Imperial College London

ICASSP 2021 TORONTO Canada June 6-11, 2021 Metro Toronto Convention Centre

UNIVERSITY OF Southampton

PEVD of Spherical Harmonics for Speech Enhancement [ID:1218]

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

*Department of Electrical and Electronic Engineering, Imperial College London, U.K.

†School of Electronics and Computer Science, University of Southampton, U.K.

Summary

- PEVD of eigenbeams for speech enhancement performs identically and sometimes better than Raw PEVD [1] at a fraction of its complexity factor
- Effective for speech enhancement in noisy, reverberant environments
- No noticeable artifacts, blind and unsupervised approach
- Robust even when eigenbeams are not steered towards the speaker
- Listening examples are available at [2]

Multi-channel Signal Model

Incorporating the reverberant channel model $\mathbf{h}_q$, the q-th microphone signal on a spherical array is

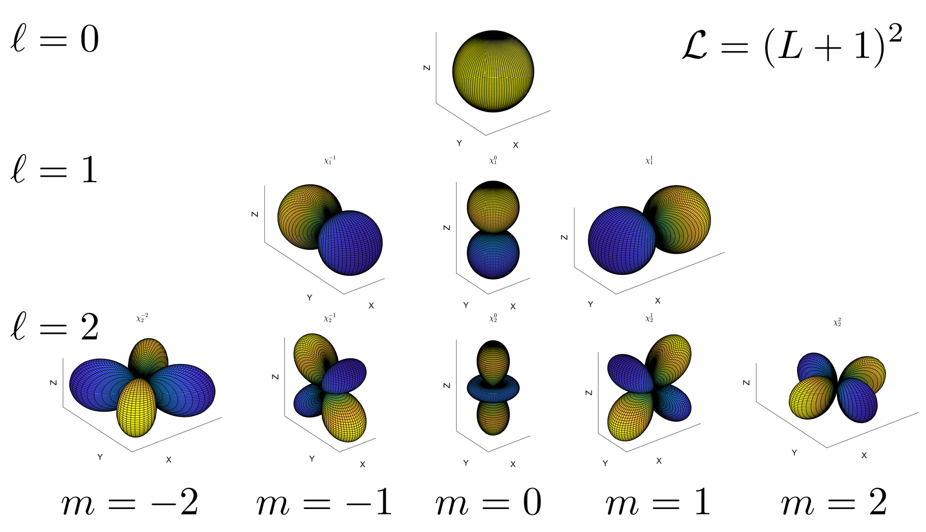
$$x_q(n, \mathbf{r}_q) = \mathbf{h}_q^T \mathbf{s}_0(n) + v_q(n) = \tilde{s}_q(n) + \tilde{v}_q(n),$$

where $\mathbf{s}_0(n)$: clean speech, $v_q(n)$: noise signal,
$\mathbf{r}_q = (r, \theta_q, \phi_q)$, r: radius, θ_q: elevation, ϕ: azimuth

Spherical Harmonic Decomposition

Spherical harmonic (SH) transform of the spatially sampled sound field:

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



α_q: quadrature weights,
L: max. SH order,
$\mathcal{L} = (L+1)^2$ eigenbeams,
$R_\ell^m(\mathbf{r}_q)$: ℓ-th order, m-th degree real SH basis

Arranging in ascending order and degree: $\boldsymbol{\chi} = [\chi_0^0, \chi_1^{-1}, \chi_1^0, \ldots, \chi_L^L]^T$

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

$\varepsilon(\%)$: Total squared error between received and L-th order approximation

$\beta = (\frac{\mathcal{L}}{Q})^3$, where $Q = 32 \Rightarrow$ More complexity savings with fewer eigenbeams.

Polynomial Eigenvalue Decomposition (PEVD)

Space-time Covariance Matrix, assuming stationarity, is:

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

Para-Hermitian Polynomial Matrix:

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

The PEVD of $\boldsymbol{\mathcal{R}}_{\mathbf{xx}}(z)$ is [3]:

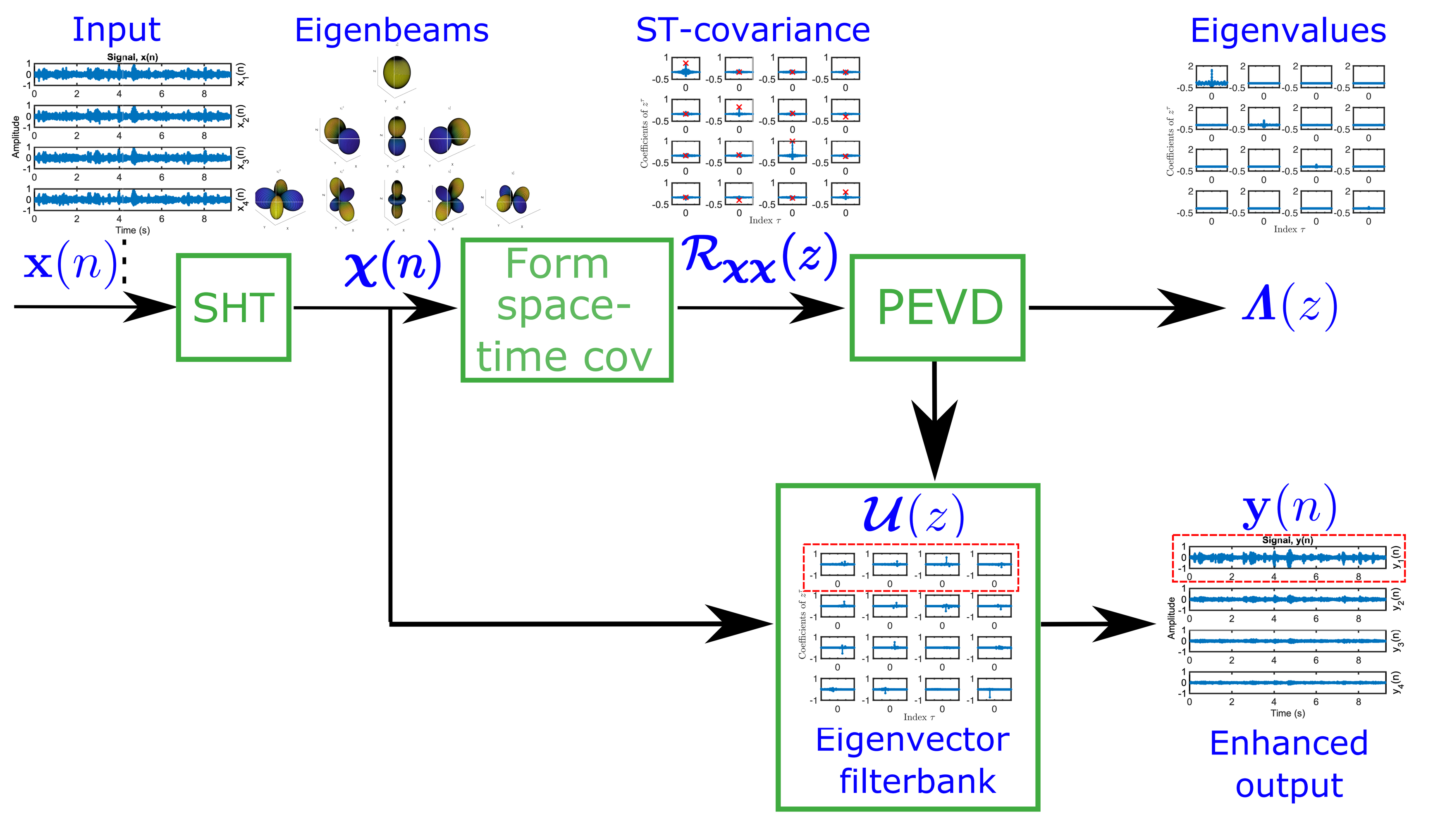
$$\boldsymbol{\mathcal{R}}_{\boldsymbol{\chi}\boldsymbol{\chi}}(z) \approx \boldsymbol{\mathcal{U}}^P(z)\boldsymbol{\Lambda}(z)\boldsymbol{\mathcal{U}}(z)$$

$$= \left[\, \boldsymbol{\mathcal{U}}_{\tilde{s}}^P(z) \mid \boldsymbol{\mathcal{U}}_{\tilde{v}}^P(z) \,\right] \begin{bmatrix} \boldsymbol{\Lambda}_{\tilde{s}}(z) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{\tilde{v}}(z) \end{bmatrix} \begin{bmatrix} \boldsymbol{\mathcal{U}}_{\tilde{s}}(z) \\ \boldsymbol{\mathcal{U}}_{\tilde{v}}(z) \end{bmatrix}$$

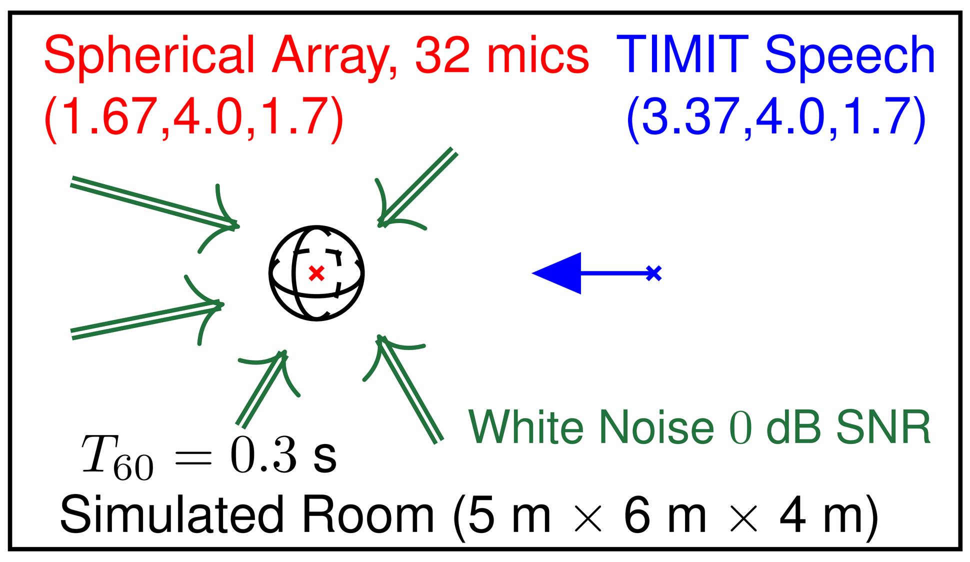
Filtering the eigenbeams through the $\boldsymbol{\mathcal{U}}(z)$ generates the enhanced signal:

$$\mathbf{y}(z) = \begin{bmatrix} \boldsymbol{\mathcal{U}}_{\tilde{s}}(z) \\ \mathbf{0} \end{bmatrix} \boldsymbol{\chi}(z)$$

PEVD of Spherical Harmonics Algorithm



Experiment Setup: Reverberant Speech in Noise



Comparative algorithms:

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

Enhancement measures:

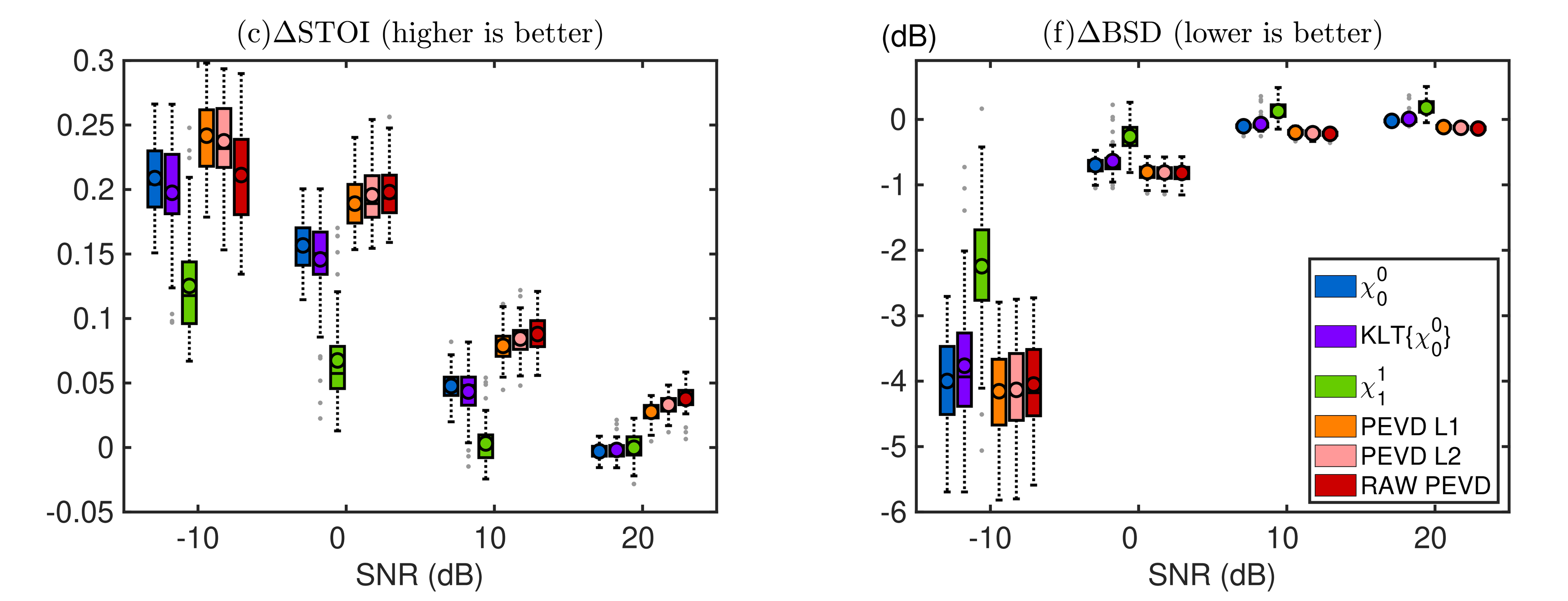
- Frequency-weighted Segmental SNR (FwSegSNR)
- Short-Time Objective Intelligibility (STOI)
- Perceptual Evaluation of Speech Quality (PESQ)
- Bark Spectral Distortion (BSD)

Speech Enhancement Results

SMIRGen Room, 0 dB White Noise

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

ACE Corpus: Lecture Room 2, Babble Noise



References

[1] V. W. Neo, C. Evers, and P. A. Naylor, "PEVD-based speech enhancement in reverberant environments," in *Proc. IEEE ICASSP*, 2020, pp. 186–190.

[2] ——, *PEVD using spherical harmonics for speech enhancement*, Oct. 2020. [Online]. Available: `https://vwn09.github.io/shd-pevd/`.

[3] J. G. McWhirter, P. D. Baxter, T. Cooper, S. Redif, and J. Foster, "An EVD algorithm for para-Hermitian polynomial matrices," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2158–2169, May 2007.