PADERBORN UNIVERSITY
The University for the Information Society

Department of Communications Engineering

Blind Speech Separation based on Complex Spherical k-Mode Clustering

Lukas Drude
Christoph Boeddeker
Reinhold Haeb-Umbach

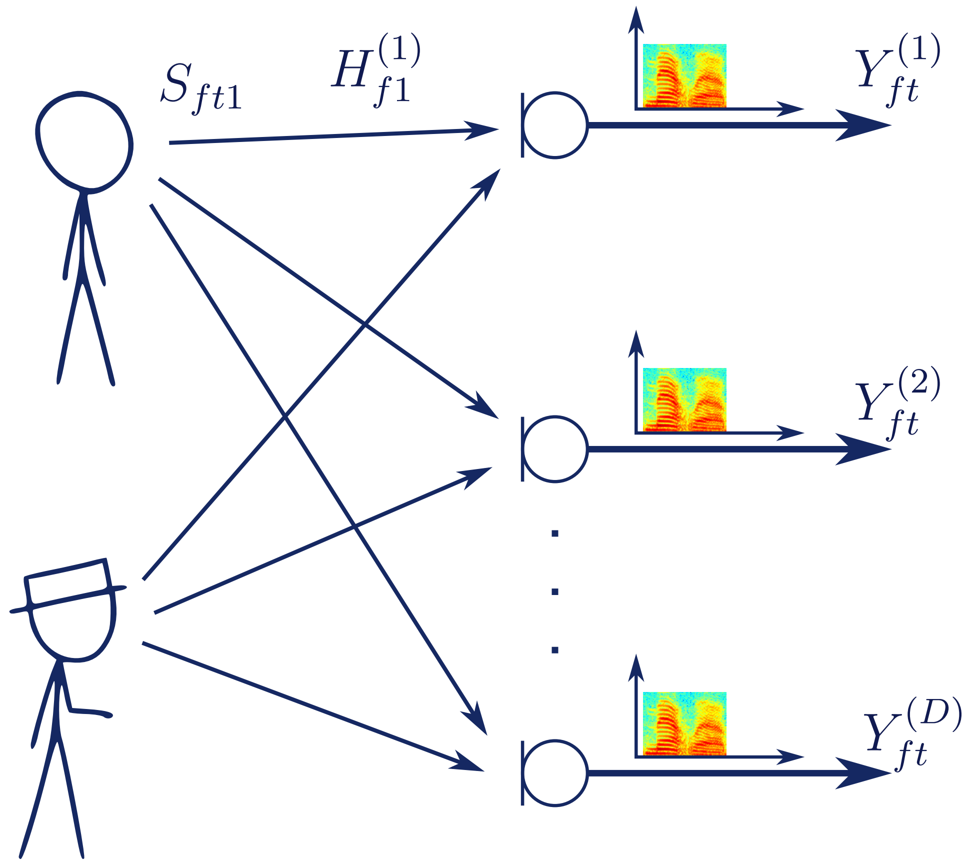
Contents

- Scenario
- Statistical model
- Solve blind source separation for each frequency
    - EM algorithm
    - Spherical k-mode clustering
- Comparison
- Results
- Summary

