Speech Dereverberation of a Polynomial Matrix Eigenvalue Decomposition Subspace Approach

Imperial College London

Speech and Audio Processing Lab

UNIVERSITY OF Southampton

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

EUSIPCO 2020

Outline

Introduction

Motivation for Speech Enhancement

Speech enhancement is important for many applications:

- Hearing aids
- Telecommunications
- Automatic speech recognition (ASR) systems
- Voice-controlled home systems

Main causes of speech degradation:

- Background noise
- Reverberation

Challenge: No prior information of target speech or acoustic environment
$\Rightarrow$ Need for blind and unsupervised approaches

Existing Approaches for Speech Enhancement

- 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, inadequate for speech

- Other methods typically use STFT to process [Cohen2002; Ephraim1984; Gannot 2001; Markovich2009]
  - Use DFT to divide broadband signal into multiple narrowband signals

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

PEVD Approach for Speech Enhancement

- Polynomial Matrices and Polynomial Eigenvalue Decomposition (PEVD)
  - Simultaneously capture correlations across space, time and frequency
  - Impose spatial decorrelation over a range of time shifts
  - No phase discontinuity

- PEVD-based Speech Enhancement [Neo2019a; Neo2020]
  - Effective for noise reduction
  - Performance approaches the Oracle Multichannel Wiener Filter (OMWF)
  - No noticeable artifacts

This Talk: Speech Dereverberation Performance

Background

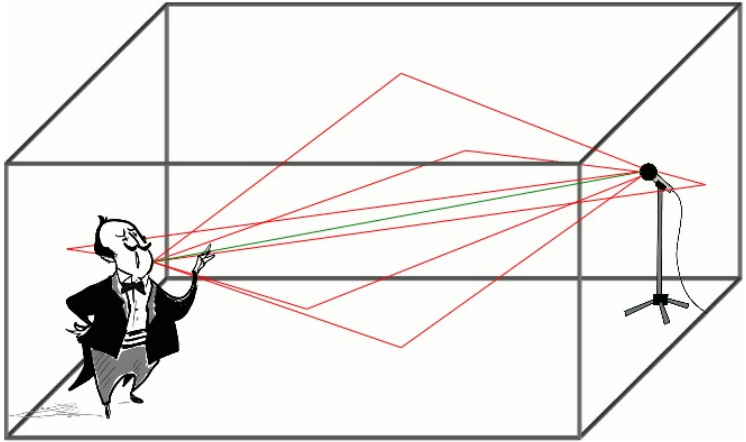
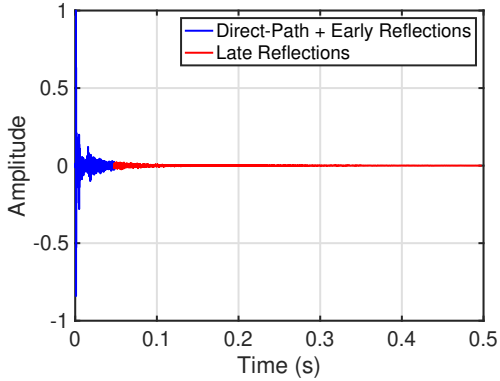


Figure is taken from the DREAMS project on the SAP website.

The m-th channel modelled as a FIR filter: $\mathbf{h}_m = \mathbf{h}_{m,dp} + \mathbf{h}_{m,er} + \mathbf{h}_{m,lr}$



An example of a room impulse response.

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

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

where

- $\tilde{s}_m(n) = (\mathbf{h}_{m,dp}^T + \mathbf{h}_{m,er}^T)\mathbf{s}_0(n)$ is the speech component,
- $\tilde{v}_m(n) = \mathbf{h}_{m,lr}^T \mathbf{s}_0(n) + v_m(n)$ is the noise component.
- $\mathbf{s}_0(n)$ is the anechoic speech signal,
- $v_m(n)$ is the noise signal at the m-th sensor.

