Probabilistic Segmentation of Auditory Cues based on a Mixture of von Mises Distributions

8th AABBA General Meeting

Christopher Schymura, Dorothea Kolossa

January 22, 2016

Outline

Introduction

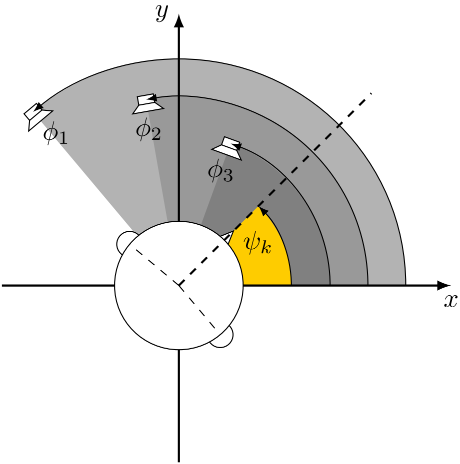


Figure: Application scenario: segmentation of auditory cues for multiple sound sources at different angular positions w.r.t. the listener.

Introduction

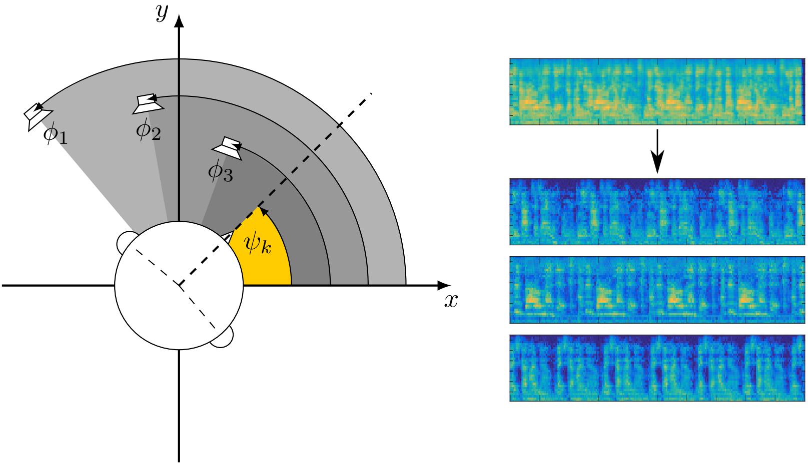


Figure: Application scenario: segmentation of auditory cues for multiple sound sources at different angular positions w.r.t. the listener.

Introduction

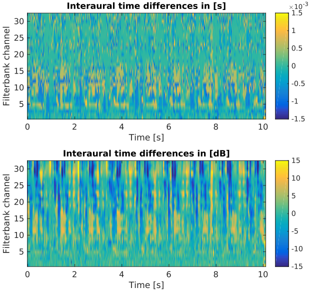


Figure: Distribution of binaural cues and estimated relative azimuth angles for three speech sources positioned at -60°, 0° and 60° w.r.t. the look direction.

Introduction

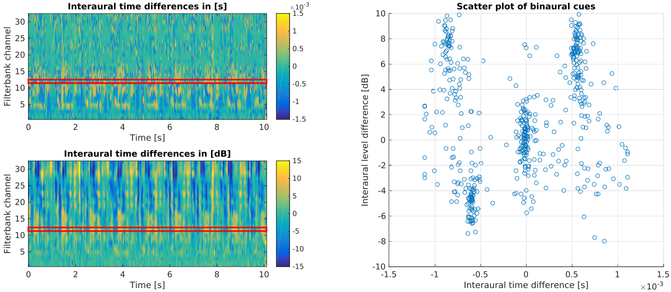


Figure: Distribution of binaural cues and estimated relative azimuth angles for three speech sources positioned at -60°, 0° and 60° w.r.t. the look direction.