Scenario

- Multiplicative transfer function approximation

$$\mathbf{Y}_{ft} = \sum_{k=1}^{K} \mathbf{H}_{fk} S_{ftk} + \mathbf{N}_{ft}$$

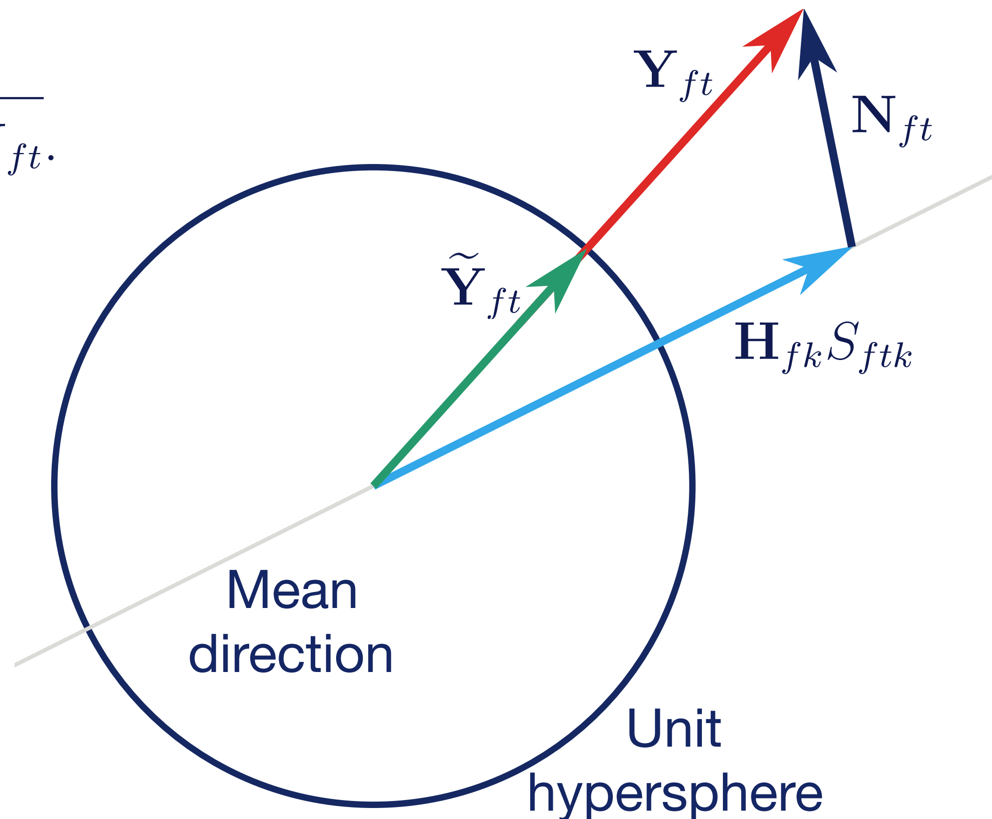


Statistical model

$$\mathbf{Y}_{ft} = \sum_{k=1}^{K} \mathbf{H}_{fk} S_{ftk} + \mathbf{N}_{ft}$$

$$\widetilde{\mathbf{Y}}_{ft} = \frac{\mathbf{Y}_{ft}}{A_{ft}}, \quad \text{where } A_{ft} = \sqrt{\mathbf{Y}_{ft}^{\mathrm{H}} \mathbf{Y}_{ft}}.$$

- Complex valued vectors
- Unit length
- Decouple transmission path and speech source