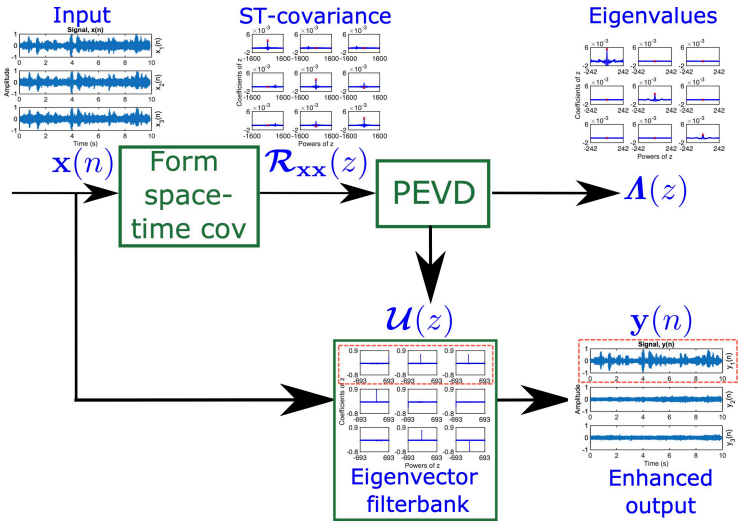
The data vector collected from M sensors is

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

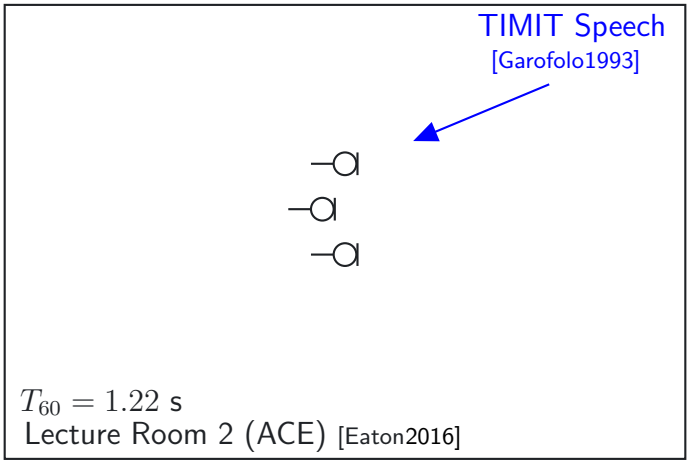
Speech Enhancement Using PEVD

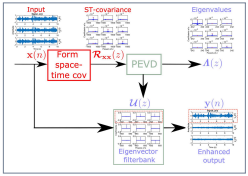
Speech Enhancement Using PEVD [Neo2020]

Imperial College London



Reverberant Speech (No Noise)





Assuming stationarity, the 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.

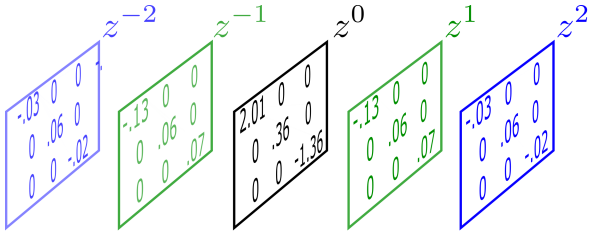
Z-transform of $\mathbf{R}_{\mathbf{xx}}(\tau)$ is a para-Hermitian polynomial matrix

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

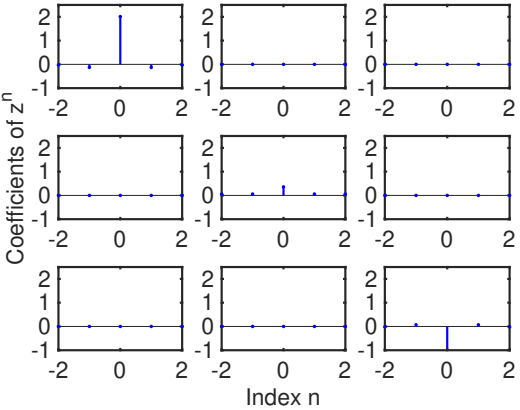
where $\mathbf{R}_{\mathbf{xx}}(\tau) \approx 0$ for $|\tau| > W$, calligraphic $\boldsymbol{\mathcal{R}}$ for polynomial matrices and regular $\mathbf{R}$ for matrices.