Introduction

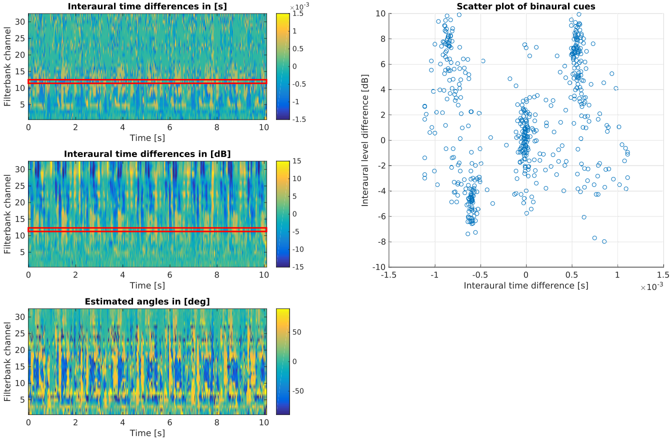


Figure: Distribution of binaural cues and estimated relative azimuth angles for three speech sources positioned at -60°, 0° and 60° w.r.t. the look direction.

Introduction

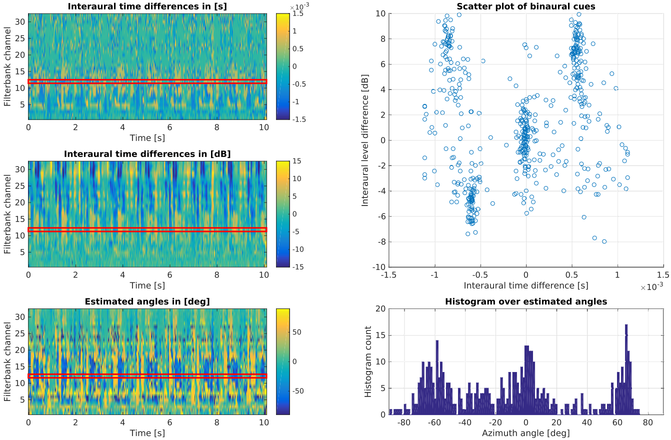
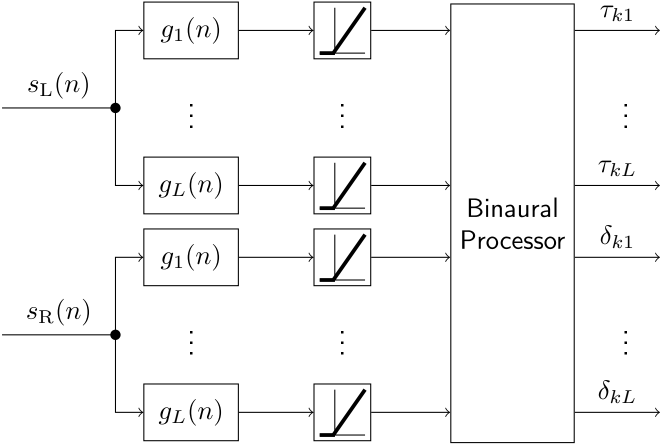


Figure: Distribution of binaural cues and estimated relative azimuth angles for three speech sources positioned at -60°, 0° and 60° w.r.t. the look direction.

Binaural Front-End



See also: `http://twoears.aipa.tu-berlin.de/doc/1.0/afe/`

Binaural Front-End

Mapping binaural cues to relative azimuth angles using polynomial regression:

$$\begin{aligned}\phi_{kl} &= w_0^{(l)} + \sum_{i=1}^{P} w_i^{(l)} \tau_{kl}^{i} + \sum_{j=1}^{P} w_{P+j}^{(l)} \delta_{kl}^{j} \\ &= \boldsymbol{w}_l^T \boldsymbol{x}_{kl}\end{aligned}$$

with $\quad \boldsymbol{w}_l^T = \begin{bmatrix} w_0^{(l)} & \cdots & w_{2P}^{(l)} \end{bmatrix}^T, \quad \boldsymbol{x}_{kl} = \begin{bmatrix} 1 & \tau_{kl} & \cdots & \tau_{kl}^{P} & \delta_{kl} & \cdots & \delta_{kl}^{P} \end{bmatrix}^T$