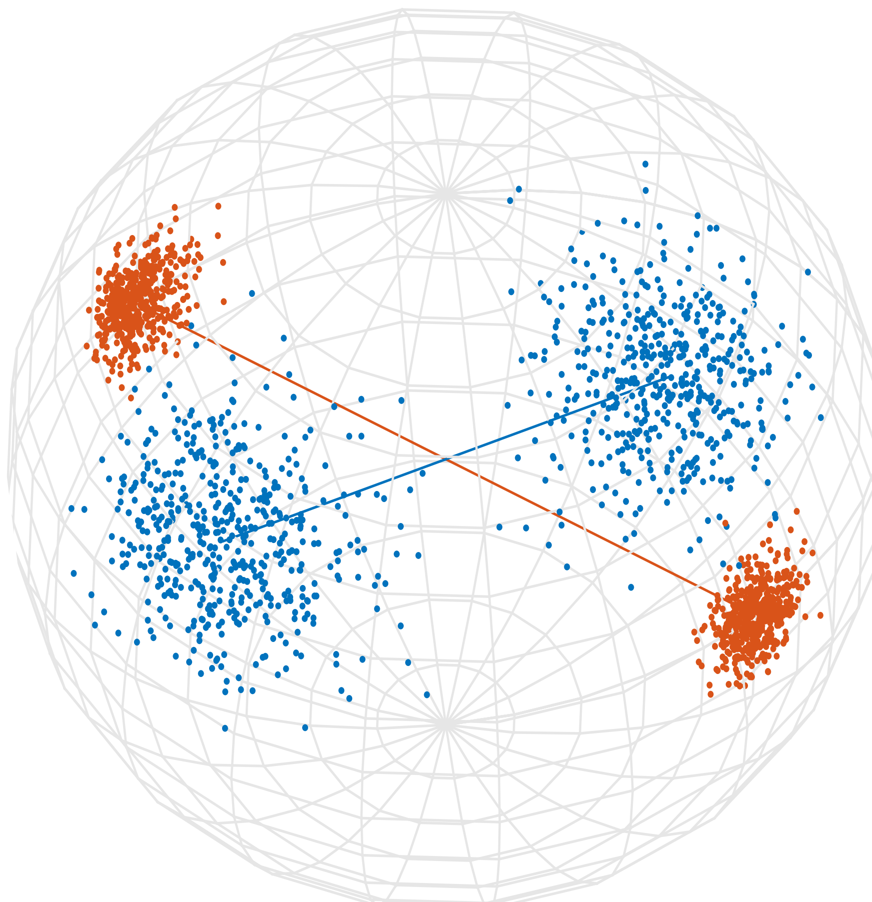
- Idea: Observations form clusters on unit-hypersphere
- Use: Complex Watson Mixture Model

Statistical model: Complex Watson Mixture Model

$$p(\widetilde{\mathbf{Y}}_t) = \sum_{k=1}^{K} \pi_k p(\widetilde{\mathbf{Y}}_t | c_t = k)$$

$$p(\widetilde{\mathbf{Y}}_t | c_t = k) = \frac{1}{c_{\mathrm{W}}(\kappa_k)} \mathrm{e}^{\kappa_k |\widetilde{\mathbf{Y}}_t^{\mathrm{H}} \mathbf{W}_k|^2}$$

- Use EM algorithm to estimate parameters



EM algorithm

- Estimation step:
  - Calculate class affiliation posterior

$$\gamma_{tk} := P(c_t = k|\widetilde{\mathbf{Y}}_t) = \frac{P(c_t = k)p(\widetilde{\mathbf{Y}}_t|c_t = k)}{\sum_{k=1}^{K} P(c_t = k)p(\widetilde{\mathbf{Y}}_t|c_t = k)}$$

- Maximization step:

$$\pi_k = N_k/T, \quad \text{where } N_k = \sum_{t=1}^{T} \gamma_{tk}$$

$$\mathbf{W}_k = \mathcal{P}\left\{\mathbf{\Phi}_k\right\}, \quad \text{where } \mathbf{\Phi}_k = \frac{1}{N_k}\sum_{t=1}^{T} \gamma_{tk}\widetilde{\mathbf{Y}}_t\widetilde{\mathbf{Y}}_t^{\mathrm{H}}$$

$$\kappa_k \quad \text{with implicit equation}$$

Spherical k-mode clustering

- Quantization of the posteriors

$$\hat{c}_{tk} = \begin{cases} 1, & k = \underset{\tilde{k}}{\operatorname{argmax}}\, \gamma_{t\tilde{k}}, \\ 0, & \text{else.} \end{cases}$$

- Equal mixture weights, shared concentrations

$$\begin{aligned} k &= \underset{\tilde{k}}{\operatorname{argmax}}\, \pi_{\tilde{k}} \frac{1}{c_{\mathrm{W}}(\kappa_{\tilde{k}})} \mathrm{e}^{\kappa_{\tilde{k}} |\widetilde{\mathbf{Y}}_t^{\mathrm{H}} \mathbf{W}_{\tilde{k}}|^2} \\ &= \underset{\tilde{k}}{\operatorname{argmax}}\, \mathrm{e}^{|\widetilde{\mathbf{Y}}_t^{\mathrm{H}} \mathbf{W}_{\tilde{k}}|^2} \\ &= \underset{\tilde{k}}{\operatorname{argmax}}\, |\widetilde{\mathbf{Y}}_t^{\mathrm{H}} \mathbf{W}_{\tilde{k}}|^2 \end{aligned}$$

