



Blind Speech Separation based on Complex Spherical k-Mode Clustering

Lukas Drude
Christoph Boeddeker
Reinhold Haeb-Umbach



Contents

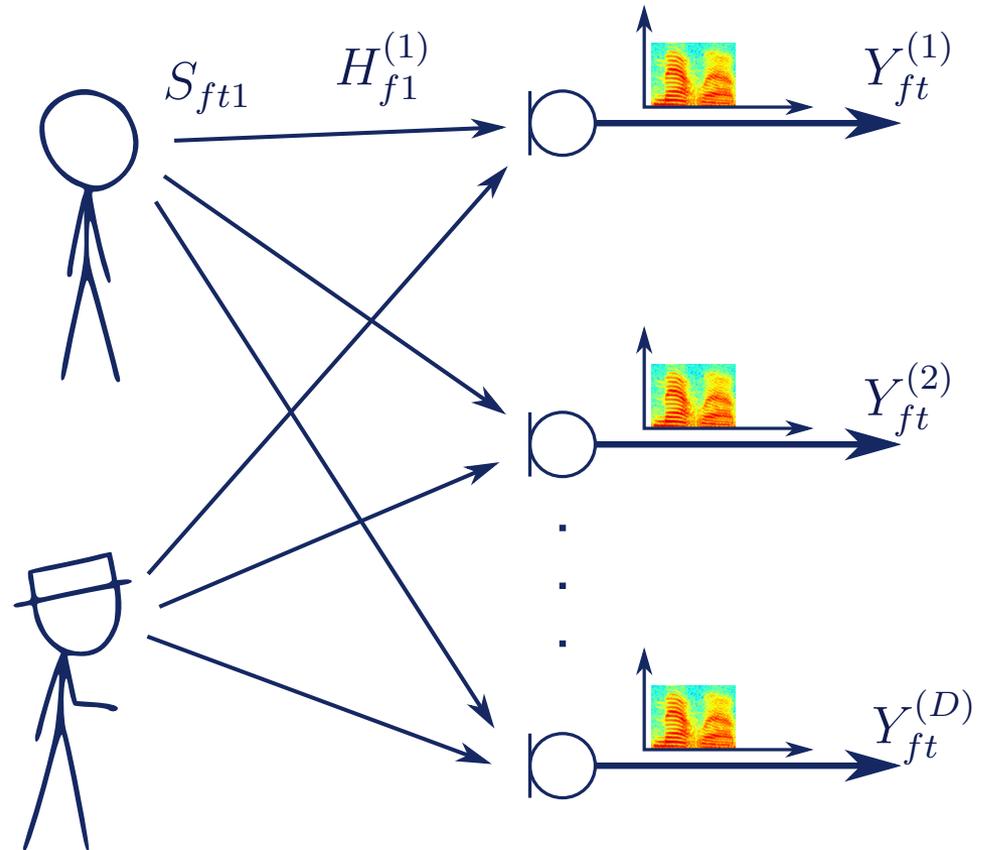
2

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

Scenario

- Multiplicative transfer function approximation

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



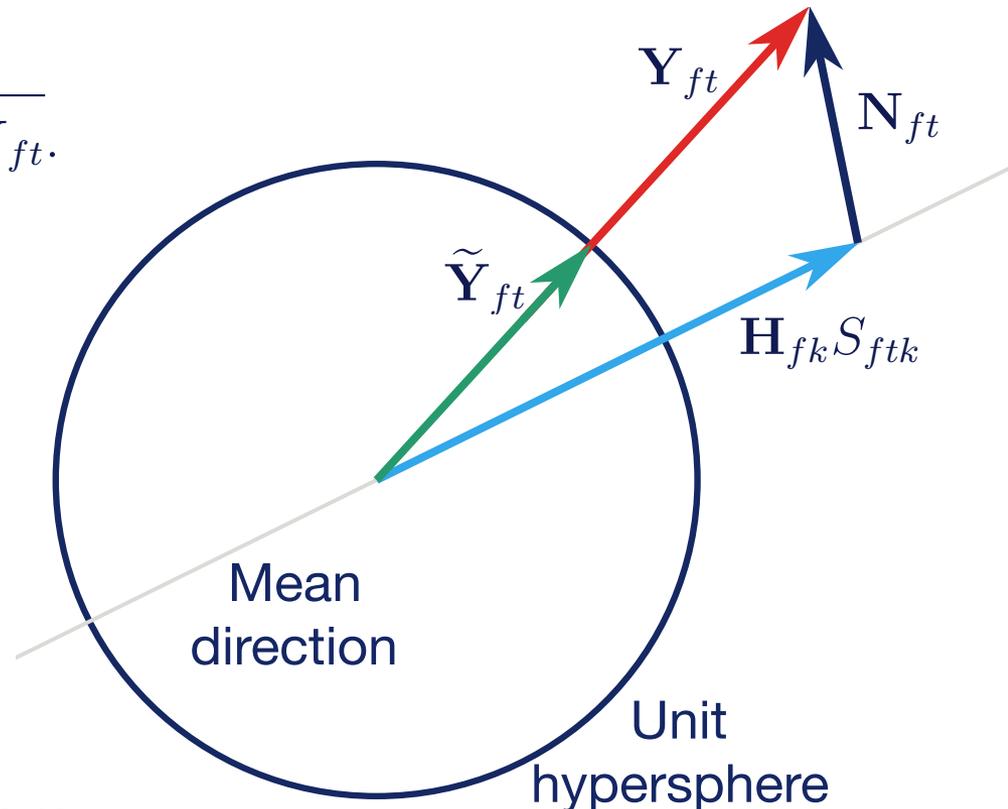
Statistical model

4

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

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

- Complex valued vectors
- Unit length
- Decouple transmission path and speech source