Binaural Front-End

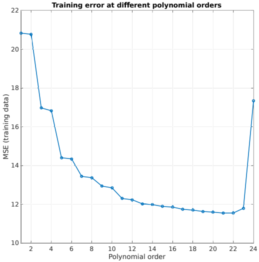
Mapping binaural cues to relative azimuth angles using polynomial regression:

$$\phi_{kl} = w_0^{(l)} + \sum_{i=1}^{P} w_i^{(l)} \tau_{kl}^i + \sum_{j=1}^{P} w_{P+j}^{(l)} \delta_{kl}^j$$
$$= \boldsymbol{w}_l^T \boldsymbol{x}_{kl}$$

with $\boldsymbol{w}_l^T = \left[w_0^{(l)} \quad \cdots \quad w_{2P}^{(l)} \right]^T, \quad \boldsymbol{x}_{kl} = \left[1 \quad \tau_{kl} \quad \cdots \quad \tau_{kl}^P \quad \delta_{kl} \quad \cdots \quad \delta_{kl}^P \right]^T$

Training setup:

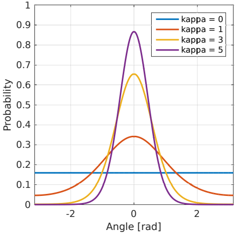
- Anechoic HRTF's (KEMAR dummy head)
- 180 relative azimuth angles (1° increment)
- White noise as stimulus signal
- Individual models are trained for each filterbank channel/center frequency



Probabilistic Circular Clustering

The von Mises distribution:

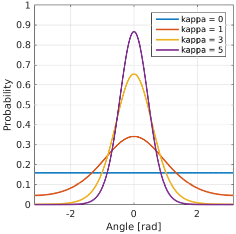
$$\mathcal{VM}(\phi \,|\, \mu,\, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\left\{\kappa \cos(\phi - \mu)\right\}$$



Probabilistic Circular Clustering

The von Mises distribution:

$$\mathcal{VM}(\phi \,|\, \mu, \, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\left\{\kappa \cos(\phi - \mu)\right\}$$



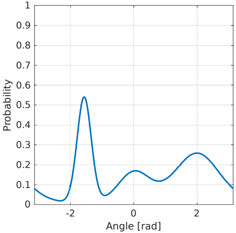
Mixture of von Mises distributions:

$$p(\phi \,|\, \boldsymbol{\pi}, \, \boldsymbol{\mu}, \, \boldsymbol{\kappa}) = \sum_{m=1}^{M} \pi_m \mathcal{VM}(\phi \,|\, \mu_m, \, \kappa_m)$$

$$\boldsymbol{\pi} = \left[\pi_1, \, \ldots, \, \pi_M\right]^T$$

$$\boldsymbol{\mu} = \left[\mu_1, \, \ldots, \, \mu_M\right]^T$$

$$\boldsymbol{\kappa} = \left[\kappa_1, \, \ldots, \, \kappa_M\right]^T$$



Probabilistic Circular Clustering

Expectation maximization for circular clustering:

Inputs:

- Number of sound sources M
- Estimated target source azimuth ϕ_T
- Estimated azimuth angles for all T-F units as a vector $\boldsymbol{\phi} \in \mathbb{R}^{N_\mathrm{S}}$, $N_\mathrm{S} = K \cdot L$

Initialization: Run circular k-means to initialize π_m, μ_m, κ_m and γ_{im}

repeat

E-Step:

Compute responsibilities $\gamma_{im} = \frac{\pi_m \mathcal{VM}(\phi_i \,|\, \mu_m, \kappa_m)}{\sum_{j=1}^{M} \pi_j \mathcal{VM}(\phi_i \,|\, \mu_j, \kappa_j)}$

M-Step:

Re-estimate circular means:

$$\mu_m = \begin{cases} \phi_T, & \text{if } m = 1 \\ \operatorname{atan2}\Big(\sum_{i=1}^{N_\mathrm{s}} \gamma_{im} \sin(\phi_i), \sum_{i=1}^{N_\mathrm{s}} \gamma_{im} \cos(\phi_i) \Big), & \text{otherwise} \end{cases}$$