Polynomial Matrix Representation

Equivalently, expressed as:



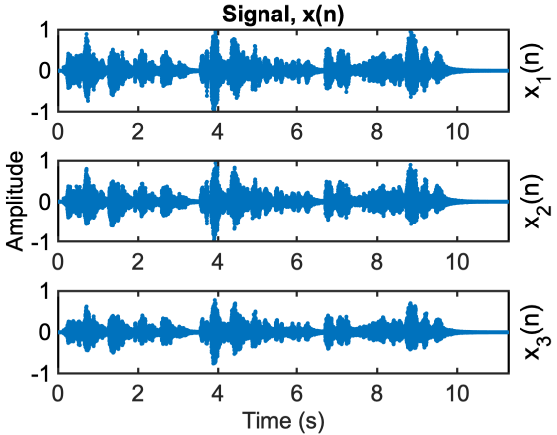
Polynomial with matrix coefficients.



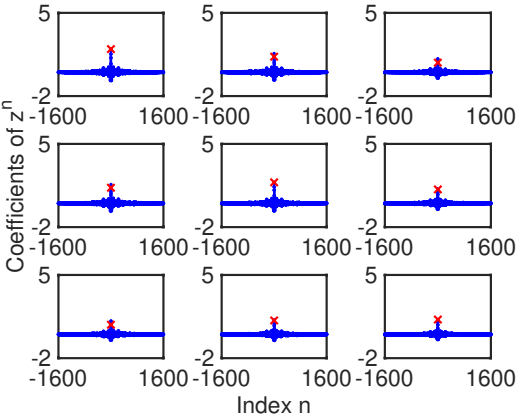
Matrix with polynomial elements.

Example: Polynomial Matrix from ST-Covariance

Imperial College London



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

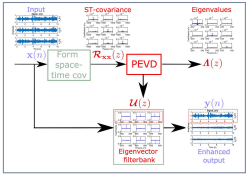


Polynomial matrix, $\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z)$.

Polynomial Matrix Eigenvalue Decomposition

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

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



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

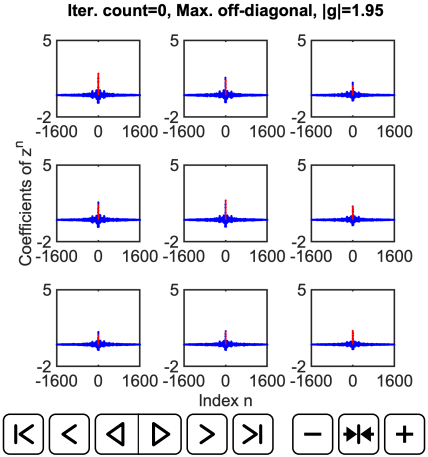
Since $\tilde{\mathbf{s}}(n)$ and $\tilde{\mathbf{v}}(n)$ are uncorrelated [Naylor2010a]

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

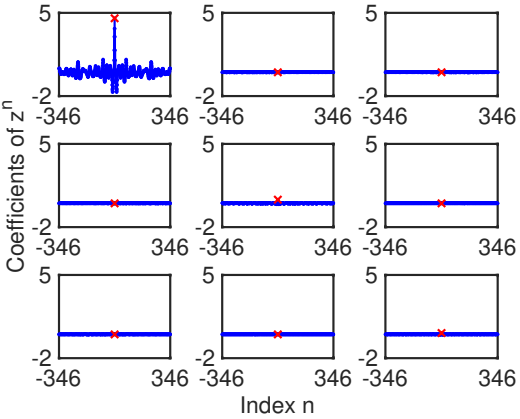
with orthogonal signal, $\{\cdot\}_{\tilde{s}}$ and noise subspaces, $\{\cdot\}_{\tilde{v}}$.

