Polynomial Matrix Eigenvalue Decomposition Based Source Separation Using Informed Spherical Microphone Arrays

Imperial College London

Speech and Audio Processing Lab

University of Southampton

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

WASPAA 2021

Introduction

Motivation for Source Separation

Audio source separation is important for many applications:

- Speech enhancement in hearing aids, telecommunications
- 3D sound rendering
- Robot audition

Main challenges:

- Interfering sources
- Background noise
- Reverberation

Existing Approaches to Source Separation

- Microphone Array Processing
  - Spatial filtering to isolate sources
- Independent Component Analysis (ICA)
  - Decomposes into independent non-Gaussian signals
- Non-Negative Matrix Factorization (NNMF)
  - Decomposes into spectral patterns and temporal activations

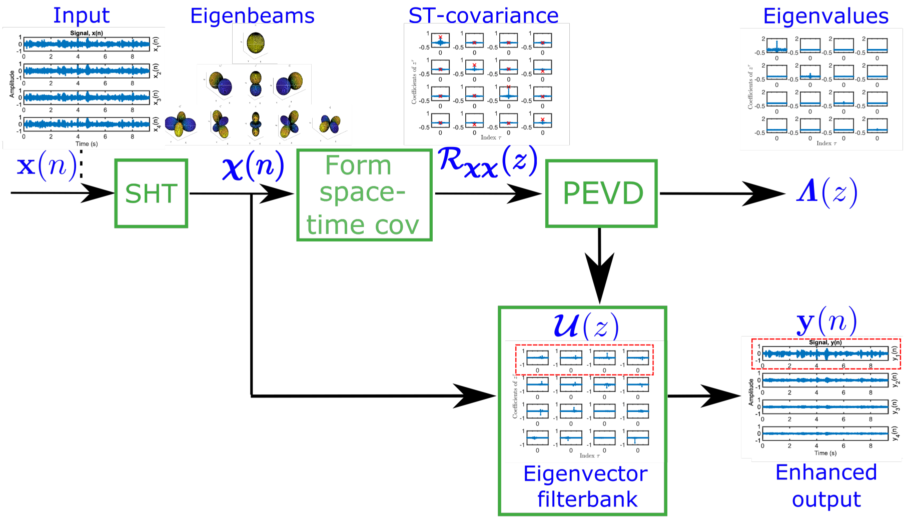
$\Rightarrow$ Performs separation but may introduce processing artefacts

- PEVD-based Speech Enhancement of Eigenbeams (Spherical Arrays)
  - Provides good speech enhancement without introducing artefacts

