Polynomial Matrix Eigenvalue Decomposition of Spherical Harmonics for Speech Enhancement

Imperial College London

Speech and Audio Processing Lab

University of Southampton

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

ICASSP 2021

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 Subspace 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

- PEVD-based speech enhancement [Neo2019a; Neo2020]
  - Use PEVD to impose spatial decorrelation over a range of time shifts
  - Effective for noise reduction and dereverberation
  - Robust for linear and arbitrary array geometries

$\Rightarrow$ Limitation: Complexity $\propto (\#\text{ of mics})^3$

This Talk: Spherical Microphone Array

Background

Multichannel Reverberant Signal Model

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

$$x_q(n) = \mathbf{h}_q^T \mathbf{s}_0(n) + v_q(n) = \tilde{s}_q(n) + \tilde{v}_q(n)$$

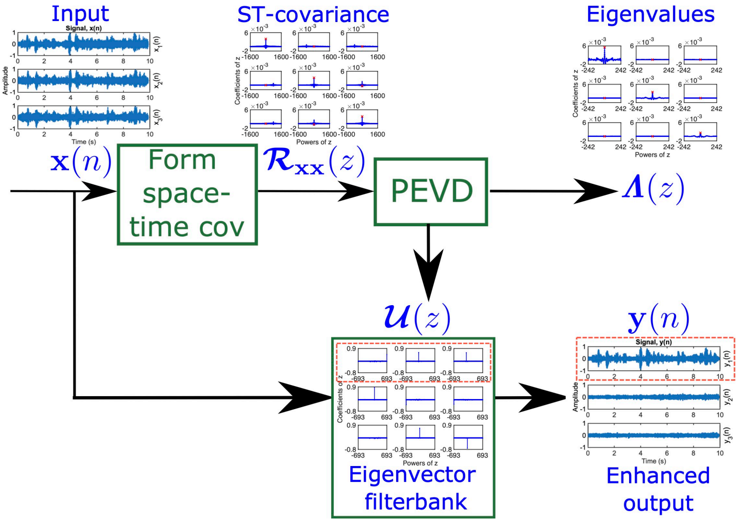
where

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

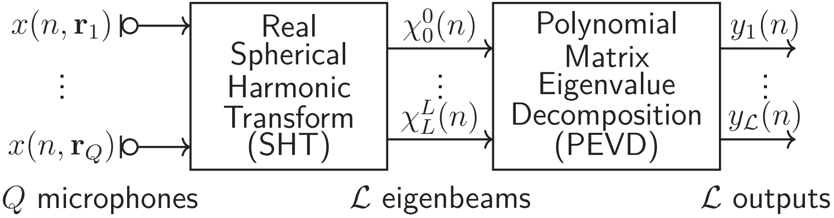
The data vector collected from Q sensors is

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

Speech Enhancement Using PEVD [Neo2020]



Sketch: PEVD of Spherical Harmonics (SH)



Spherical Harmonics Decomposition

The ℓ-th order, m-th degree eigenbeam signal, associated with the real SH basis function $R_\ell^m(\mathbf{r}_q)$ and quadrature sampling weight α_q, is

$$\chi_\ell^m(n) \approx \sum_{q=1}^{Q} \alpha_q x(n, \mathbf{r}_q) R_\ell^m(\mathbf{r}_q).$$

Recovery of each microphone signal uses a weighted sum of the SH

$$x(n, \mathbf{r}_q) = \sum_{\ell=1}^{L} \sum_{m=-\ell}^{\ell} \chi_\ell^m(n) R_\ell^m(\mathbf{r}_q)$$

and alias-free spatial reconstruction requires $Q \geq (L+1)^2$ where L is the maximum SH order of the sound field and $\mathcal{L} \triangleq (L+1)^2$ eigenbeams.