Example: PEVD Algorithm

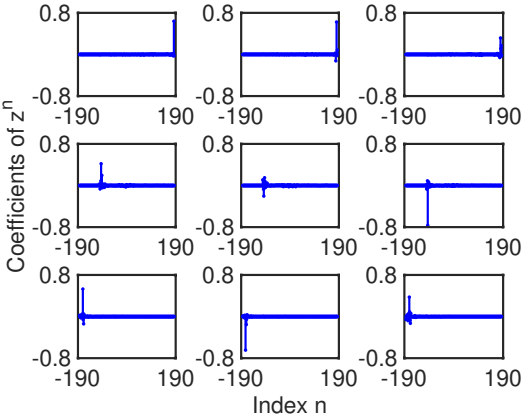
Algorithm converges when $|g| < 1.68 \times 10^{-2}$



Example: PEVD Algorithm Outputs



Eigenvalue polynomial matrix, $\boldsymbol{\Lambda}(z)$.

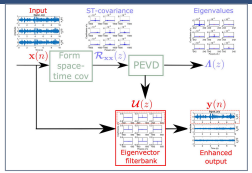


Eigenvector polynomial matrix, $\boldsymbol{\mathcal{U}}(z)$.

PEVD Algorithms

PEVD algorithms include:

- Second-order Sequential Best Rotation (SBR2) [McWhirter2007]
- Sequential Matrix Diagonalization (SMD) [Redif2015]
- Householder-like PEVD [Redif2011]
- Tridiagonal PEVD [Neo2019b]
- Multiple-shift SBR2/SMD [Wang2015; Corr2014]



$\boldsymbol{\mathcal{U}}(z)$ is a filterbank for $\mathbf{x}(z)$ which produces outputs,

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

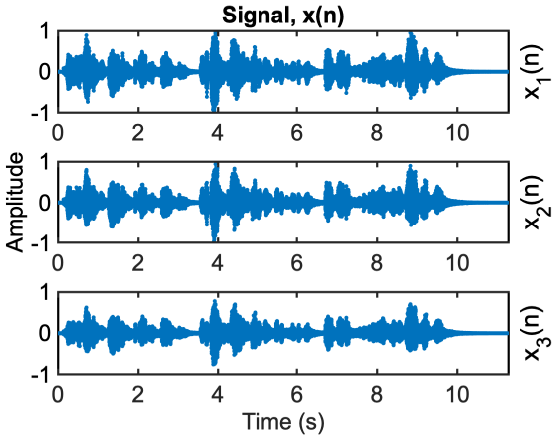
that are strongly decorrelated.

The output in the first channel, $y_1(z)$, is the enhanced and dereverberated speech signal with space-time covariance matrix

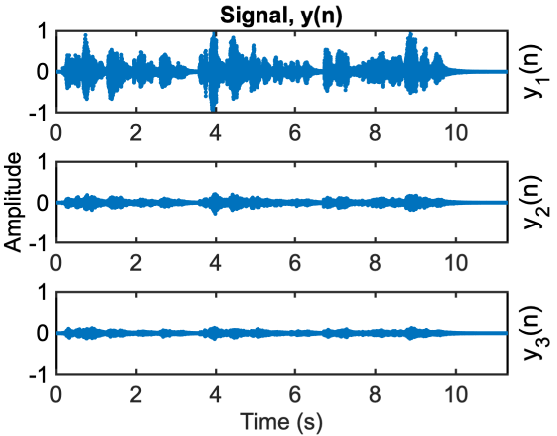
$$\boldsymbol{\mathcal{R}}_{y_1 y_1} = \left[\begin{array}{c|c} \boldsymbol{\mathcal{U}}^P_{\tilde{s}}(z) & \mathbf{0} \end{array} \right] \left[\begin{array}{c|c} \boldsymbol{\Lambda}_{\tilde{s}}(z) & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} \end{array} \right] \left[\begin{array}{c} \boldsymbol{\mathcal{U}}_{\tilde{s}}(z) \\ \hline \mathbf{0} \end{array} \right].$$

Example: PEVD-based Enhancement Outputs

Imperial College London



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



Enhanced signals, $\mathbf{y}(z)$.

