PEVD-based Speech Enhancement in Reverberant Environments

Imperial College London

Speech and Audio Processing Lab

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

7 May 2020

Outline

1. Introduction
   - Motivation for PEVD
2. Background
   - Multichannel Signal Model
   - Polynomial Matrices
   - PEVD
3. Proposed Methodology
4. Experimental Results
5. Conclusion

Introduction

Motivation

- 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; Gannot 2001; Markovich2009]
  - Use DFT to divide broadband into multiple narrowband signals

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

Motivation for PEVD

- Polynomial Matrices and 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 Speech Enhancement [Neo2019a]
  - Effective for anechoic environments
  - Performance approaches the Oracle Multichannel Wiener Filter (OMWF)
  - No noticeable artifacts

This Talk: Speech Enhancement in Reverberant Environments

Background

Multichannel Signal Model

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

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

where

- $\mathbf{s}_0(n)$ is the anechoic speech signal,
- $\mathbf{h}_m$ is the channel modelled as an order J FIR filter,
- $v_m(n)$ is the noise signal at the m-th sensor.

The data vector collected from M sensors is

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

Space-time Covariance Polynomial Matrix

Assuming stationarity, the space-time covariance matrix is

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

where $(i,j)^{\mathrm{th}}$ 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 tensor and regular $\mathbf{R}$ 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 $\boldsymbol{\mathcal{R}}_{\mathbf{xx}}^{P}(z) = \boldsymbol{\mathcal{R}}_{\mathbf{xx}}^{H}(z^{-1})$.

$\boldsymbol{\mathcal{U}}(z)$ is a filterbank for $\mathbf{x}(z) \in \mathbb{C}^{M\times 1\times T}$ 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.

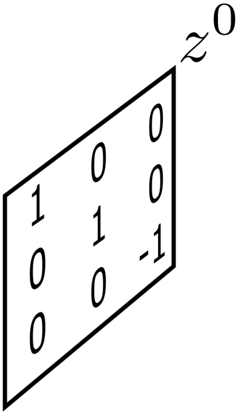
PEVD Algorithms

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]

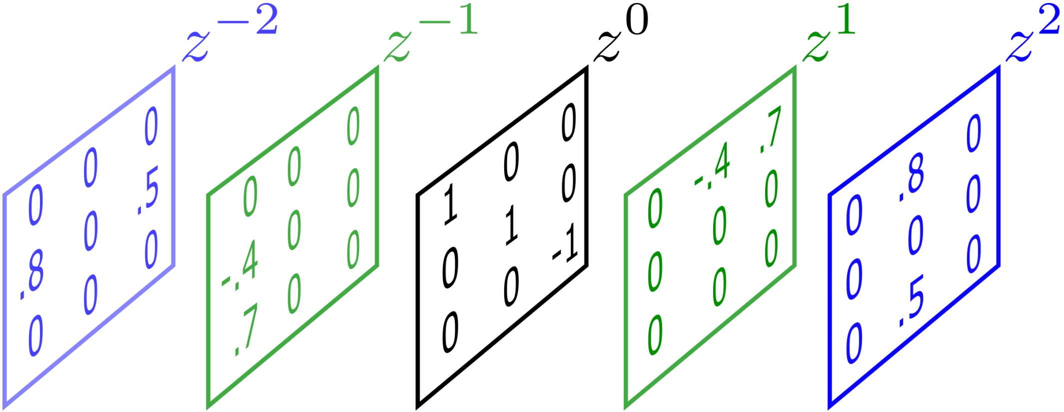
Example of a Polynomial Matrix

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



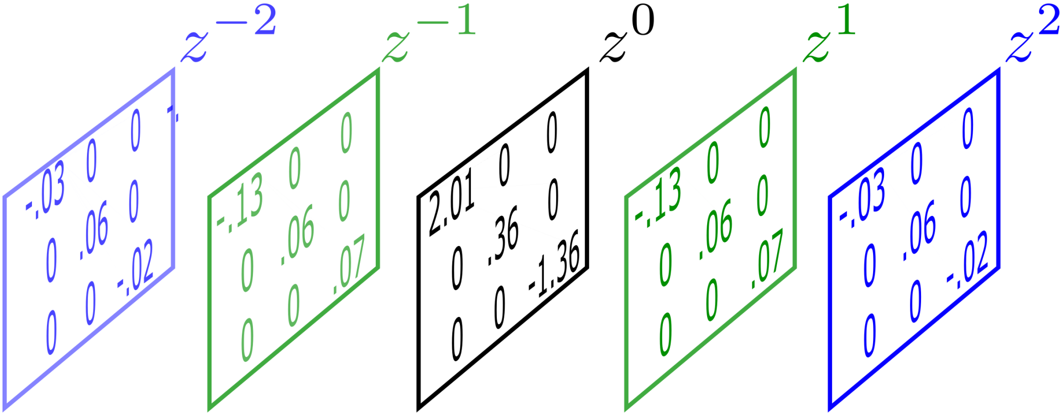
$\mathbf{R}_{\mathbf{xx}}(0)$: instantaneous (spatial) covariance matrix / coefficient of z^0.