- Idea: Observations form clusters on unit-hypersphere
- Use: Complex Watson Mixture Model





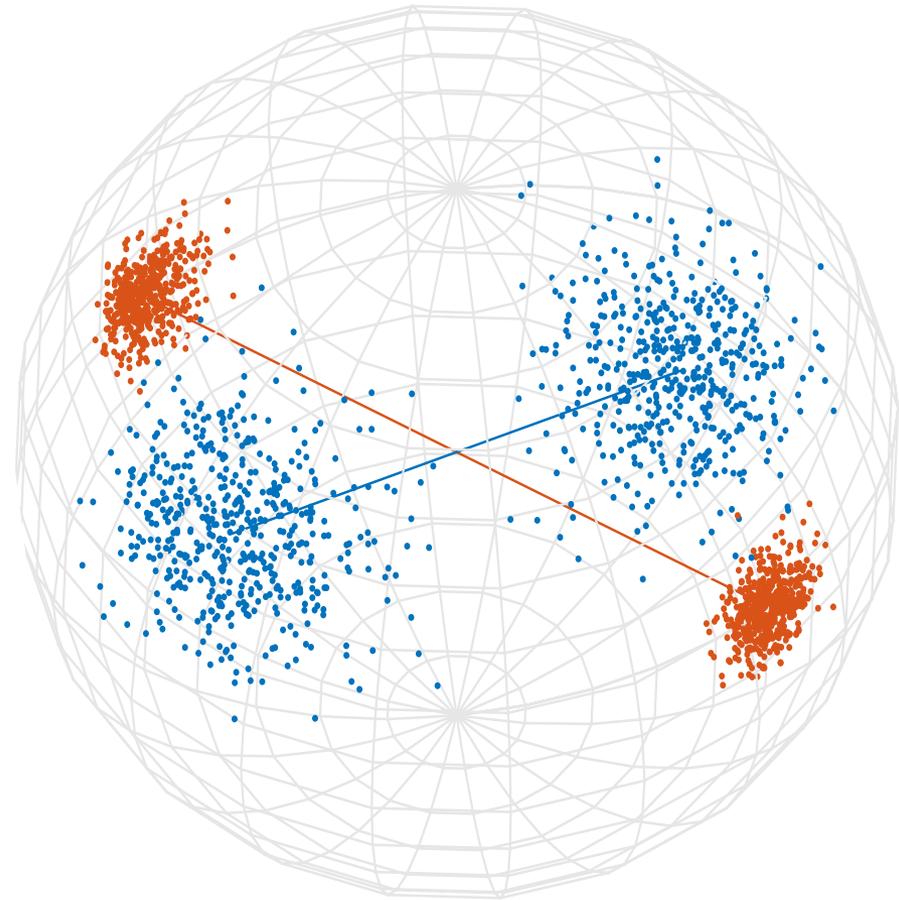
Statistical model: Complex Watson Mixture Model

5

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

$$p(\tilde{\mathbf{Y}}_t | c_t = k) = \frac{1}{c_W(\kappa_k)} e^{\kappa_k |\tilde{\mathbf{Y}}_t^H \mathbf{w}_k|^2}$$

$$p(\mathbf{Y}_t) = p(\mathbf{Y}_t e^{j\varphi})$$



- Use EM algorithm to estimate parameters



EM algorithm

6

- Estimation step:
 - Calculate class affiliation posterior

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

- Maximization step:

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

$$\mathbf{W}_k = \mathcal{P}\{\Phi_k\}, \quad \text{where } \Phi_k = \frac{1}{N_k} \sum_{t=1}^T \gamma_{tk} \tilde{\mathbf{Y}}_t \tilde{\mathbf{Y}}_t^H$$

κ_k with implicit equation



Spherical k-mode clustering

7

- Quantization of the posteriors

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

- Equal mixture weights, shared concentrations

$$\begin{aligned} k &= \operatorname{argmax}_{\tilde{k}} \pi_{\tilde{k}} \frac{1}{c_W(\kappa_{\tilde{k}})} e^{\kappa_{\tilde{k}} |\tilde{\mathbf{Y}}_t^H \mathbf{w}_{\tilde{k}}|^2} \\ &= \operatorname{argmax}_{\tilde{k}} e^{|\tilde{\mathbf{Y}}_t^H \mathbf{w}_{\tilde{k}}|^2} \\ &= \operatorname{argmax}_{\tilde{k}} |\tilde{\mathbf{Y}}_t^H \mathbf{w}_{\tilde{k}}|^2 \end{aligned}$$



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

9

	EM	Spherical k-mode	k-means
Likelihood calculations	T K I	0	0
Eigenvalue decomposition	K I	K I	0
Implicit equations	K I	0	0

T: Time frames

K: Sources

I: EM iterations



Comparison: Duality

10

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

11

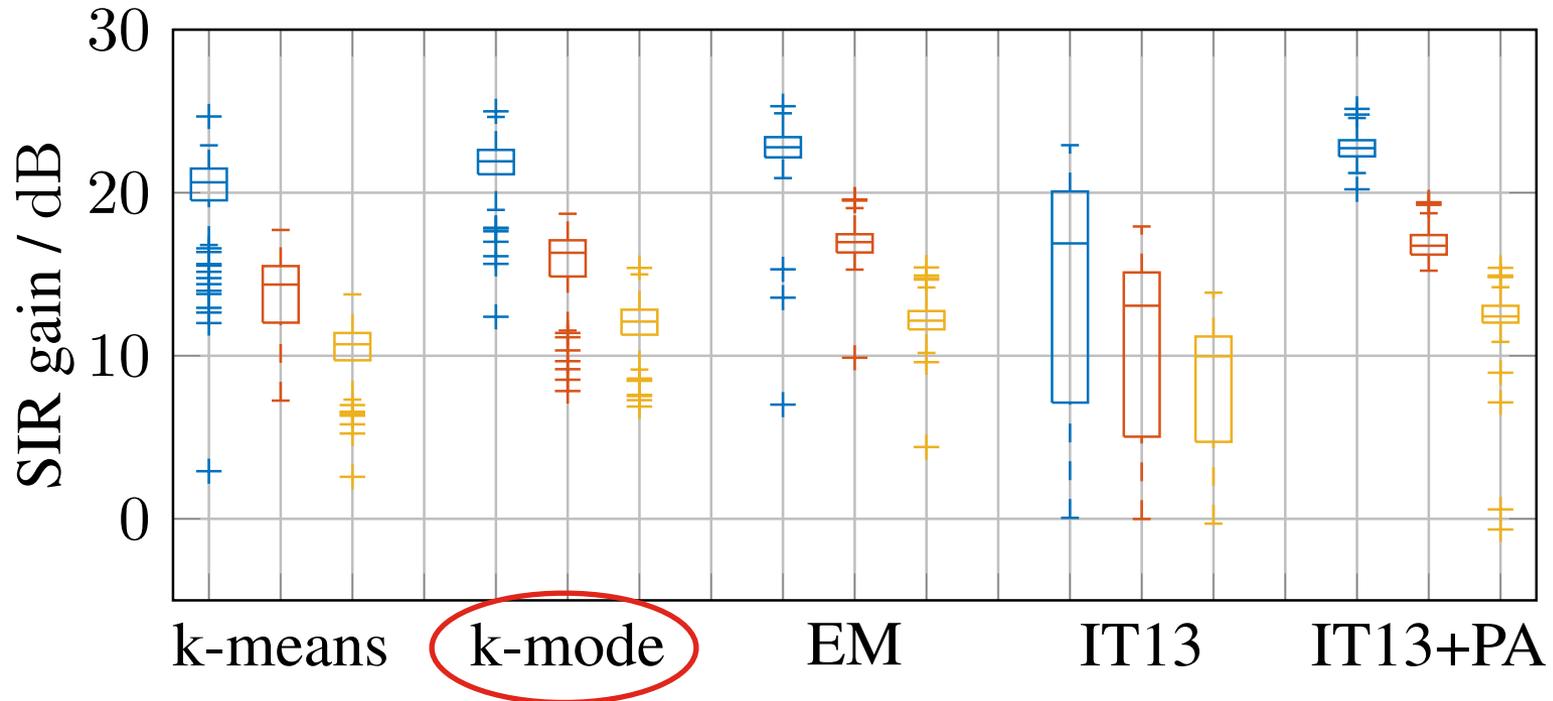
- 5 s utterances from TIMIT DB
- Room impulse responses from MIRD DB
- Spherically isotropic noise
- LCMV-Beamformer to separate sources with given masks
 - PSD matrix from mask
 - Relative transfer functions using Mode Vectors