Comparative Results

Speech Dereverberation Evaluation

Comparative algorithms:

1. Generalized weighted prediction error (GWPE) [Yoshioka2012]
2. Multichannel Subspace (MCSUB) - Uses an EVD [Huang2008]
3. Oracle-MWF (OMWF) - Given clean speech [Doclo2002]

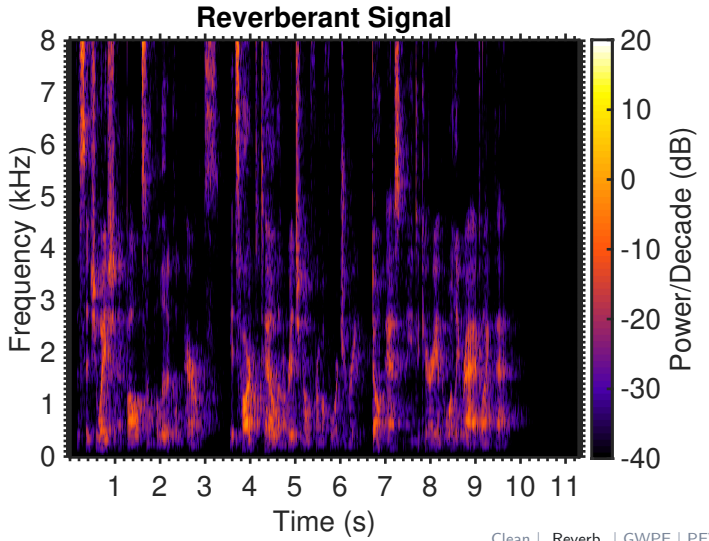
Dereverberation measures:

- Normalized Signal to Reverberant Ratio (NSRR) [Naylor2010b]
- Bark Spectral Distortion (BSD)

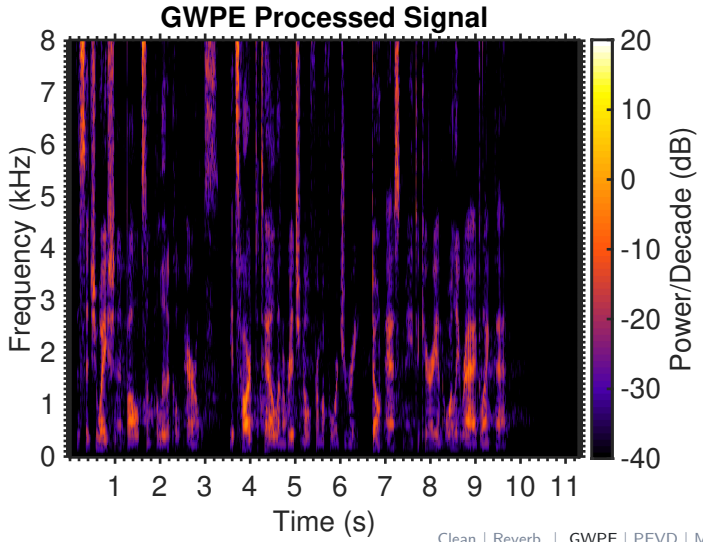
Noise reduction and speech quality measures:

- Frequency-weighted Segmental SNR (FwSegSNR) [Hu2006]
- Perceptual Evaluation of Speech Quality (PESQ) [ITU-T P.862]