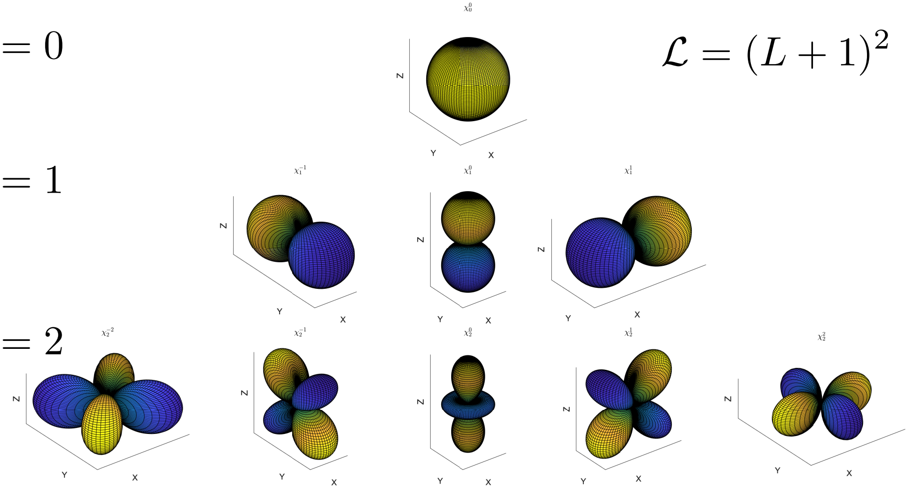
Spherical Harmonics Decomposition

$\ell = 0$

$\ell = 1$

$\ell = 2$

$$\mathcal{L} = (L+1)^2$$



$m=-2$ $\quad m=-1$ $\quad m=0$ $\quad m=1$ $\quad m=2$

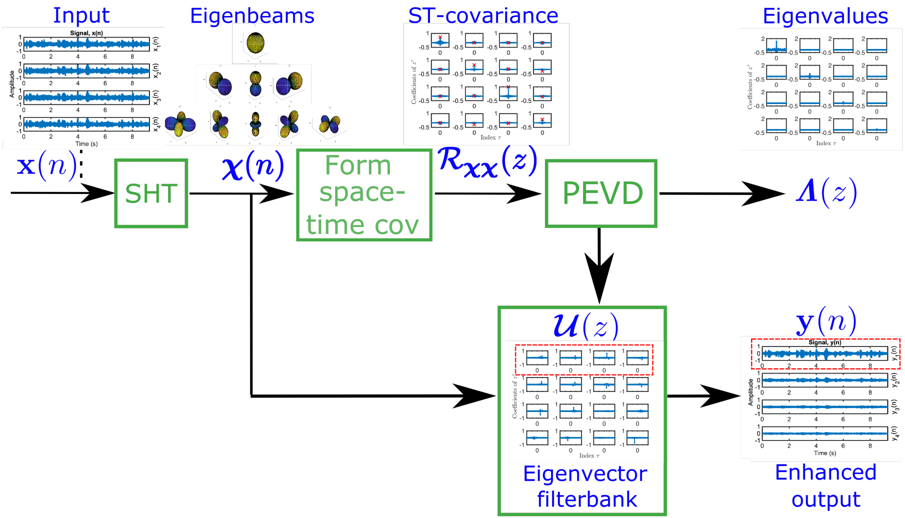
Complexity Analysis of a Single Example

Imperial College London

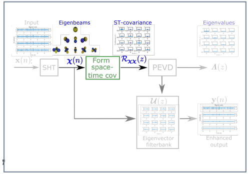
SH Order, L	0	1	2	3	4
# Eigenbeams $\mathcal{L}$	1	4	9	16	25
Approx. Error, $\varepsilon(\%)$	3.82	3.77	3.45	2.74	1.38
Complexity Factor, β	-	0.002	0.022	0.125	0.477

*$\beta = (\frac{\mathcal{L}}{Q})^3$, where $Q = 32$ microphones.

