

POLYNOMIAL EIGENVALUE DECOMPOSITION-BASED TARGET SPEAKER VOICE ACTIVITY DETECTION IN THE PRESENCE OF COMPETING TALKERS

Vincent W. Neo*, Stephan Weiss†, Simon W. McKnight*, Aidan O. T. Hogg*, Patrick A. Naylor*

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

†Department of Electronic and Electrical Engineering, University of Strathclyde, Scotland

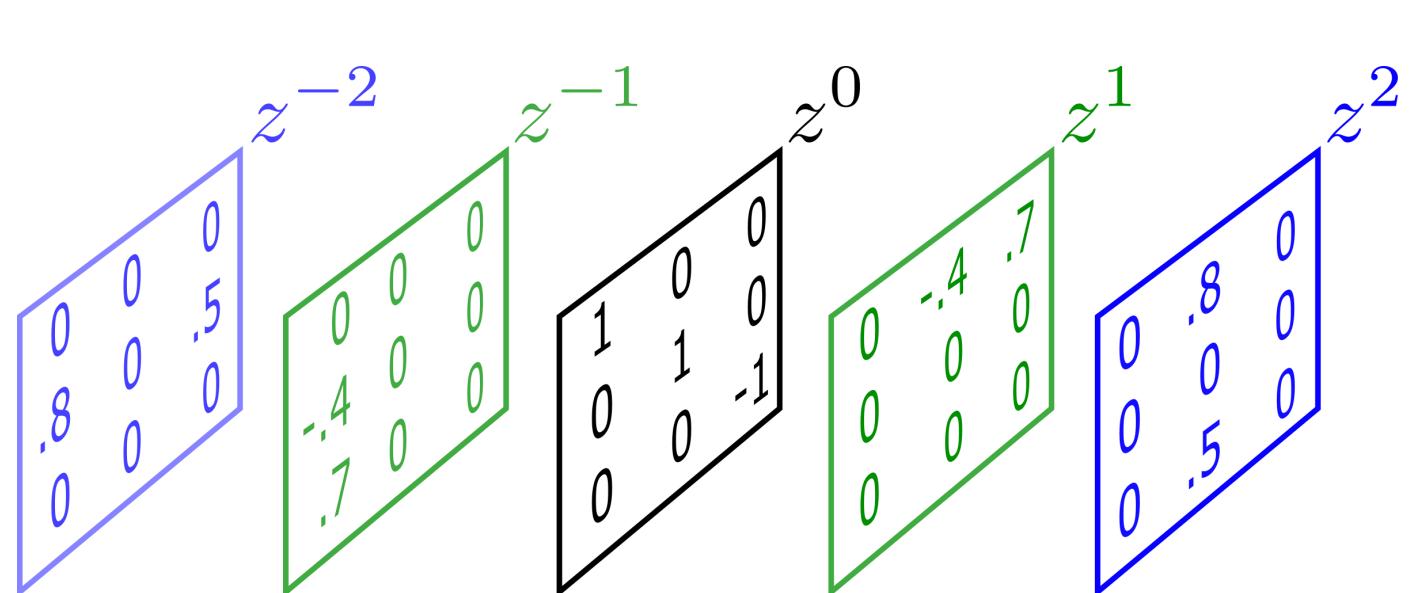


Summary

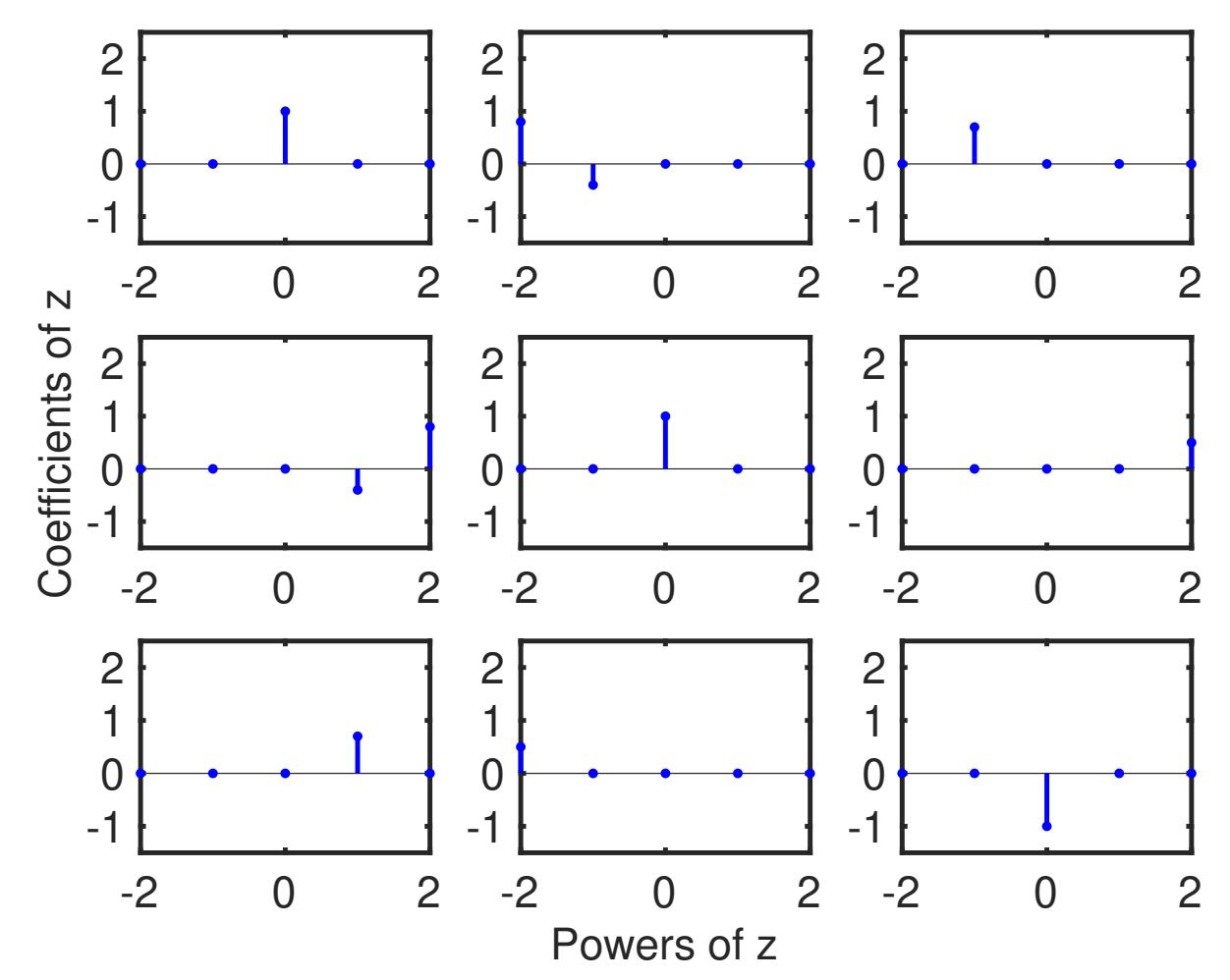
- PEVD-based VAD algorithm, which is inspired by [1], exploits multi-microphone processing to detect target speaker voice activity
- Consistently among the best in F1 and BACC scores even when the target and interfering speaker are of the same gender

What is a Polynomial Matrix?

Polynomial with matrix coefficients



Matrix with polynomial elements



How Do Polynomial Matrices Arise?

Multichannel model (P sources, Q microphones):

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

From Q sensors:

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

Assuming stationarity, space-time covariance matrix:

$$\mathbf{R}(\tau) = \mathbb{E}[\mathbf{x}(n)\mathbf{x}^T(n-\tau)]$$

Para-Hermitian polynomial matrix:

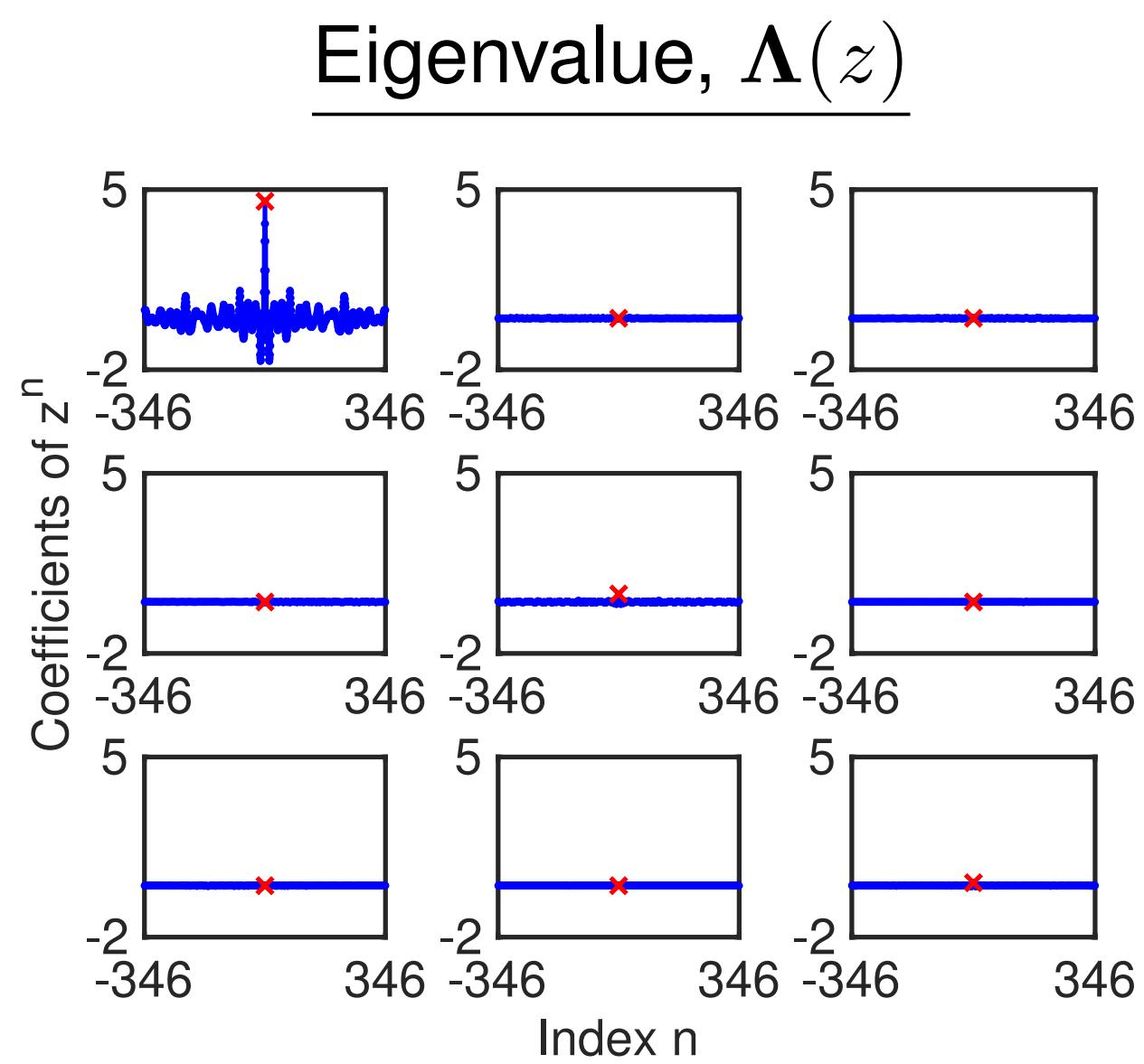
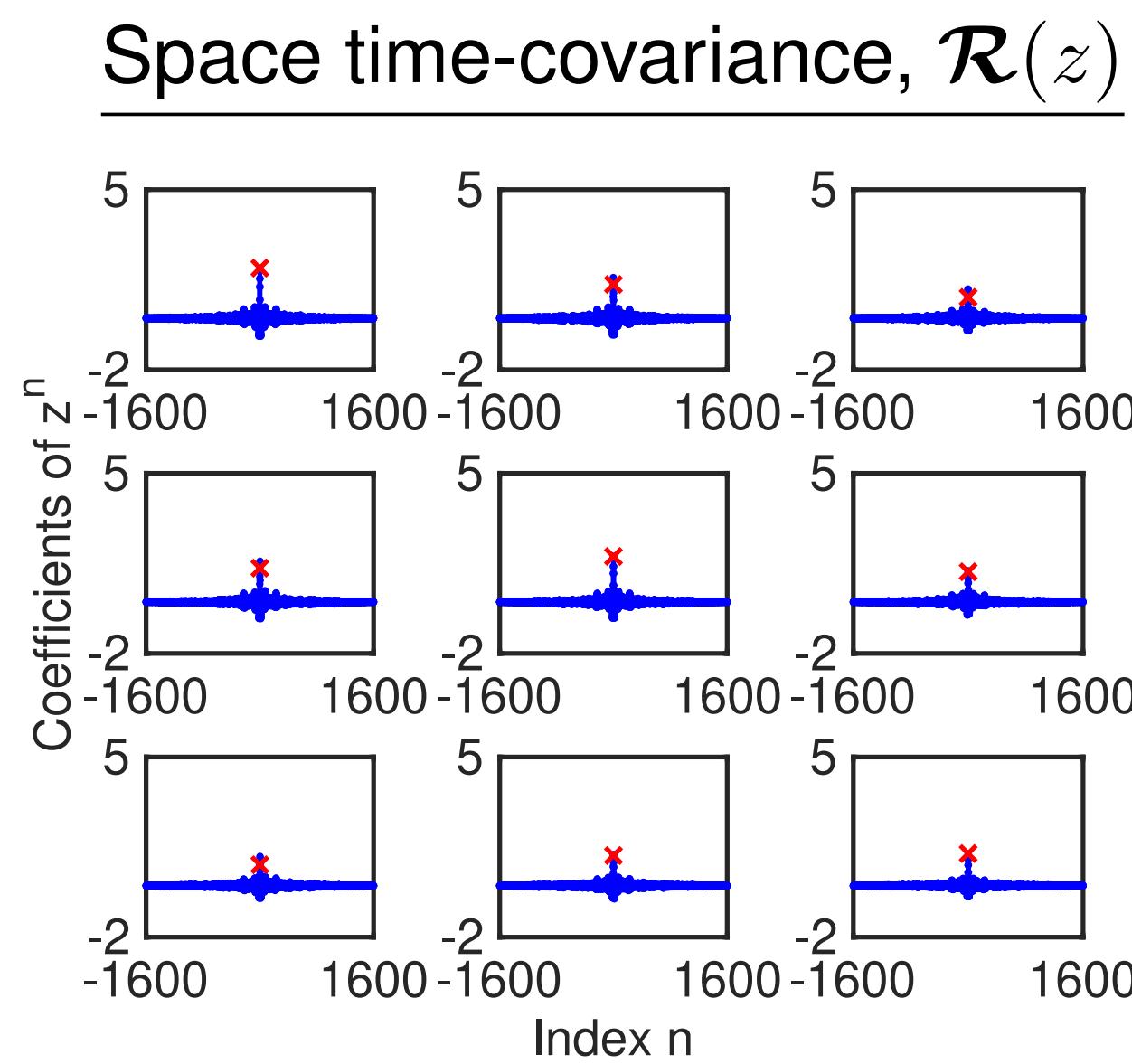
$$\mathcal{R}(z) = \sum_{\tau=-W}^W \mathbf{R}(\tau) z^{-\tau}$$