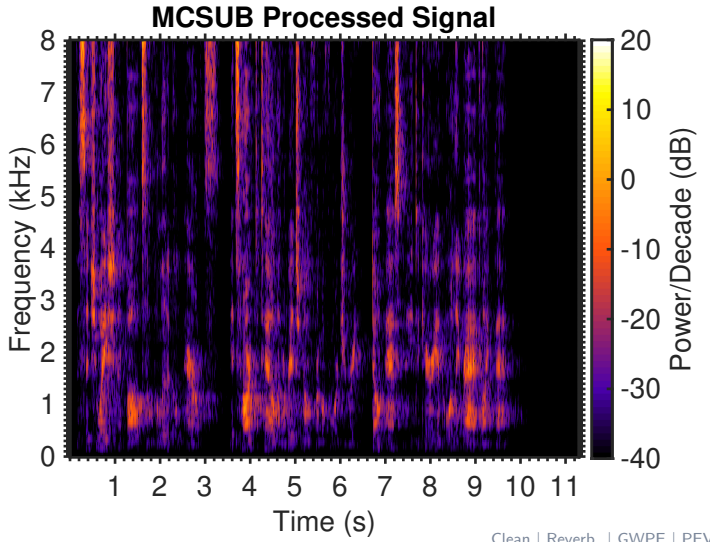
Reverberant Speech Spectrogram (No Noise)



