymbol{\mathcal{R}}_{\mathbf{xx}}^{P}(z) = \boldsymbol{\mathcal{R}}_{\mathbf{xx}}^{H}(z^{-1})$.

$\boldsymbol{\mathcal{U}}(z)$ is a filterbank for $\mathbf{x}(z) \in \mathbb{C}^{Q\times 1\times T}$ so that the outputs in

$$\mathbf{y}(z) = \boldsymbol{\mathcal{U}}(z)\mathbf{x}(z) \implies \boldsymbol{\mathcal{R}}_{\mathbf{yy}}(z) \approx \boldsymbol{\Lambda}(z),$$

are strongly decorrelated.

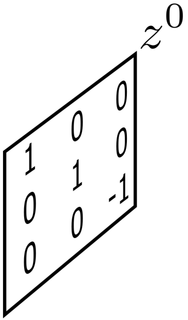
PEVD Algorithms

PEVD algorithms include:

- Second-order Sequential Best Rotation (SBR2) [McWhirter2007]
- Sequential Matrix Diagonalization (SMD) [Redif2015]
- Householder-like PEVD [Redif2011]
- Tridiagonal PEVD [Neo2019]
- Multiple-shift SBR2/SMD [Wang2015; Corr2014]

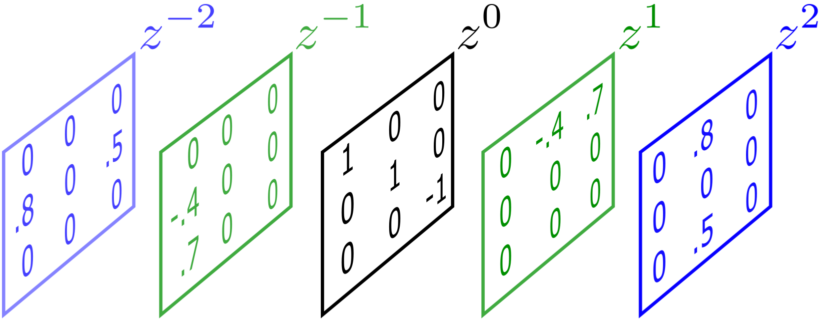
Example of a Polynomial Matrix

Typically compute $\mathbf{R}_{\mathbf{xx}}(0) = \mathbb{E}[\mathbf{x}(n)\mathbf{x}^H(n)]$:



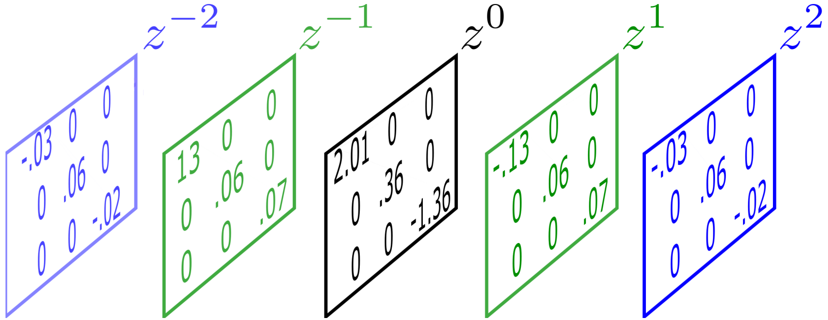
$\mathbf{R}_{\mathbf{xx}}(0)$: instantaneous spatial covariance matrix / coefficient of z^0.

Before diagonalization, $\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z)$:



In this example, z^0 plane is diagonal but not at other planes.