Enhancement Algorithm: PEVD of Eigenbeams



Space-time Covariance Polynomial Matrix



Assuming stationarity, the space-time covariance matrix is

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

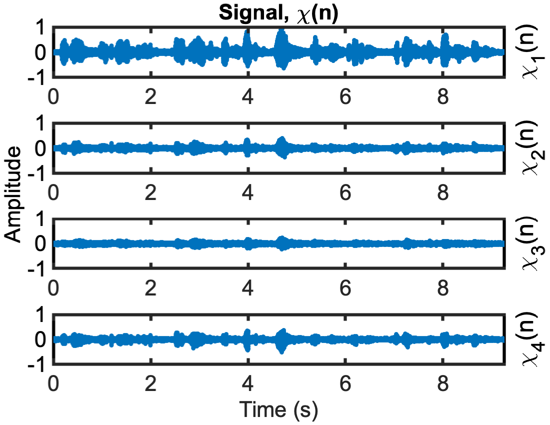
where $(i,j)^{\text{th}}$ element is the correlation function $r_{ij}(\tau) = \mathbb{E}[\chi_i(n)\chi_j^*(n-\tau)]$, τ is the time-shift and $\boldsymbol{\chi} = [\chi_0^0, \chi_1^{-1}, \chi_1^0, \ldots, \chi_L^L]^T$ is arranged in ascending order and degree.

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

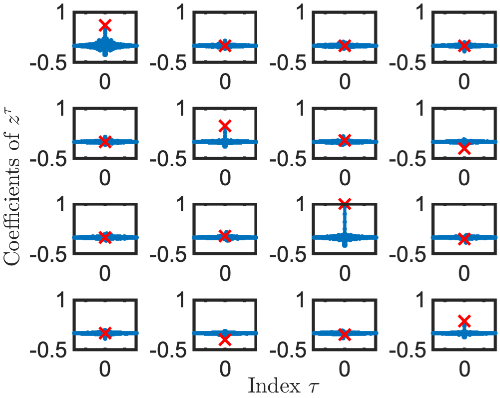
$$\boldsymbol{\mathcal{R}}_{\chi\chi}(z) = \sum_{\tau=-W}^{W} \mathbf{R}_{\chi\chi}(\tau) z^{-\tau},$$

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

Example: Polynomial Matrix from ST-Covariance



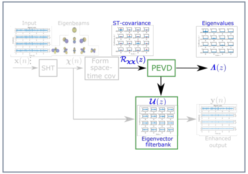
Eigenbeam signals $\boldsymbol{\chi}(n)$.



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

Polynomial Matrix Eigenvalue Decomposition

Imperial College London



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

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

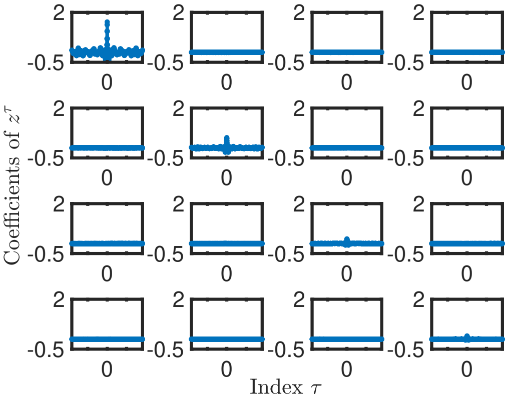
where $\boldsymbol{\Lambda}(z)$,$\boldsymbol{\mathcal{U}}(z)$ contain the eigenvalues and eigenvectors and $\boldsymbol{\mathcal{R}}^{P}_{\boldsymbol{\chi}\boldsymbol{\chi}}(z) = \boldsymbol{\mathcal{R}}^{H}_{\boldsymbol{\chi}\boldsymbol{\chi}}(z^{-1})$.

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

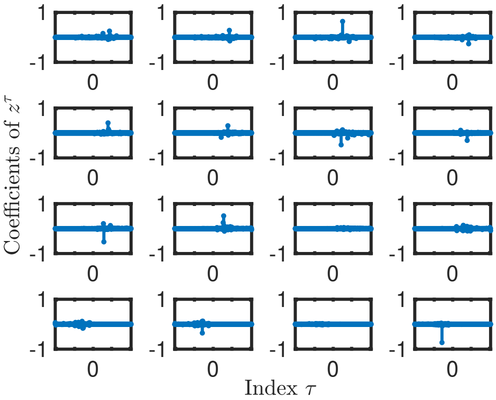
$$\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z) = \left[\ \boldsymbol{\mathcal{U}}^{P}_{\tilde{s}}(z) \;\middle|\; \boldsymbol{\mathcal{U}}^{P}_{\tilde{v}}(z)\ \right] \left[\begin{array}{c|c} \boldsymbol{\Lambda}_{\tilde{s}}(z) & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{\Lambda}_{\tilde{v}}(z) \end{array}\right]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

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

Example: PEVD Algorithm Outputs



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

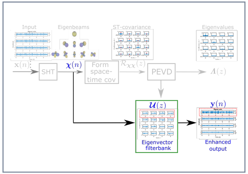


Eigenvector polynomial matrix, $\boldsymbol{\mathcal{U}}(z)$.