Comparison

EM	Spherical k-mode	k-means
Evaluate PDF	Squared cosine distance	Euclidean distance
Estimate mode (PCA)	Estimate mode (PCA)	Estimate mean
Estimate concentration		
		Additional normalization

Comparison: Number of iterations

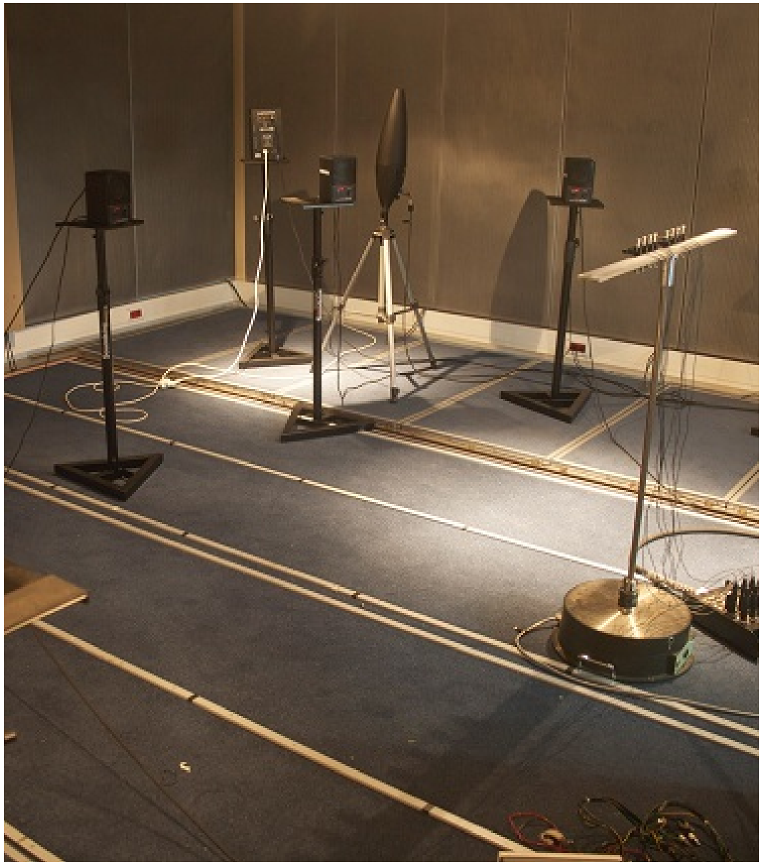
	EM	Spherical k-mode	k-means
Likelihood calculations	TKI	0	0
Eigenvalue decomposition	IK	IK	0
Implicit equations	IK	0	0

Comparison: Duality

EM for complex Watson Mixture Model	→	Complex spherical k-mode
EM for Gaussian Mixture Model	→	k-means
EM for von-Mises-Fisher Mixture Model	→	Spherical k-means