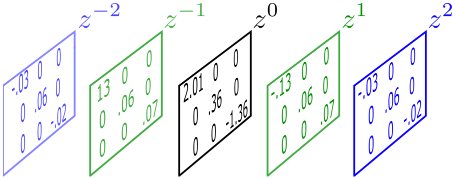
After diagonalization using PEVD, $\boldsymbol{\Lambda}(z)$:



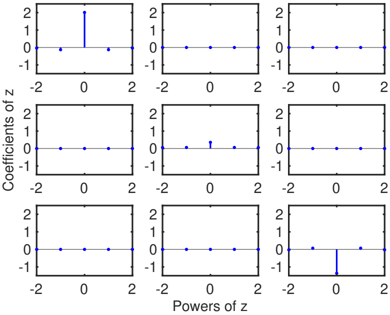
Alternate Representation of Example

Imperial College London

Equivalently, expressed as:



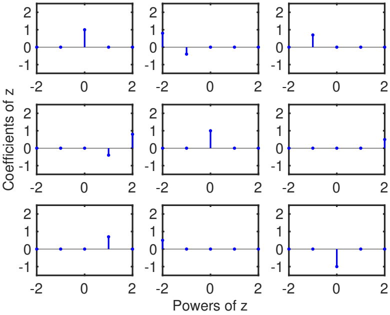
Polynomial with matrix coefficients.



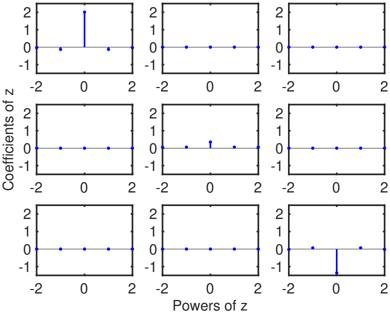
Matrix with polynomial elements.

Alternate Representation of Example

The same example can be represented as:



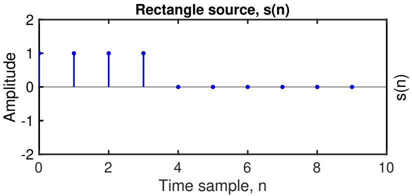
Original $\mathcal{R}_{\mathbf{xx}}(z)$.



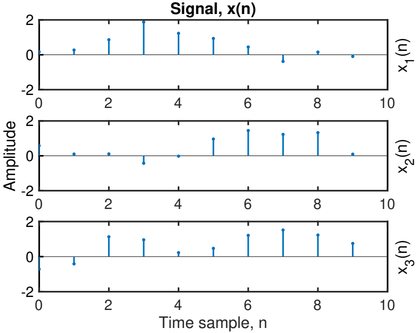
Diagonalized $\boldsymbol{\Lambda}(z)$.

Application Examples

A rectangular pulse source signal arriving at the 3 sensors, corrupted by i.i.d. sensor noise: $\mathcal{N}(0, 0.1^2)$.



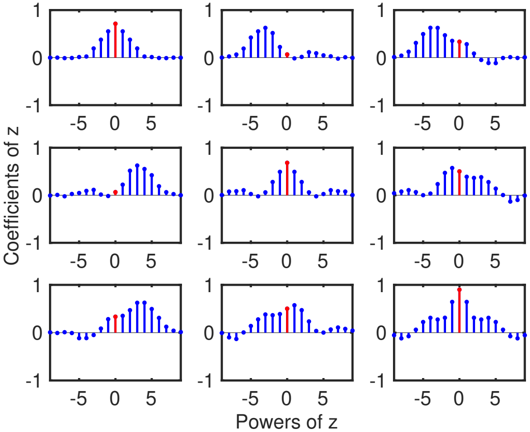
Source signal, $s(n)$.



Received signals, $\mathbf{x}(n)$.