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 $\boldsymbol{\chi}(z)$ which produces outputs,

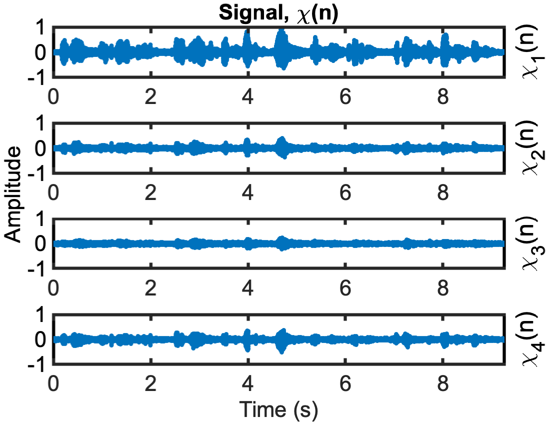
$$\mathbf{y}(z) = \boldsymbol{\mathcal{U}}(z)\boldsymbol{\chi}(z) \implies \boldsymbol{\mathcal{R}}_{\mathbf{yy}}(z) \approx \boldsymbol{\Lambda}(z),$$

that are strongly decorrelated.

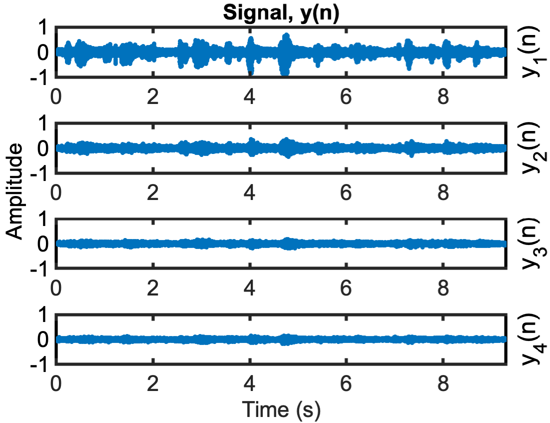
First channel output, $y_1(z)$, is the enhanced speech with ST-covariance

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

Example: Filterbank Output

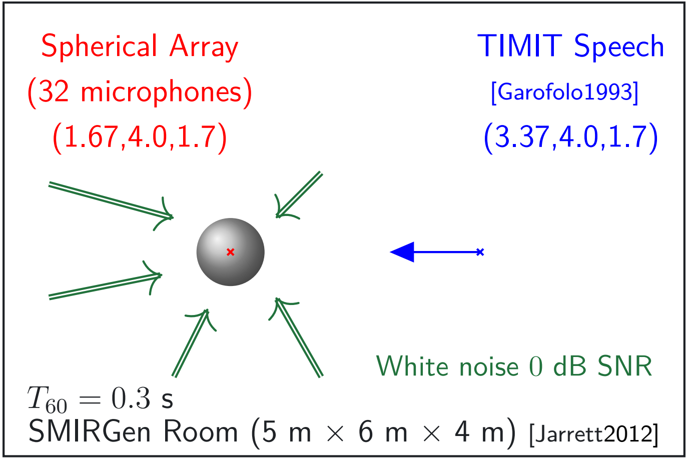


Eigenbeam signals $\boldsymbol{\chi}(n)$.



Enhanced signals, $y(n)$.