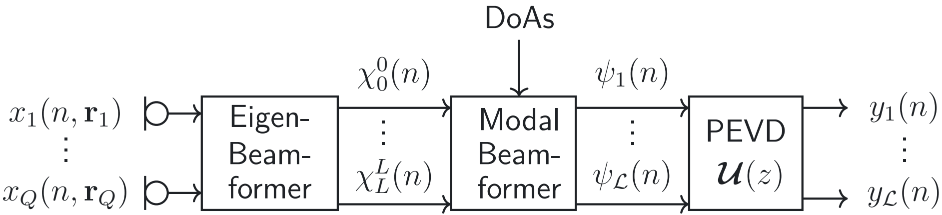
This Talk: PEVD-based Source Separation

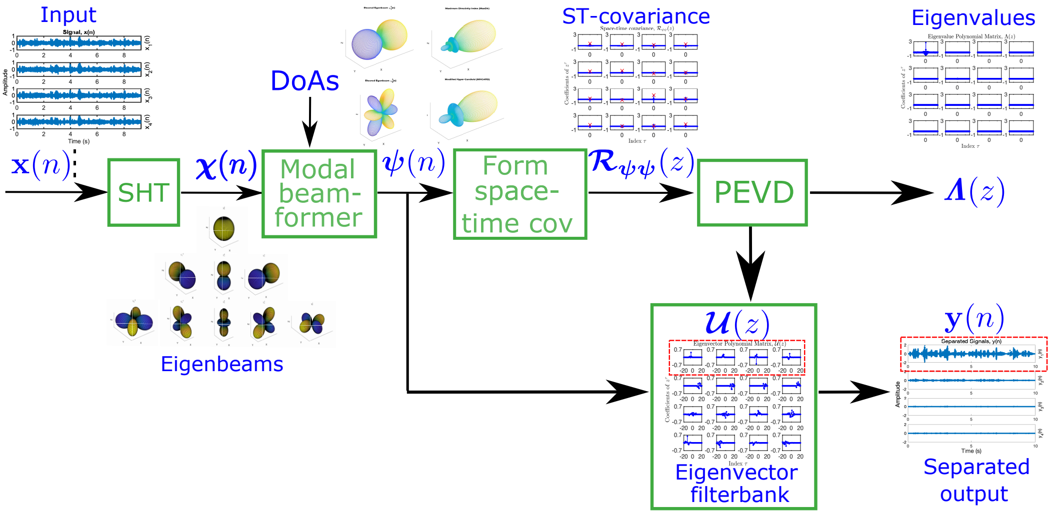
Background

Enhancement Algorithm: PEVD of Eigenbeams [Neo2021]



Sketch: PEVD of Eigenbeams for Separation





Input
Signal, x(n)
Time (s)
Amplitude
$\mathbf{x}(n)$
SHT
$\boldsymbol{\chi}(n)$
DoAs
Modal beam-former
$\boldsymbol{\psi}(n)$
Eigenbeams
ST-covariance
Space-time covariance, $\mathcal{R}_{\psi\psi}(z)$
Coefficients of z^{τ}
Index τ
Form space-time cov
$\mathcal{R}_{\psi\psi}(z)$
PEVD
Eigenvalues
Eigenvalue Polynomial Matrix, $\Lambda(z)$
$\boldsymbol{\Lambda}(z)$
$\mathcal{U}(z)$
Eigenvector Polynomial Matrix, $\mathcal{U}(z)$
Eigenvector filterbank
$\mathbf{y}(n)$
Separated Signals, y(n)
Separated output

Multichannel Multisource Signal Model

The received signal at the q-th microphone on a spherical array with time index n:

$$x_q(n, \mathbf{r}_q) = \sum_{p=1}^{P} \mathbf{h}_{p,q}^T \mathbf{s}_p(n)$$

where

- $\mathbf{h}_{p,q}^T$ is the room impulse response from pth source to qth microphone,
- $\mathbf{s}_p(n)$ is the pth localized source signal,
- $\mathbf{r}_q = (r, \theta_q, \phi_q)$ is the spherical coordinate with radius r, elevation angle θ_q, and azimuth angle ϕ_q.

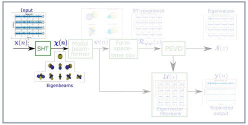
The data vector collected from Q sensors:

$$\mathbf{x}(n, \mathbf{r}) = [x_1(n, \mathbf{r}_1), x_2(n, \mathbf{r}_2), \ldots, x_Q(n, \mathbf{r}_Q)]^T.$$

Spherical Harmonics (SH) Decomposition

Imperial College London

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



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

For an order L sound field, each microphone signal is a weighted sum of SH

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

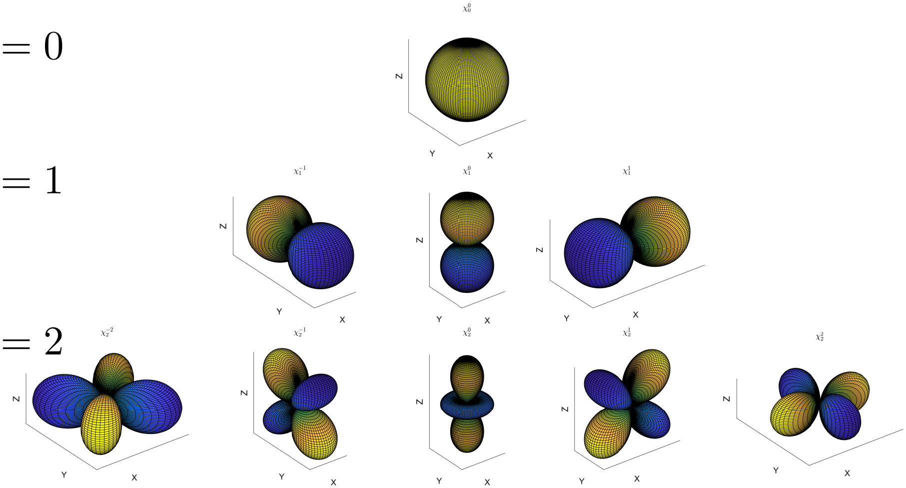
and alias-free spatial reconstruction requires $Q \geq (L+1)^2$.