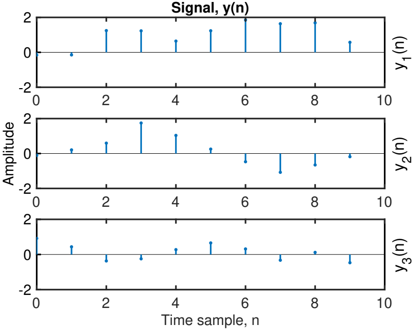
Broadband Example: ST-Covariance

Corresponding space-time covariance matrix, $\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z)$

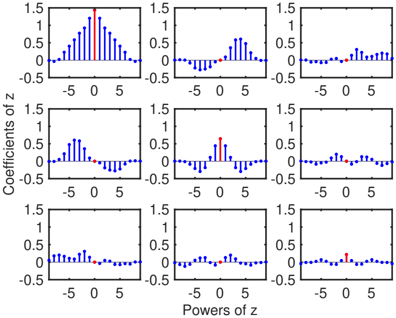


- instantaneous covariance, $\mathbb{E}[\mathbf{x}(n)\mathbf{x}^H(n)]$, marked in red.

Using $\mathbf{U}$ from EVD gives:

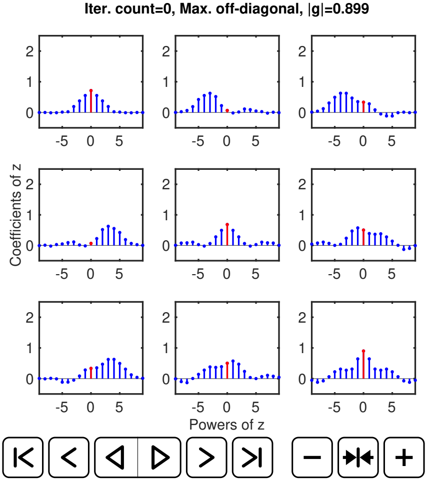


Weighted output, $\mathbf{y}(n)$.



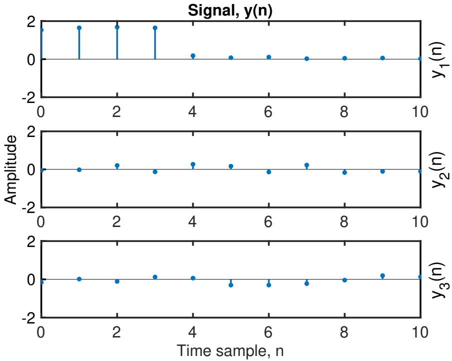
ST-covariance, $\boldsymbol{\mathcal{R}}_{\mathbf{yy}}(z)$.

Diagonalization using PEVD with $\delta = 0.0077$ gives:

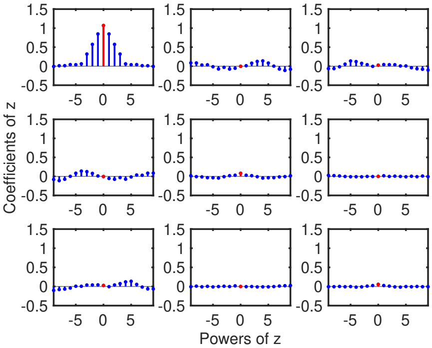


Broadband Example: PEVD

Using $\boldsymbol{\mathcal{U}}(z)$ from PEVD using $\delta = 0.0077$ gives:



Weighted output, $\mathbf{y}(n)$, with arbitrary delays compensated.



ST-covariance, $\boldsymbol{\mathcal{R}}_{\mathbf{yy}}(z)$.

Proposed Methodology

Application: Speech Enhancement

Imperial College London

If $s(n)$ is a speech signal, uncorrelated with noise

$$\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z) = \left[\, \boldsymbol{\mathcal{U}}_S^P(z) \,\middle|\, \boldsymbol{\mathcal{U}}_V^P(z) \,\right] \left[\begin{array}{c|c} \boldsymbol{\Lambda}_S(z) & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{\Lambda}_V(z) \end{array}\right] \left[\begin{array}{c} \boldsymbol{\mathcal{U}}_S(z) \\ \hline \boldsymbol{\mathcal{U}}_V(z) \end{array}\right],$$

with orthogonal signal, $\{\cdot\}_S$ and noise subspaces, $\{\cdot\}_V$.