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



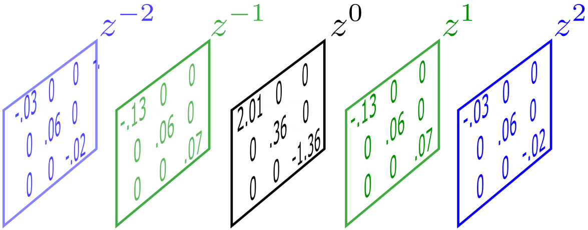
In this example, z^0 plane is diagonal but not at other planes.

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

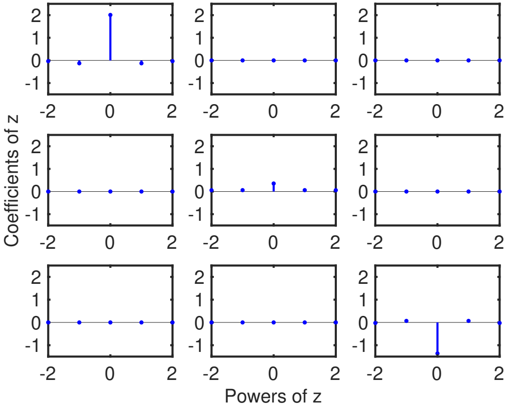


Alternate Representation of Example

Equivalently, expressed as:



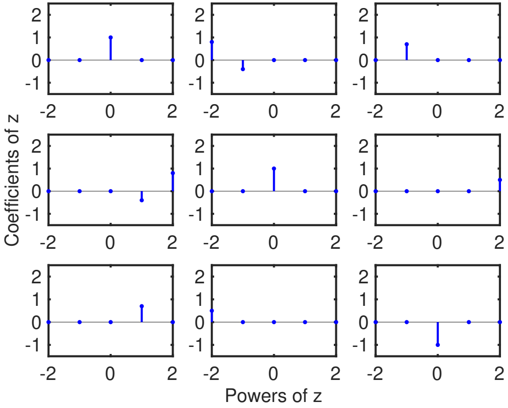
Polynomial with matrix coefficients.



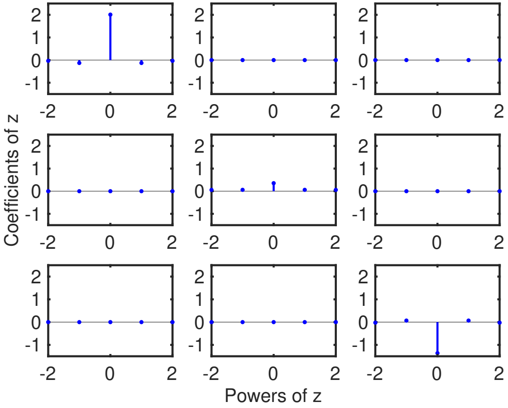
Matrix with polynomial elements.

Alternate Representation of Example

The same example can be represented as:



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

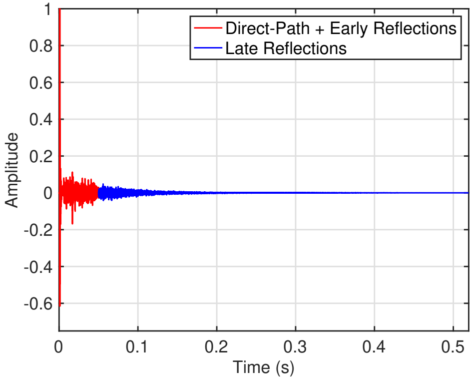


Diagonalized $\boldsymbol{\Lambda}(z)$.

Proposed Methodology

Reverberant Channel Model

The m-th channel: $\mathbf{h}_m = \tilde{\mathbf{h}}_{m,dp} + \tilde{\mathbf{h}}_{m,er} + \tilde{\mathbf{h}}_{m,lr}$



An example of a room impulse response.

Application:Speech Enhancement

Rewriting the signal model using the reverberant channel model gives

$$x_m(n) = \tilde{s}_m(n) + \tilde{v}_m(n)$$

where

- $\tilde{s}_m(n) = (\tilde{\mathbf{h}}^T_{m,dp} + \tilde{\mathbf{h}}^T_{m,er})\mathbf{s}_0(n)$ is the speech component,
- $\tilde{v}_m(n) = \tilde{\mathbf{h}}^T_{m,lr}\mathbf{s}_0(n) + v_m(n)$ is the noise component.