GWPE Processed Spectrogram

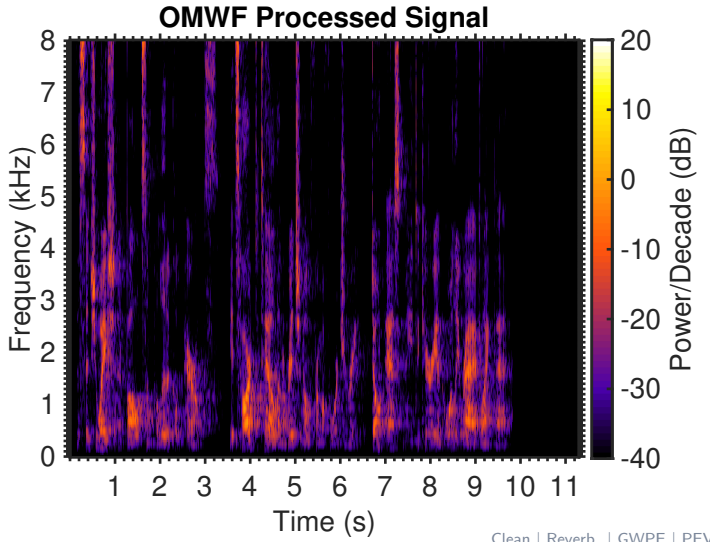


MCSUB Processed Spectrogram

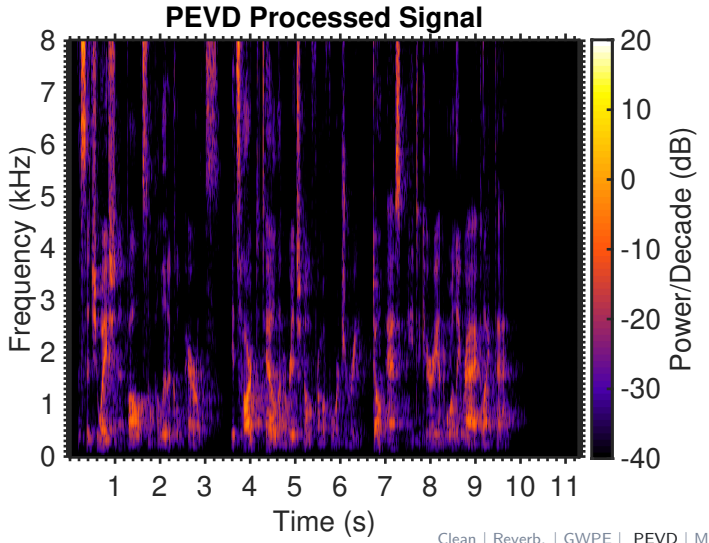


Clean | Reverb. | GWPE | PEVD | MCSUB | OMWF | Table

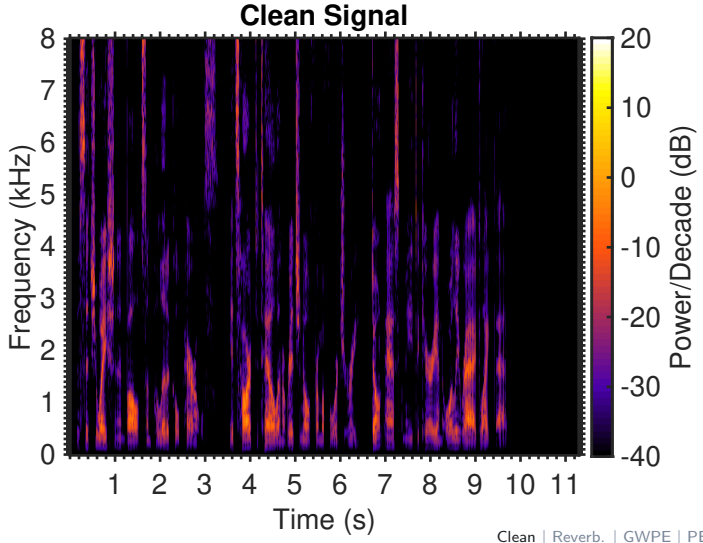
OMWF Processed Spectrogram



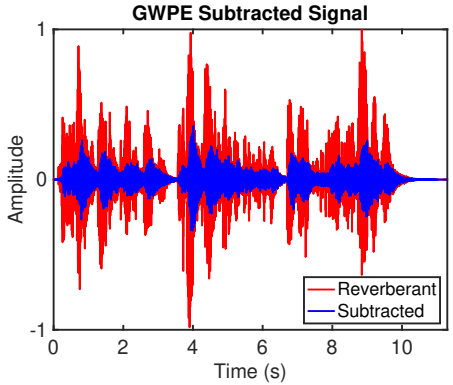
PEVD Processed Spectrogram

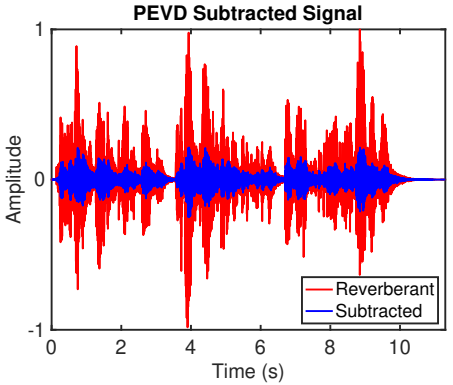


Clean Speech Spectrogram



Comparison of Subtracted Signals

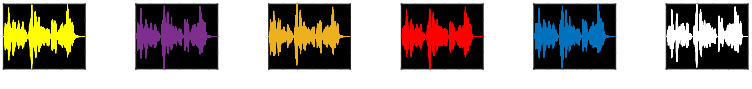


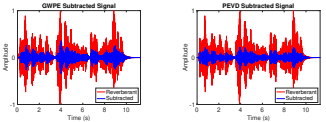


Dereverberation Performance (No Noise)