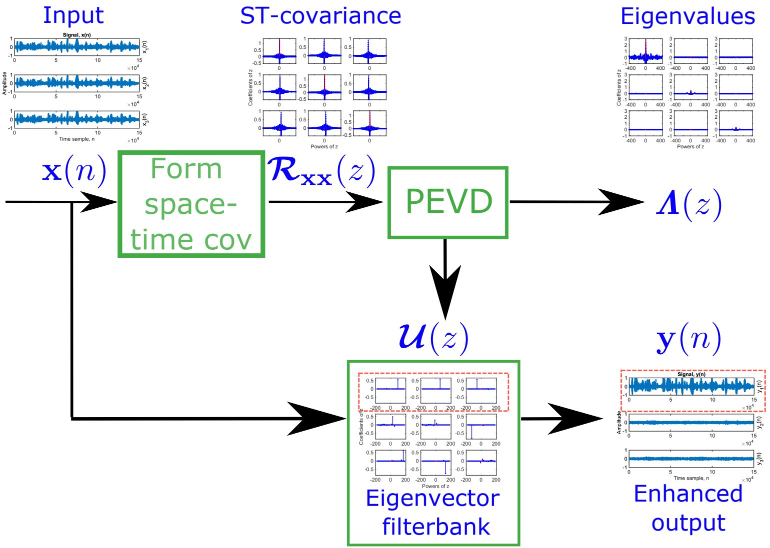
The output

$$\mathbf{y}(z) = \boldsymbol{\mathcal{U}}(z)\mathbf{x}(z),$$

has the first element, $y_1(z) \in \mathbb{R}^{1\times1\times T}$, as the denoised speech signal with space-time covariance matrix

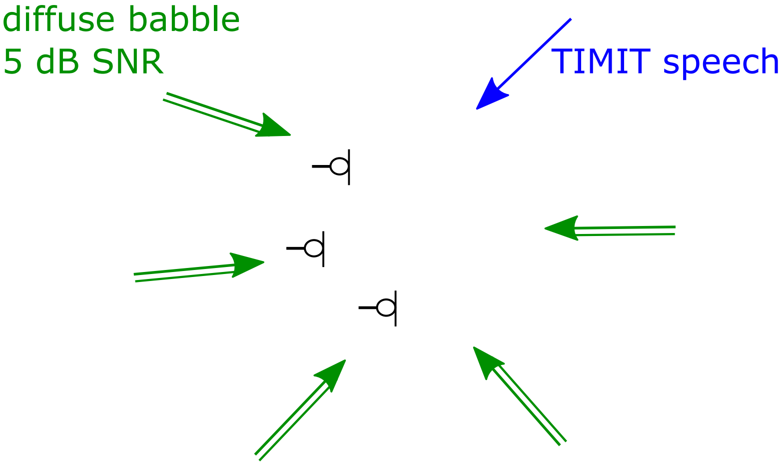
$$\boldsymbol{\mathcal{R}}_{y_1y_1} = \left[\, \boldsymbol{\mathcal{U}}_S^P(z) \,\middle|\, \mathbf{0} \,\right] \left[\begin{array}{c|c} \boldsymbol{\Lambda}_S(z) & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \left[\begin{array}{c} \boldsymbol{\mathcal{U}}_S(z) \\ \hline \mathbf{0} \end{array}\right].$$

Speech Enhancement Using PEVD



Experimental Results

Speech in Noise (Anechoic)