Spherical Harmonics Decomposition

$\ell = 0$

$\ell = 1$

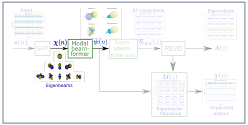
$\ell = 2$

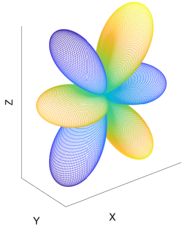


$m = -2$ $\quad m = -1$ $\quad m = 0$ $\quad m = 1$ $\quad m = 2$

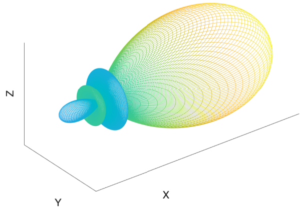
Eigenbeamformer Outputs

Given target source DoA (θ_p, ϕ_p), eigenbeam signals $\chi_\ell^m(n)$ are steered or used to form $\mathcal{L}$ beamformer outputs $\boldsymbol{\psi}(n) = [\psi_1(n), \ldots, \psi_{\mathcal{L}}(n)]^T$.

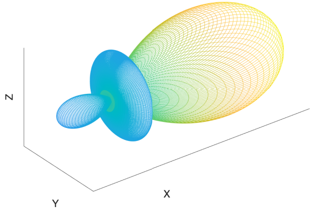




Steered eigenbeam.



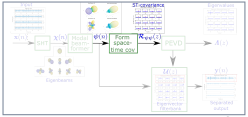
Maximum directivity index (MaxDir) beamformer.



Modified hyper-cardioid (MHCARD) beamformer.

Space-time Covariance Polynomial Matrix

Assuming stationarity, the space-time covariance matrix of $\mathcal{L}$ modal outputs is



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

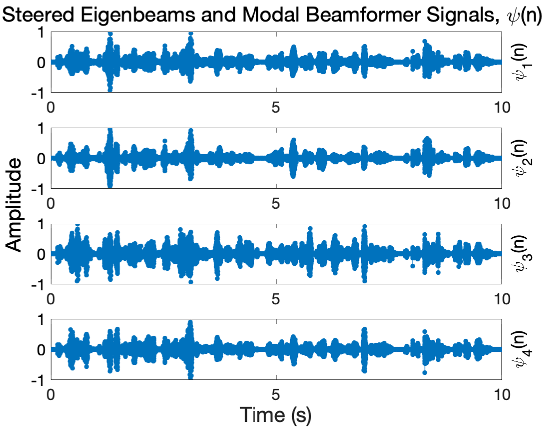
where $(i,j)^{\text{th}}$ element is the correlation function $r_{ij}(\tau) = \mathbb{E}[\psi_i(n)\psi_j^*(n-\tau)]$ and τ is the time-shift.

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

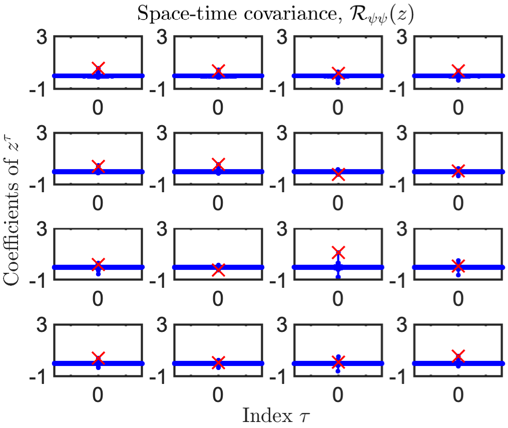
$$\boldsymbol{\mathcal{R}}_{\psi\psi}(z) = \sum_{\tau=-W}^{W} \mathbf{R}_{\psi\psi}(\tau) z^{-\tau},$$

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

Example: Polynomial Matrix from ST-Covariance

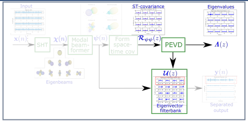


Modal signals $\psi(n)$.