Polynomial Eigenvalue Decomposition (PEVD)

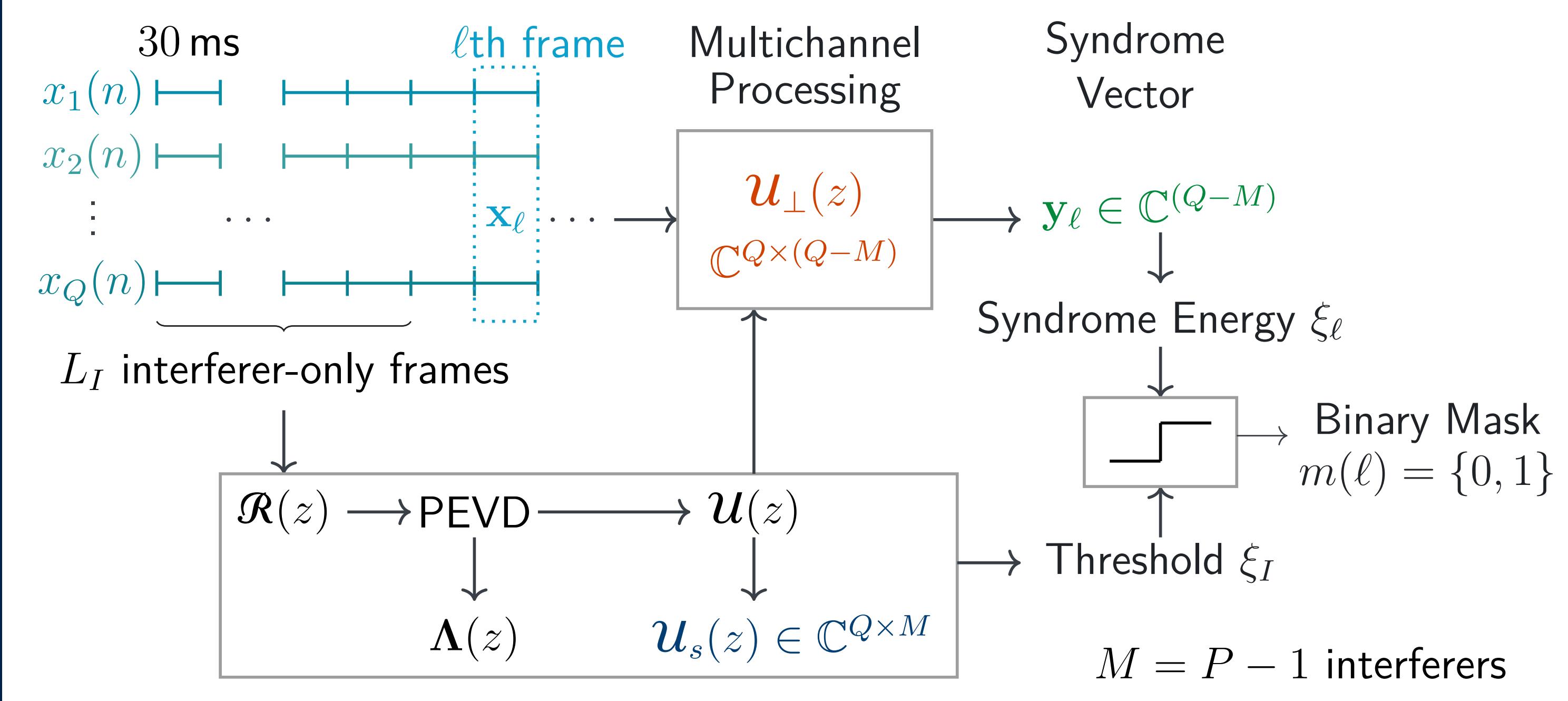
The PEVD of $\mathcal{R}(z)$ is [2]:

$$\mathcal{R}(z) = [\mathbf{U}_s(z) \quad \mathbf{U}_{\perp}(z)] \begin{bmatrix} \Lambda_s(z) & \mathbf{0} \\ \mathbf{0} & \Lambda_{\perp}(z) \end{bmatrix} [\mathbf{U}_s^P(z) \quad \mathbf{U}_{\perp}^P(z)], \quad (1)$$

associated with signal, $\{\cdot\}_s$ and orthogonal complement, $\{\cdot\}_{\perp}$ subspaces.