Results: Evaluation Setup

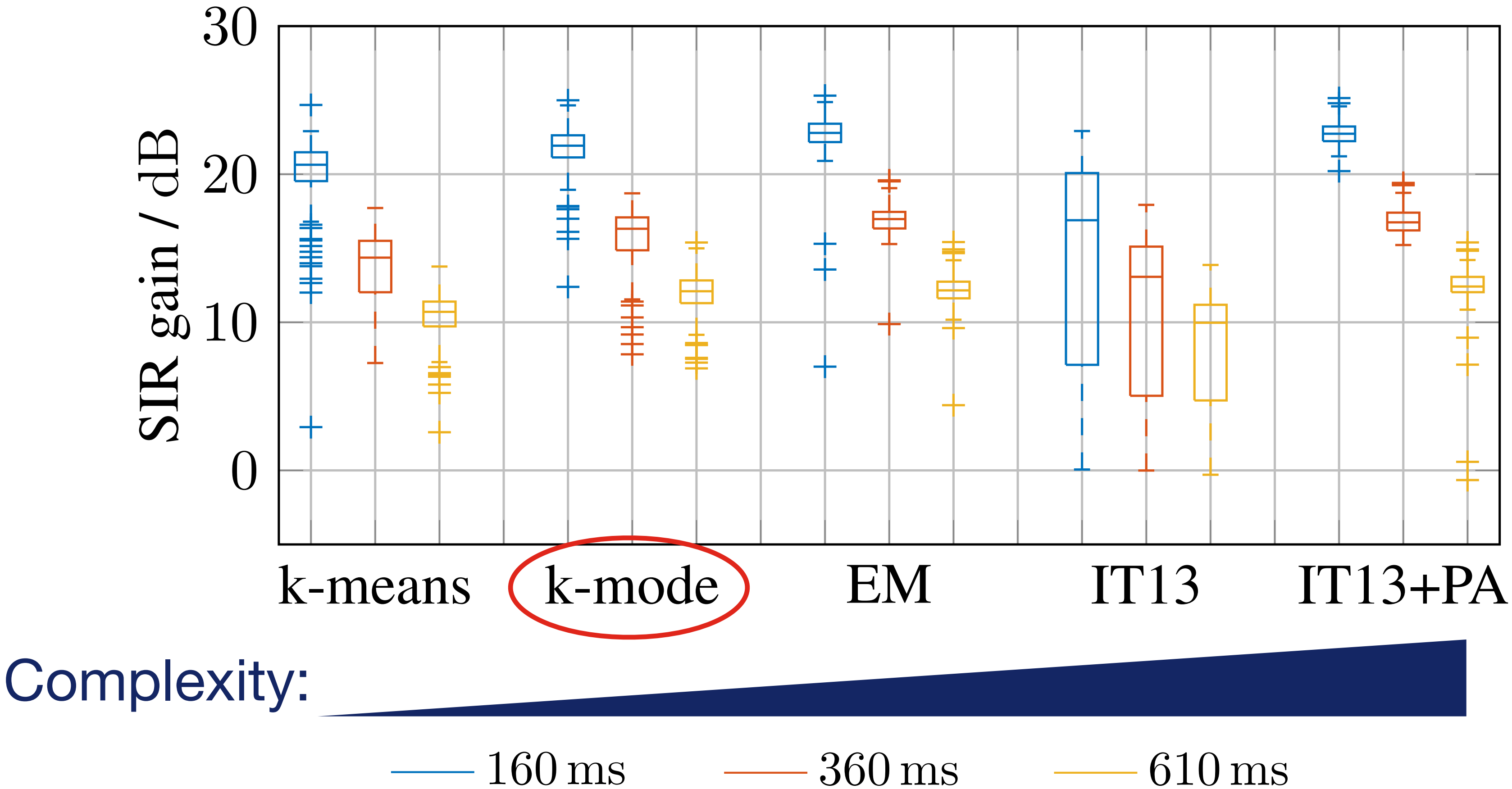
- 5 s utterances from TIMIT DB
- Room impulse responses from MIRD DB
- Spherically isotropic noise
- LCMV-Beamformer to separate sources with given masks
  - PSD matrix from mask
  - Relative transfer functions using Mode Vectors