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

Polynomial Matrix Eigenvalue Decomposition



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

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

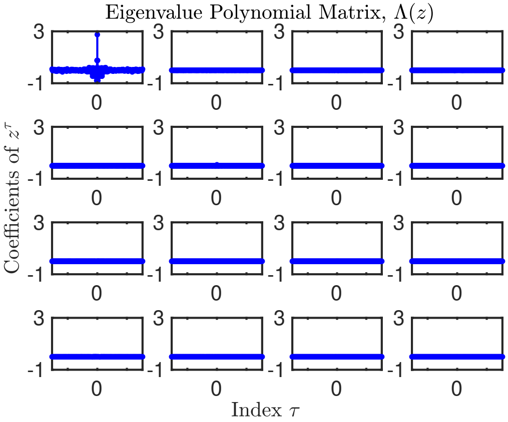
where $\boldsymbol{\Lambda}(z)$,$\boldsymbol{\mathcal{U}}(z)$ contain the eigenvalues and eigenvectors and $\boldsymbol{\mathcal{R}}^{P}_{\psi\psi}(z) = \boldsymbol{\mathcal{R}}^{H}_{\psi\psi}(z^{-1})$.

Subspace decomposition by the PEVD generates strongly decorrelated outputs:

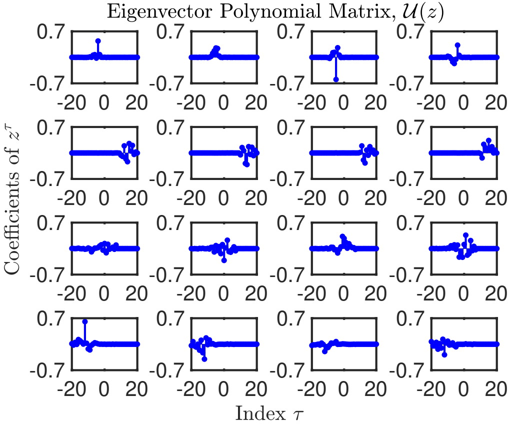
$$\boldsymbol{\mathcal{R}}_{\psi\psi}(z) = \left[\; \boldsymbol{\mathcal{U}}^{P}_{s}(z) \;\middle|\; \boldsymbol{\mathcal{U}}^{P}_{s\perp}(z) \;\right] \left[\begin{array}{c|c} \boldsymbol{\Lambda}_{s}(z) & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{\Lambda}_{s\perp}(z) \end{array}\right] \left[\begin{array}{c}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

associated with orthogonal target source, $\{\cdot\}_{s}$ and interferer, $\{\cdot\}_{s\perp}$ subspaces.

Example: PEVD Algorithm Outputs



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



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

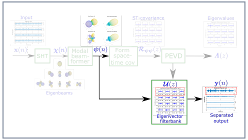
PEVD Algorithms

PEVD algorithms include:

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

Filterbank for Target Source Extraction

Imperial College London



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

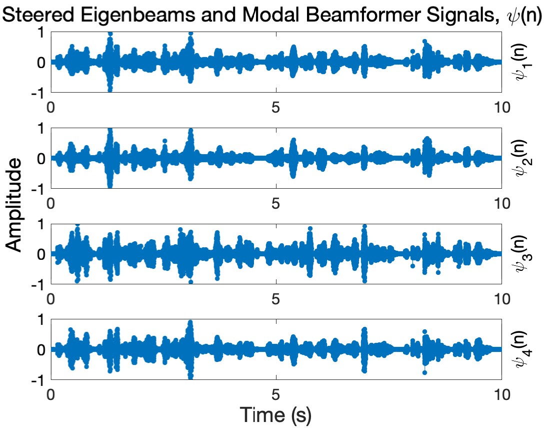
$$\mathbf{y}(z) = \boldsymbol{\mathcal{U}}(z)\boldsymbol{\psi}(z) \implies \boldsymbol{\mathcal{R}}_{\mathbf{yy}}(z) \approx \boldsymbol{\Lambda}(z),$$

that are strongly decorrelated.