Experiment Setup: Reverberant 0 dB White Noise



Comparative Results

Speech Enhancement Evaluation

Comparative algorithms:

1. Eigenbeams χ_0^0, χ_1^1 [Rafaely 2015; Jarrett2017]
2. KLT$\{\chi_0^0\}$ - Uses an EVD on eigenbeam [Ephraim1995]
3. Raw PEVD - 32 microphone signals for PEVD [Neo2020]
4. PEVD L1, L2 - Use SH order 1, 2 eigenbeams $\Rightarrow$ 4, 9 signals for PEVD

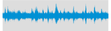
Enhancement measures:

- Frequency-weighted Segmental SNR (FwSegSNR) [Hu2006]
- Short-Time Objective Intelligibility (STOI) [Taal2011]
- Perceptual Evaluation of Speech Quality (PESQ) [ITU-T P.862]
- Bark Spectral Distortion (BSD) [Naylor2010]

Speech Enhancement Performance

Algorithm	ΔFwSegSNR	ΔSTOI	ΔPESQ	ΔBSD
χ_0^0	4.86 dB	0.055	0.42	-1.53 dB
KLT$\{\chi_0^0\}$	5.56 dB	0.054	**0.51**	-1.65 dB
χ_1^1	0.89 dB	0.122	0.44	-0.65 dB
PEVD L1	5.72 dB	0.110	0.47	-1.68 dB
PEVD L2	**5.92 dB**	**0.125**	**0.51**	**-1.71 dB**
RAW PEVD	5.59 dB	0.119	0.49	-1.62 dB

Noisy



χ_0^0

KLT$\{\chi_0^0\}$

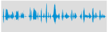
χ_1^1

PEVD L1

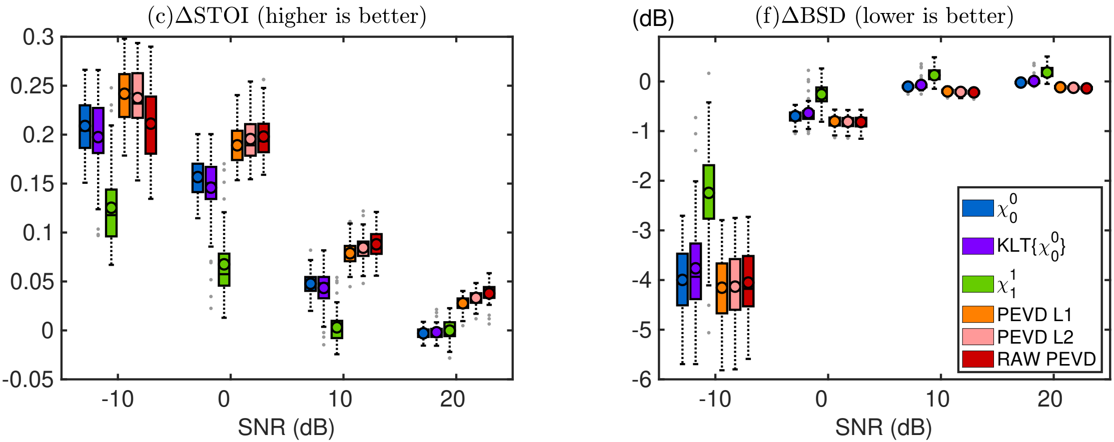
PEVD L2

RAW PEVD

Clean



ACE: Lecture Room 2, Babble Noise [Eaton2016]



Conclusion

Conclusion

- PEVD of eigenbeams remains effective for speech enhancement in noisy, reverberant environments
  - Performs almost identically, and sometimes even better, than Raw PEVD
  - Complexity factor is fraction of Raw PEVD: 0.002 to 0.477 times

- Robust even when eigenbeams are not steered towards the speaker
  - 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.
- 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.
- 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 H. L. Van Trees (July 1995). "A Signal Subspace Approach for Speech Enhancement". In: *IEEE Trans. Speech Audio Process.* 3.4, pp. 251–266.
- 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, USA: 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.
- Jarrett, D. P., E. A. P. Habets, and P. A. Naylor (2017). *Theory and Applications of Spherical Microphone Array Processing*. Springer Topics in Signal Processing.

References

- Jarrett, D. P., E. A. P. Habets, M. R. P. Thomas, and P. A. Naylor (Sept. 2012). "Rigid Sphere Room Impulse Response Simulation: Algorithm and Applications". In: *J. Acoust. Soc. Am.* 132.3, pp. 1462–1472.
- 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.
- 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).
- Rafaely, B. (2015). *Fundamentals of Spherical Array Processing*. Springer Topics in Signal Processing.

References

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.

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 Trans. Audio, Speech, Lang. Process.* 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://vwn09.github.io/shd-pevd/
Webpage: https://vwn09.github.io