Evaluation

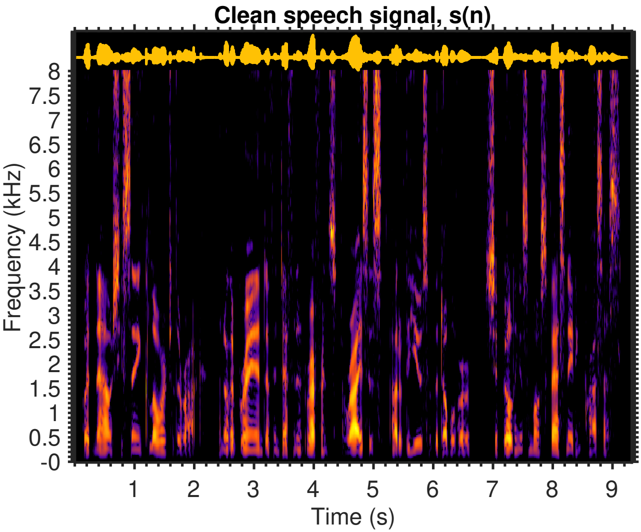
Comparative algorithms:

1. Log-Minimum Mean Square Error (Log-MMSE) [Ephraim 1984]
2. Multichannel Wiener Filter (MWF) - Relative Transfer Function (RTF) and noise estimator [Kuklasiński2016]
3. Oracle-MWF (O-MWF) - Given clean speech [Doclo2002]

Evaluation measures:

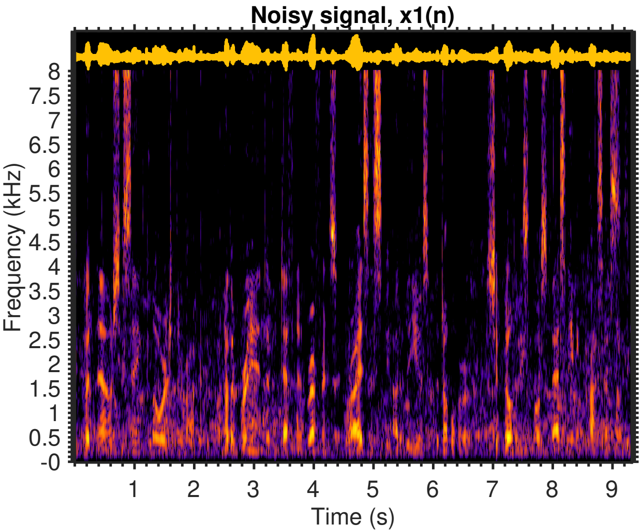
- Segmental SNR (SegSNR)
- Frequency weighted SegSNR (fwSegSNR) [Hu2008]
- STOI [Taal2011]
- PESQ [ITU-T P.862]

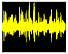
Clean Spectrogram



Clean | Noisy | Log-MMSE | PEVD

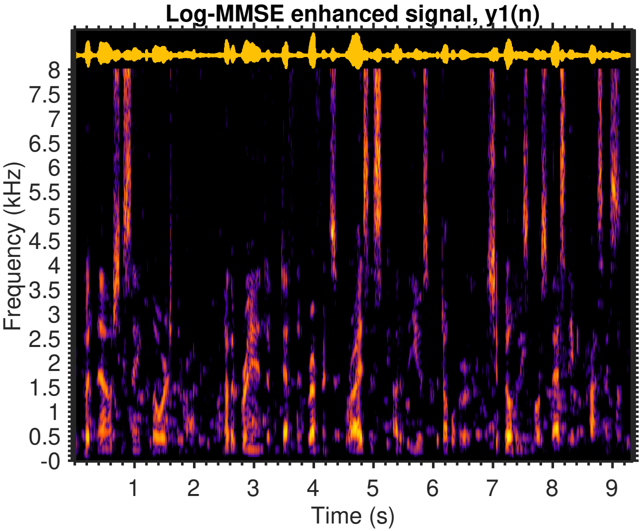
Noisy Spectrogram (5 dB diffuse babble)