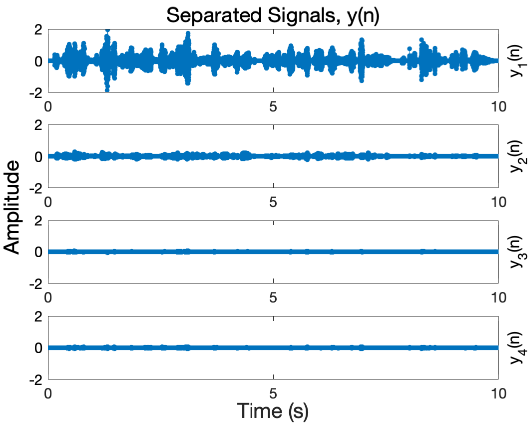
First channel output, $y_1(z)$, is the target source with ST-covariance

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

Example: Filterbank Output

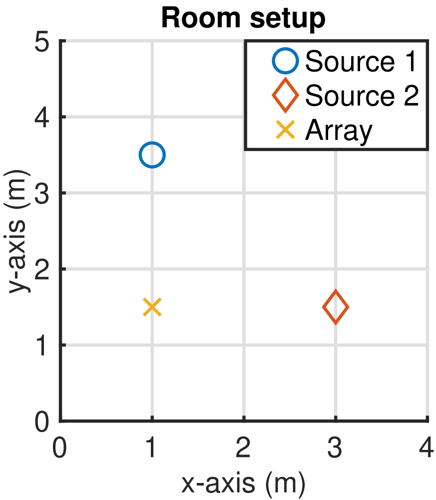


Modal signals $\psi(n)$.



Extracted source signals, $y(n)$.

Setup: 2 Female Speakers in Anechoic Room



Comparative Results

Source Separation Evaluation

Comparative algorithms:

1. Maximum Directivity index (MaxDir) [Meyer2002]
2. Modified Hyper-Cardioid (MHCARD)
3. AuxIVA [Ono 2011]
4. ILRMA [Makino 2018]
5. FastMNMF [Sekiguchi2019]

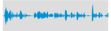
Evaluation measures:

- Separation metrics: SDR, SIR, SAR [Vincent2006]
- Short-Time Objective Intelligibility (STOI) [Taal2011]
- Perceptual Evaluation of Speech Quality (PESQ) [ITU-T P.862]

Source Separation Performance - Anechoic

Algorithm	ΔSDR	ΔSIR	ΔSAR	ΔSTOI	ΔPESQ
AuxIVA	17.7 dB	25.3 dB	11.4 dB	0.21	1.05
FastMNMF	20.6 dB	**35.2 dB**	13.8 dB	0.21	1.28
ILRMA	19.5 dB	31.3 dB	12.8 dB	0.21	1.21
MaxDir	3.9 dB	3.4 dB	4.7 dB	0.07	0.22
MHCARD	16.9 dB	17.8 dB	13.4 dB	0.21	0.93
PEVD	**21.8 dB**	25.3 dB	**16.4 dB**	**0.24**	**1.39**