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

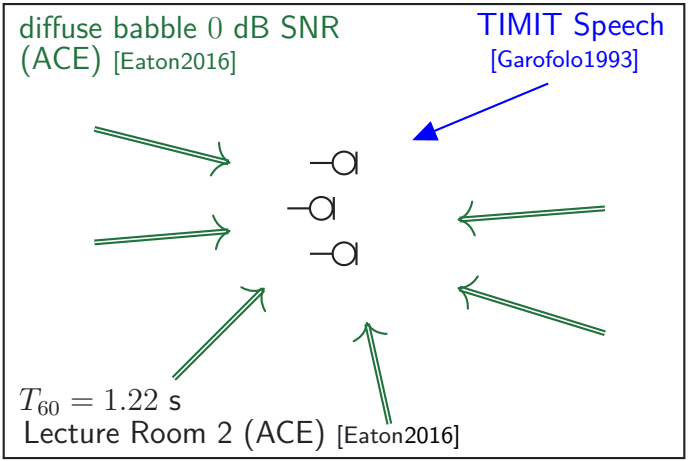
Noisy GWPE MCSUB OMWF PEVD Clean





Clean | Reverb. | GWPE | PEVD | MCSUB | OMWF | Table

Reverberant Speech in Noise



Dereverberation Performance (0 dB Babble Noise)

Algorithm	ΔNSRR	ΔBSD	ΔFwSegSNR	ΔPESQ
GWPE	0.22 dB	-0.12 dB	0.28 dB	0.05
MCSUB	-3.29 dB	0.21 dB	0.64 dB	0.21
OMWF	0.26 dB	-0.25 dB	3.12 dB	**0.29**
PEVD	**5.38 dB**	**-0.52 dB**	**3.56 dB**	0.20

Noisy GWPE MCSUB OMWF PEVD Clean

Conclusion

Conclusion

- Polynomial matrices and PEVD as a tool for processing broadband multichannel signals
- PEVD-based speech enhancement algorithm is effective for dereverberation
  - Performs well even in the presence of noise
  - No noticeable artifacts
  - Completely blind and unsupervised

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 Processing*, 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.

Eaton, J., N. D. Gaubitch, A. H. Moore, and P. A. Naylor (Oct. 2016). "Estimation of Room Acoustic Parameters: The ACE Challenge". In: *IEEE/ACM Trans. Audio, Speech, Lang. Process.* 24.10, pp. 1681–1693.

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.

References

Garofolo, J. S., L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, and V. Zue (1993). *TIMIT Acoustic-Phonetic Continuous Speech Corpus*. Corpus. Philadelphia: Linguistic Data Consortium (LDC).

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.

— (2006). "Evaluation of Objective Measures for Speech Enhancement". In: *Proc. Conf. of Intl. Speech Commun. Assoc. (INTERSPEECH)*, pp. 1447–1450.

Huang, Y., J. Benesty, and J. Chen (July 2008). "Analysis and Comparison of Multichannel Noise Reduction Methods in a Common Framework". In: *IEEE Trans. Audio, Speech, Lang. Process.* 16.5, pp. 957–968.

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.

Naylor, P. A. and N. D. Gaubitch, eds. (2010a). *Speech Dereverberation*. Springer-Verlag.

Naylor, P. A., N. D. Gaubitch, and E. A. P. Habets (2010b). "Signal-Based Performance Evaluation of Dereverberation Algorithms". In: *J. of Elect. and Comput. Eng.* 2010, pp. 1–5.

References

Neo, V. W., C. Evers, and P. A. Naylor (2019a). "Speech Enhancement Using Polynomial Eigenvalue Decomposition". In: *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 125–129.

— (2020). "PEVD-Based Speech Enhancement in Reverberant Environments". In: *Proc. IEEE Intl. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, pp. 186–190.

Neo, V. W. and P. A. Naylor (2019b). "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)*, pp. 8043–8047.

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)*, pp. 421–425.

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

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.

Yoshioka, T. and T. Nakatani (Dec. 2012). "Generalization of Multi-Channel Linear Prediction Methods for Blind MIMO Impulse Response Shortening". In: *IEEE Trans. Audio, Speech, Lang. Process.* 20.10, pp. 2707–2720.

Thank you

Listening Examples: https://www.commsp.ee.ic.ac.uk/~sap/pevddrb
Webpage: https://www.commsp.ee.ic.ac.uk/~sap/vincent-w-neo