Hadad et al. 2014

Results

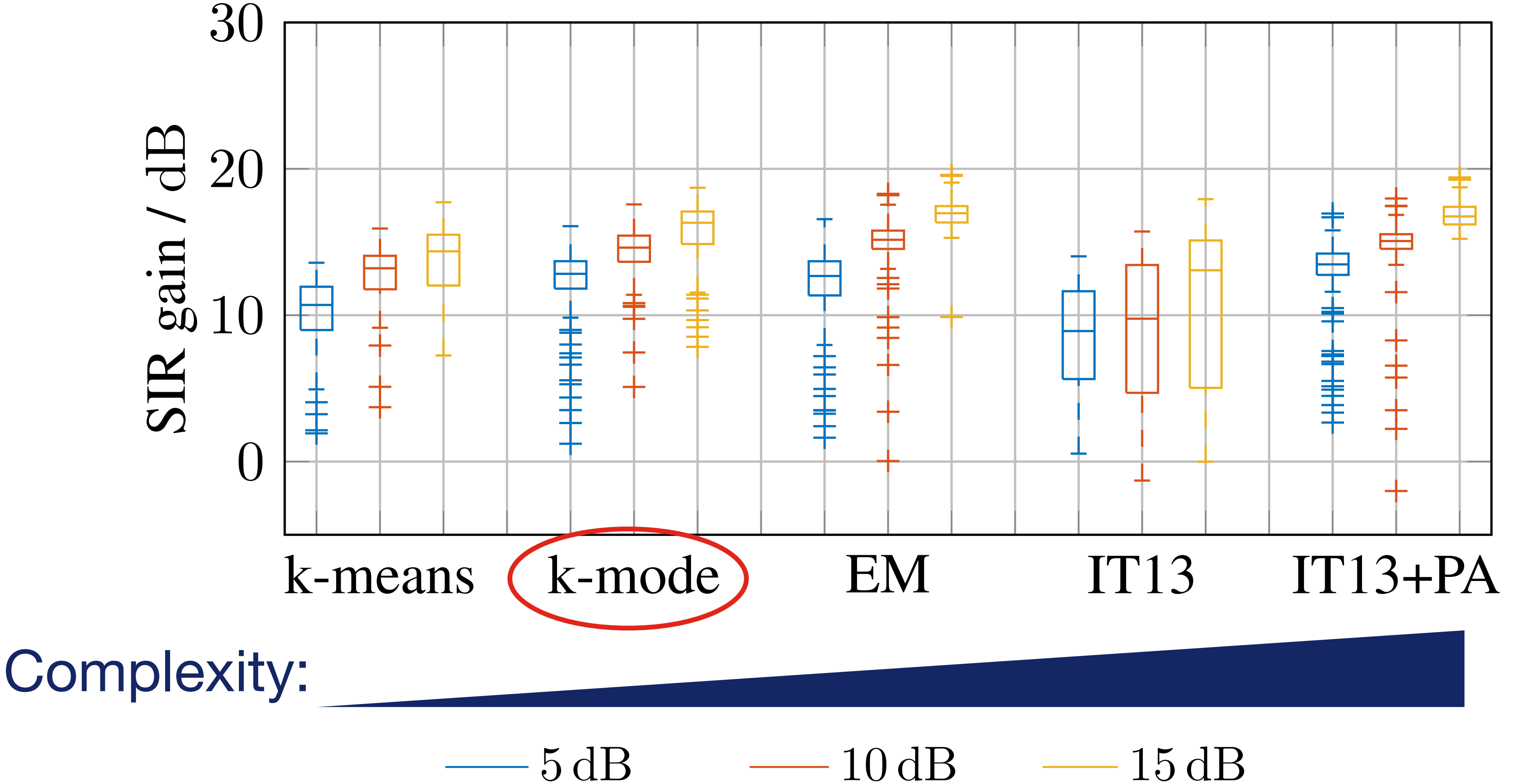
- SNR 15 dB, varying T_{60}



- IT13: Ito et al. 2013: Permutation-free convolutive blind source separation via full-band clustering based on frequency-independent source presence priors

Results

- Varying SNR, T_{60} 360 ms



- IT13: Ito et al. 2013: Permutation-free convolutive blind source separation via full-band clustering based on frequency-independent source presence priors

Summary

- Simplified algorithm provides comparable performance.
- Spherical k-mode relates to EM for complex Watson mixture model just as k-means relates to EM for GMMs.
- Spherical k-mode extends the limited tools for complex-valued directional data.

Thank you for listening!