Received



MaxDir

MHCARD

AuxIVA

FastMNMF

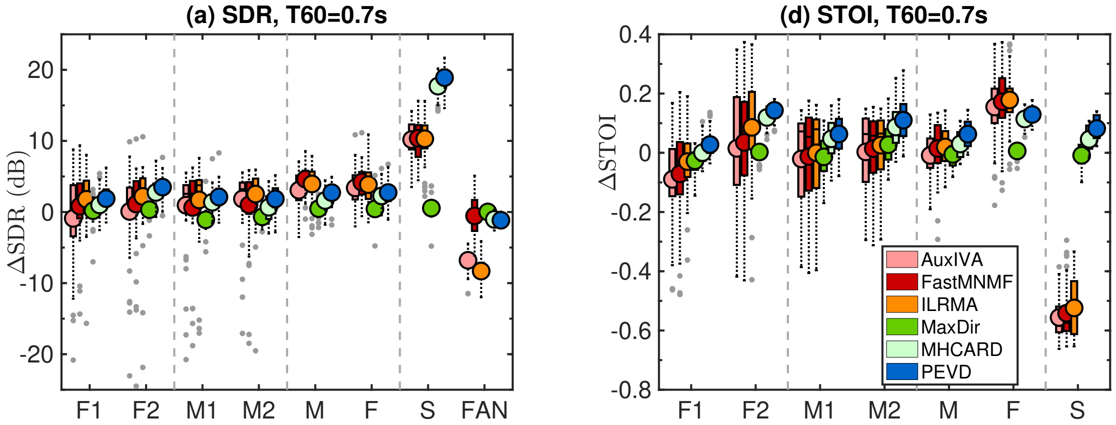
ILRMA

PEVD

Clean

Source Separation in Reverberant Room

Imperial College London



*F1 and M1 are at different positions.

Conclusion

Conclusion

- PEVD-based Source Separation Using Informed Spherical Arrays
  - Uses prior DoA information to steer eigenbeam signals and form modal beamformer outputs to extract target sources
- Performance of proposed PEVD-based approach
  - Among the best but does not always outperform other approaches
  - Achieves separation without introducing audible artefacts

References

- Corr, J., K. Thompson, S. Weiss, J. G. McWhirter, S. Redif, and I. K. Proudler (2014). "Multiple Shift Maximum Element Sequential Matrix Diagonalisation for Parahermitian Matrices". In: *Proc. IEEE/SP Workshop on Statistical Signal Processing*, pp. 844–848.
- Makino, S. (2018). *Audio Source Separation*. Signals and Communication Technology. Springer-Verlag.
- 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.
- Meyer, J. and G. Elko (May 2002). "A Highly Scalable Spherical Microphone Array Based on an Orthonormal Decomposition of the Soundfield". In: *Proc. IEEE Intl. Conf. on Acoust., Speech and Signal Process. (ICASSP)*. Vol. 2, pp. 1781–1784.
- Neo, V. W., C. Evers, and P. A. Naylor (June 2021). "Polynomial Matrix Eigenvalue Decomposition of Spherical Harmonics for Speech Enhancement". In: *Proc. IEEE Intl. Conf. on Acoust., Speech and Signal Process. (ICASSP)*.
- Neo, V. W. and P. A. Naylor (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)*, pp. 8043–8047.
- Ono, N. (Oct. 2011). "Stable and Fast Update Rules for Independent Vector Analysis Based on Auxiliary Function Technique". In: *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 189–192.
- *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).

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.

Redif, S., S. Weiss, and J. G. McWhirter (Jan. 2015). "Sequential Matrix Diagonalisation Algorithms for Polynomial EVD of Parahermitian Matrices". In: *IEEE Trans. Signal Process.* 63.1, pp. 81–89.

Scheibler, Robin, Eric Bezzam, and Ivan Dokmanić (Apr. 2018). "Pyroomacoustics: A Python Package for Audio Room Simulation and Array Processing Algorithms". In: *Proc. IEEE Intl. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, pp. 351–355.

Sekiguchi, K., A. A. Nugraha, Y. Bando, and K. Yoshii (2019). "Fast Multichannel Source Separation Based on Jointly Diagonalizable Spatial Covariance Matrices". In: *Proc. European Signal Process. Conf. (EUSIPCO)*.

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.

Vincent, E., R. Gribonval, and C. Févotte (July 2006). "Performance Measurement in Blind Audio Source Separation". In: *IEEE Trans. Audio, Speech, Lang. Process.* 14.4, pp. 1462–1469.

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/pevd-separate/
Webpage: https://vwn09.github.io