PEVD-based Target Speaker Voice Activity Detection



Experiment Setup: Reverberant Speech and Interferer

Comparative Algorithms

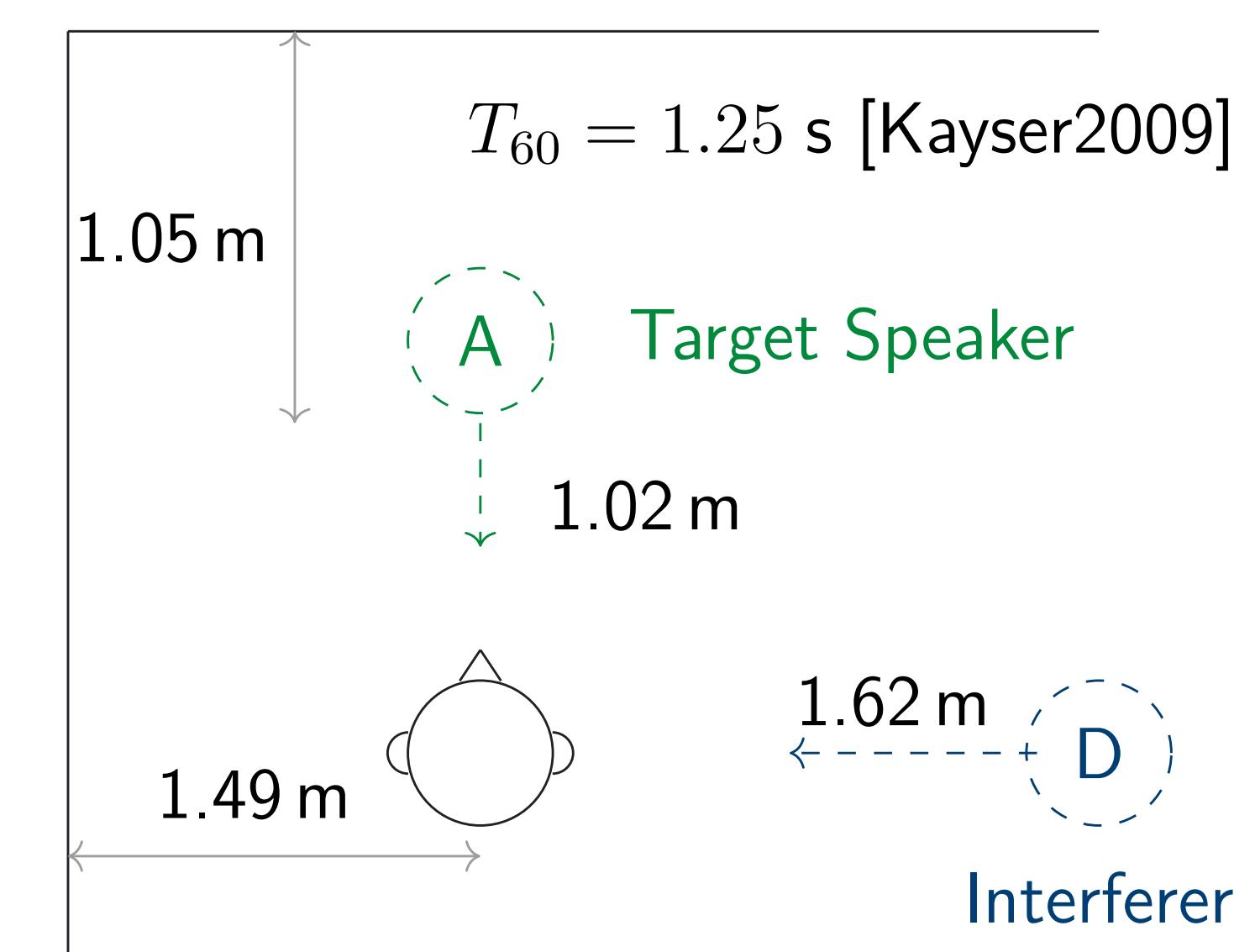
- Sohn
- WebRTC: G0 (Least aggressive)
- WebRTC: G3 (Most aggressive)
- PEVD (Proposed)

Evaluation Measures

- Label evaluation metrics
 - True Positive (TP)
 - True Negative (TN)
 - False Positive (FP)
 - False Negative (FN)
- F1 score
- Balanced Accuracy (BACC)