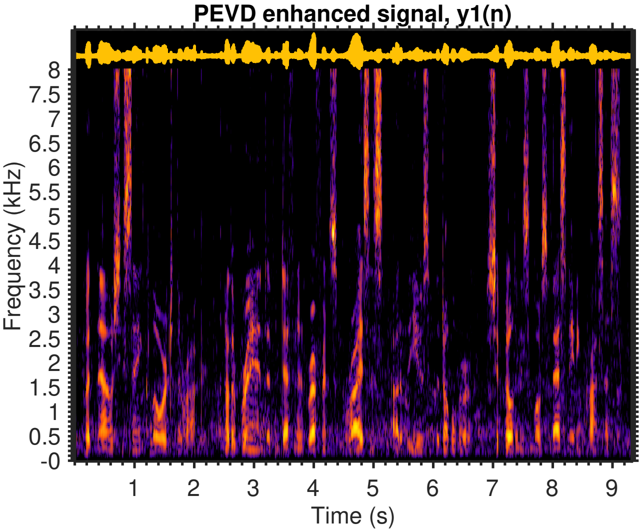
Clean | Noisy | Log-MMSE | PEVD

Log-MMSE-Enhanced Spectrogram



Clean | Noisy | Log-MMSE | PEVD

PEVD-Enhanced Spectrogram



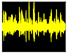
Clean | Noisy | Log-MMSE | PEVD

Comparison of Enhancement Algorithms

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

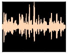
Clean

Noisy

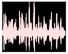


log-MMSE

MWF



O-MWF



PEVD

Conclusion

Conclusion

- Polynomial covariance matrices and PEVD as a tool for processing broadband multichannel signals
  - Polynomial matrices can simultaneously capture the correlation across space, time and frequency
  - PEVD can impose stronger decorrelation than the EVD
- Proposed a speech enhancement algorithm using PEVD
  - Performance approaches oracle MWF when SNR$\geq 5\ dB$
  - No noticeable artifacts

References

Asano, F., S. Hayamizu, T. Yamada, and S. Nakamura (2000). "Speech Enhancement Based on the Subspace Method". In: *IEEE Trans. Speech Audio Process.* 8.5, pp. 497–507.

Cohen, I. and B. Berdugo (Jan. 2002). "Noise Estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement". In: *IEEE Signal Process. Lett.* 9.1, pp. 12–15.

Corr, J., K. Thompson, S. Weiss, J. G. McWhirter, S. Redif, and I. K. Proudler (2014). "Multiple Shift Maximum Element Sequential Matrix Diagonalisation for Para-Hermitian Matrices". In: *Proc. IEEE/SP Workshop on Statistical Signal Process.* Australia, pp. 844–848.

Doclo, S. and M. Moonen (Sept. 2002). "GSVD-Based Optimal Filtering for Single and Multimicrophone Speech Enhancement". In: *IEEE Trans. Signal Process.* 50.9, pp. 2230–2244.

Ephraim, Y. and D. Malah (Dec. 1984). "Speech Enhancement Using a Minimum-Mean Square Error Short-Time Spectral Amplitude Estimator". In: *IEEE Trans. Acoust., Speech, Signal Process.* 32.6, pp. 1109–1121.

Ephraim, Y. and H. L. Van Trees (July 1995). "A Signal Subspace Approach for Speech Enhancement". In: *IEEE Trans. Speech Audio Process.* 3.4, pp. 251–266.

Gannot, S., D. Burshtein, and E. Weinstein (Aug. 2001). "Signal Enhancement Using Beamforming and Nonstationarity with Applications to Speech". In: *IEEE Trans. Signal Process.* 49.8, pp. 1614–1626.

Hu, Y. and P. C. Loizou (July 2002). "A Subspace Approach for Enhancing Speech Corrupted by Colored Noise". In: *IEEE Signal Process. Lett.* 9.7, pp. 204–206.

— (Jan. 2008). "Evalution of Objective Quality Measures for Speech Enhancement". In: 16.1.

References

Kuklasiński, A., S. Doclo, S. H. Jensen, and J. Jensen (Sept. 2016). "Maximum Likelihood PSD Estimation for Speech Enhancement in Reverberation and Noise". In: *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 24.9, pp. 1599–1612.