Hadad et al. 2014

Results

- SNR 15 dB, varying T_{60}



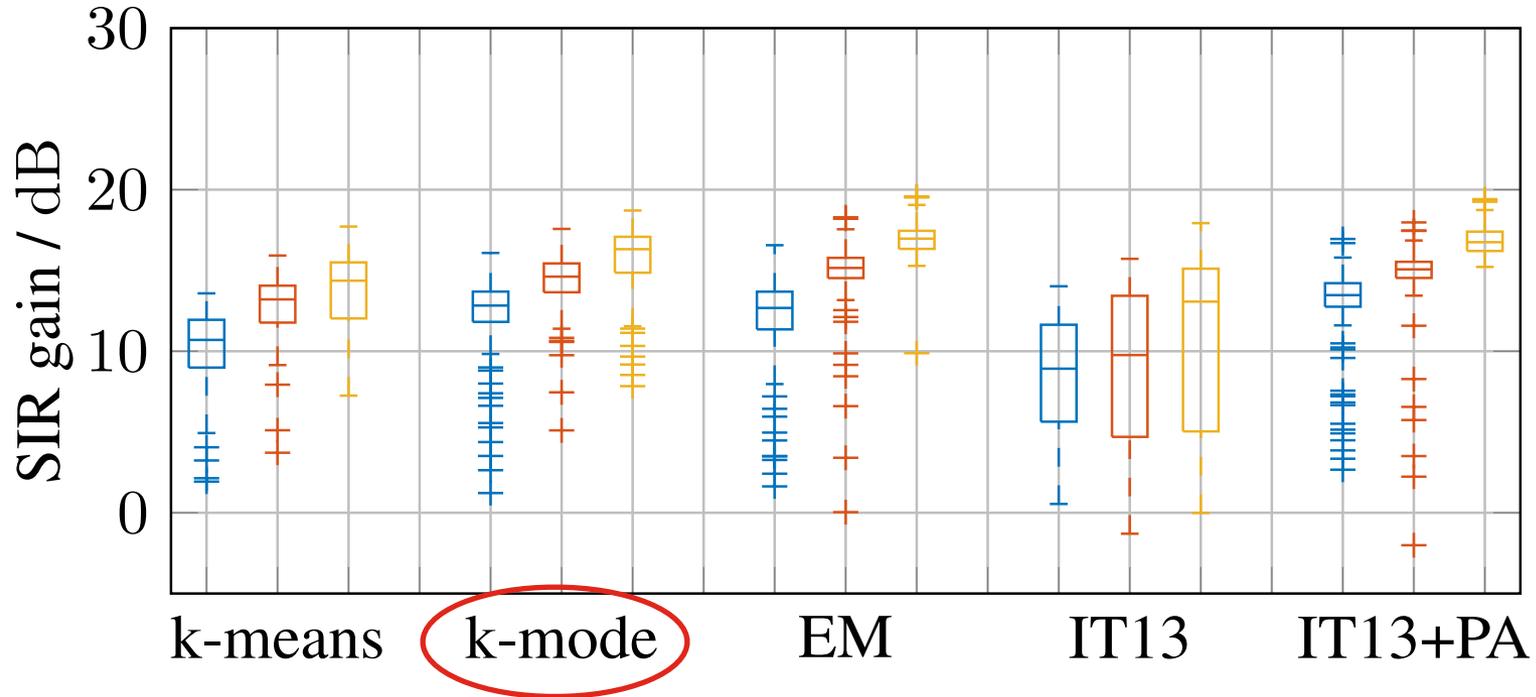
Complexity:

— 160 ms — 360 ms — 610 ms

- 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



Complexity:

— 5 dB — 10 dB — 15 dB

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



Summary

14

- 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!