Setup

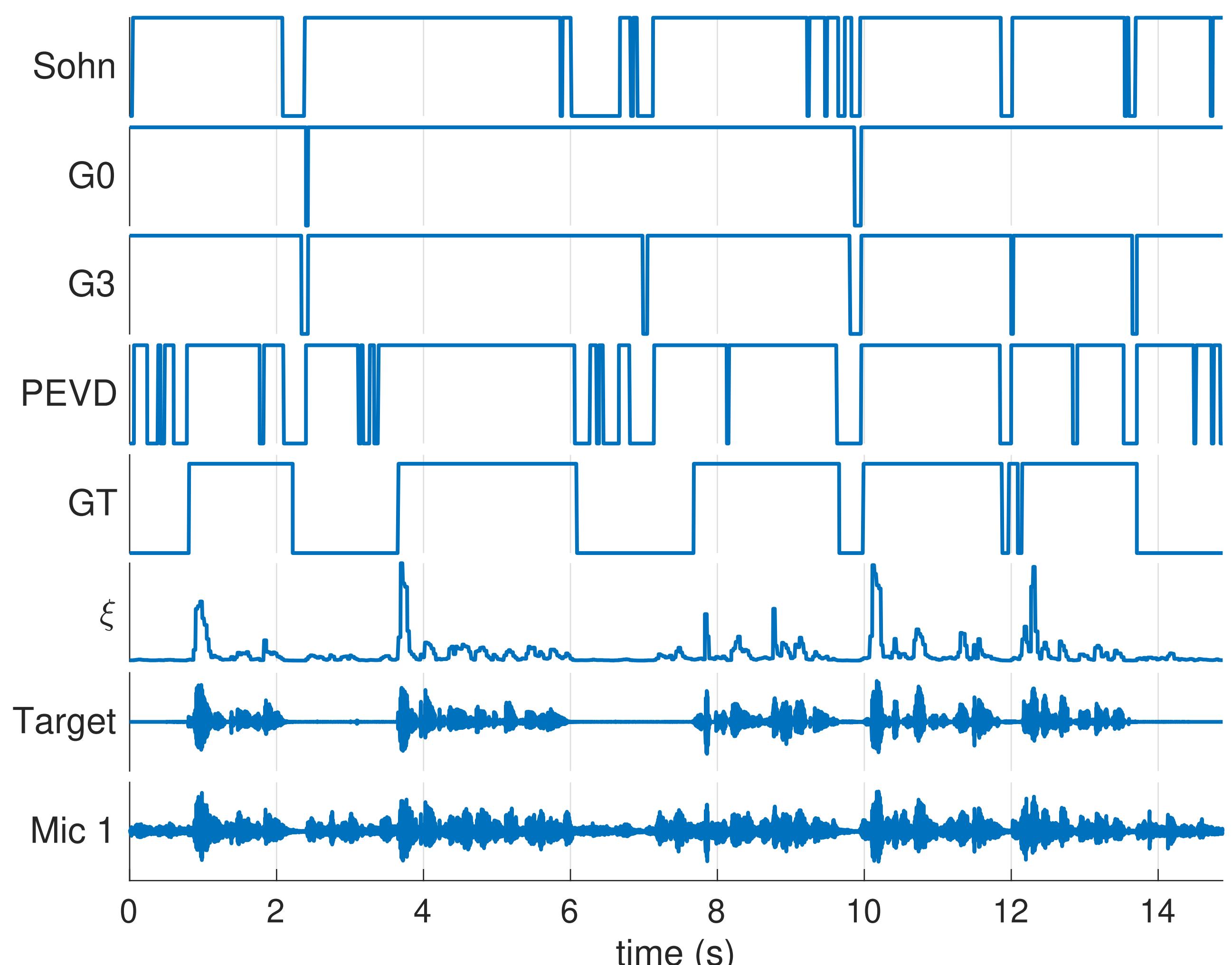
$P = 2$ Sources, $Q = 2$ Microphones



Simulation Results

VAD performance of female target and male interferer at 5 dB SIR:

Metric	Sohn	G0	G3	PEVD
TP	295	313	310	294
TN	43	4	10	72
FP	139	178	172	110
FN	18	0	3	19
F1	0.790	0.779	0.780	0.820
BACC	0.589	0.511	0.523	0.667



Listening examples are available [3].

References

- S. Weiss, C. Delaosa, J. Matthews, I. K. Proudler, and B. A. Jackson, "Detection of weak transient signals using a broadband subspace approach," in *Sensor Signal Process. for Defence Conf. (SSPD)*, 2021.
- S. Weiss, J. Pestana, and I. K. Proudler, "On the existence and uniqueness of the eigenvalue decomposition of a para-Hermitian matrix," *IEEE Trans. Signal Proc.*, vol. 66, no. 10, pp. 2659–2672, May 2018.
- V. W. Neo, *PEVD-based target speaker VAD*, Apr. 2022. [Online]. Available: <https://vwn09.github.io/research/pevd-tsvad-iwaenc>.