Goal: Obtain some enhanced version of speech using observations, $\mathbf{x}(n)$.

Enhancement targets:

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

Application: Speech Enhancement

Since $\tilde{\mathbf{s}}(n)$ and $\tilde{\mathbf{v}}(n)$ are uncorrelated [Naylor2010]

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

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

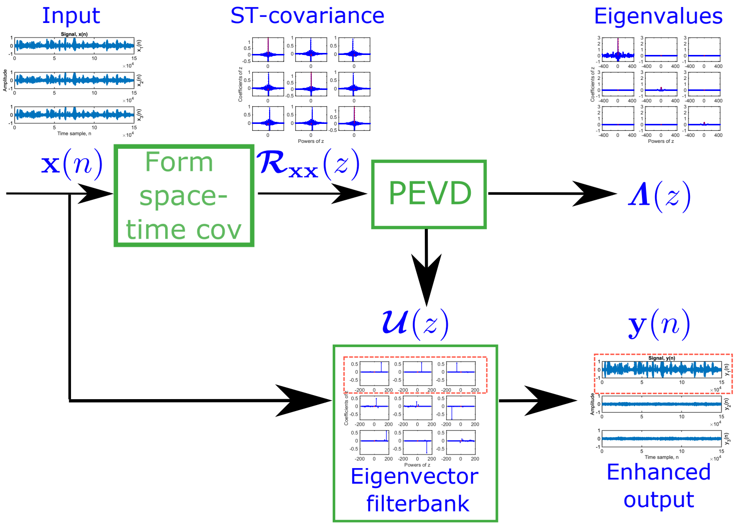
The output

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

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

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

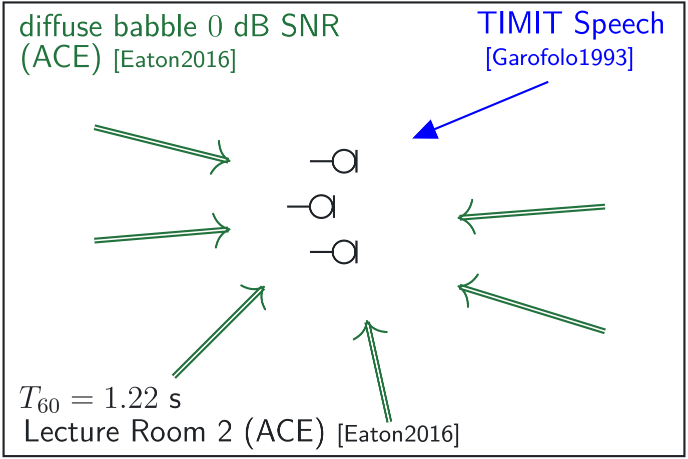
Speech Enhancement Using PEVD



Experimental Results

Experiment Setup

Reverberant Speech in Noise



Evaluation

Comparative algorithms:

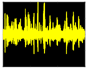
1. Coloured Noise Subspace (COLSUB) [Hu2002]
2. Log-Minimum Mean Square Error (Log-MMSE) [Ephraim1984]
3. Multichannel Subspace (MCSUB) [Huang2008]
4. Multichannel Wiener Filter (MWF) - Relative Transfer Function (RTF) and noise estimator [Kuklasiński2016]
5. Oracle-MWF (OMWF) - Given clean speech [Doclo2002]

Evaluation measures:
ΔSegSNR, ΔFwSegSNR, ΔSTOI, ΔPESQ

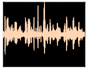
Comparison of Enhancement Algorithms

Algorithm	ΔSegSNR	ΔFwSegSNR	ΔSTOI	ΔPESQ
COLSUB	8.35 dB	7.67 dB	-0.018	-0.20
log-MMSE	3.67 dB	3.05 dB	-0.058	-0.12
MCSUB	-1.52 dB	-1.04 dB	-0.010	-0.03
MWF	1.06 dB	0.78 dB	-0.005	0.02
OMWF	0.57 dB	-0.44 dB	0.084	0.17
PEVD	2.96 dB	2.88 dB	0.078	0.11

Noisy

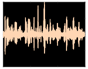


COLSUB

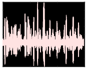


log-MMSE

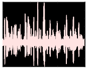
MCSUB



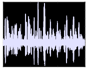
MWF



OMWF



PEVD



Clean

Conclusion

Conclusion

Imperial College London

- Polynomial matrices and PEVD as a tool for processing broadband multichannel signals
- Proposed a PEVD-based speech enhancement algorithm designed for a reverberant signal model
  - Exploits the lack of correlation between the speech and late reflections to provide further noise reduction
  - Incorporates the early reflections to further improve speech intelligibility and quality
  - No noticeable processing artifacts
  - No noise estimation is required

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.* 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.

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.

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

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.

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.

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

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).*

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

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.

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.

Thank you

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