Markovich, S., S. Gannot, and I. Cohen (Aug. 2009). "Multichannel Eigenspace Beamforming in a Reverberant Noisy Environment with Multiple Interfering Speech Signals". In: *IEEE Trans. Audio, Speech, Lang. Process.* 17.6, pp. 1071–1086.

McWhirter, J. G., P. D. Baxter, T. Cooper, S. Redif, and J. Foster (May 2007). "An EVD Algorithm for Para-Hermitian Polynomial Matrices". In: *IEEE Trans. Signal Process.* 55.5, pp. 2158–2169.

Neo, V. W. and P. A. Naylor (May 2019). "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).*

Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for End-to-End Speech Quality Assessment of Narrowband Telephone Networks and Speech Codecs (Nov. 2003). Recommendation P.862. Intl. Telecommun. Union (ITU-T).

Redif, S., S. Weiss, and J. G. McWhirter (2011). "An Approximate Polynomial Matrix Eigenvalue Decomposition Algorithm for Para-Hermitian Matrices". In: *Proc. Intl. Symp. on Signal Process. and Inform. Technology (ISSPIT).* Bilbao, Spain, pp. 421–425.

— (Jan. 2015). "Sequential Matrix Diagonalisation Algorithms for Polynomial EVD of Para-Hermitian Matrices". In: *IEEE Transactions Signal Processing* 63.1, pp. 81–89.

References

Redif, S., S. Weiss, and J. G. McWhirter (May 2017). "Relevance of Polynomial Matrix Decompositions to Broadband Blind Signal Separation". In: *Signal Processing* 134, pp. 76–86.

Taal, C. H., R. C. Hendriks, R. Heusdens, and J. Jensen (Sept. 2011). "An Algorithm for Intelligibility Prediction of Time-Frequency Weighted Noisy Speech". In: *IEEE Transactions on Audio, Speech, and Language Processing* 19.7, pp. 2125–2136.

Wang, Z., J. G. McWhirter, J. Corr, and S. Weiss (2015). "Multiple Shift Second Order Sequential Best Rotation Algorithm for Polynomial Matrix EVD". In: *Proc. European Signal Process. Conf. (EUSIPCO)*, pp. 844–848.

Weiss, S., S. Bendoukha, A. Alzin, F. K. Coutts, I. K. Proudler, and J. Chambers (2015). "MVDR Broadband Beamforming Using Polynomial Matrix Techniques". In: *Proc. European Signal Process. Conf. (EUSIPCO)*, pp. 839–843.

Weiss, S., N. J. Goddard, S. Somasundaram, I. K. Proudler, and P. A. Naylor (2017). "Identification of Broadband Source-Array Responses from Sensor Second Order Statistics". In: *Sensor Signal Process. for Defence Conf. (SSPD)*.

Thank you

Monte Carlo: Babble Noise

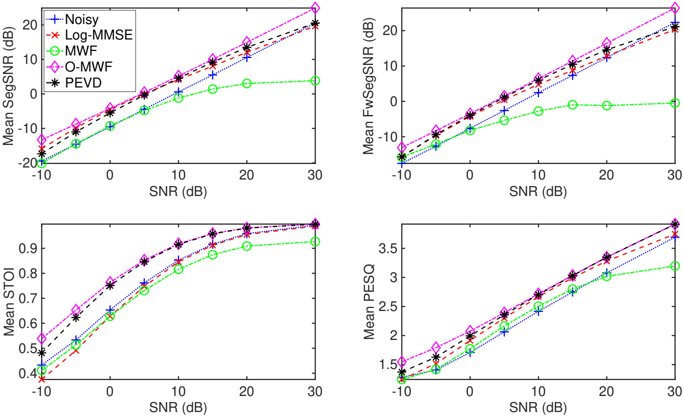


Figure Mean of the results for babble noise involving 150 trials.