Re-estimate concentration parameters $\kappa_m = A^{-1}\left(\frac{\sum_{i=1}^{N_\mathrm{s}} \gamma_{im} \cos(\phi_i - \mu_m)}{\sum_{i=1}^{N_\mathrm{s}} \gamma_{im}} \right)$

Re-estimate mixture proportions $\pi_m = \frac{1}{N_\mathrm{s}} \sum_{i=1}^{N_\mathrm{s}} \gamma_{im}$

Evaluate the log-likelihood $\mathcal{L}(\boldsymbol{\phi} \,|\, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\kappa}) = \log \Big(p(\boldsymbol{\phi} \,|\, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\kappa}) \Big)$

until $\mathcal{L}(\boldsymbol{\phi} \,|\, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\kappa})$ converges

Probabilistic Circular Clustering

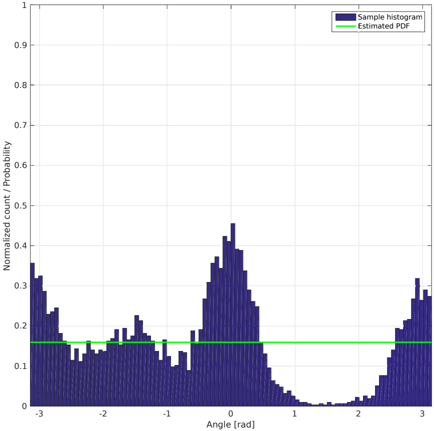


Figure: Starting point for EM algorithm ($\kappa_i = 0 \; \forall \, i$).

Probabilistic Circular Clustering

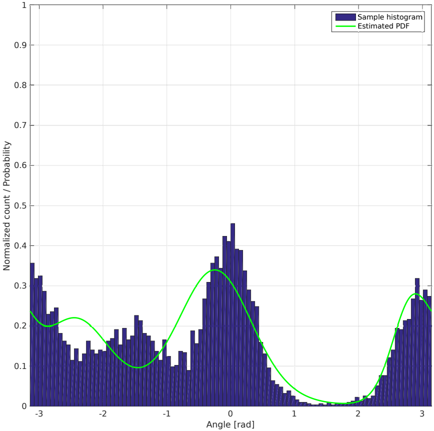


Figure: EM algorithm initialized with circular k-means.

Probabilistic Circular Clustering

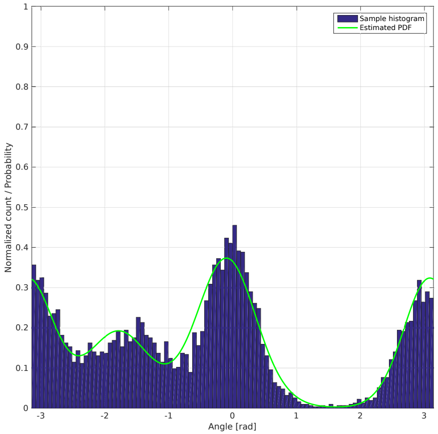


Figure: EM algorithm after one iteration.

Probabilistic Circular Clustering

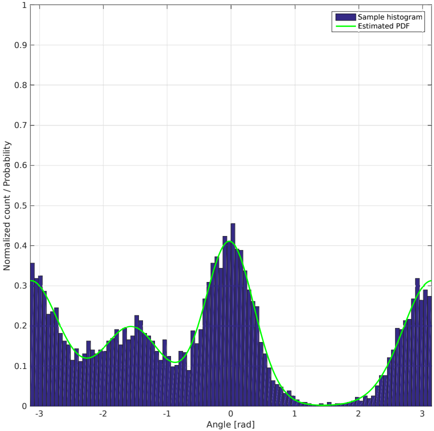


Figure: EM algorithm after five iterations.

Probabilistic Circular Clustering

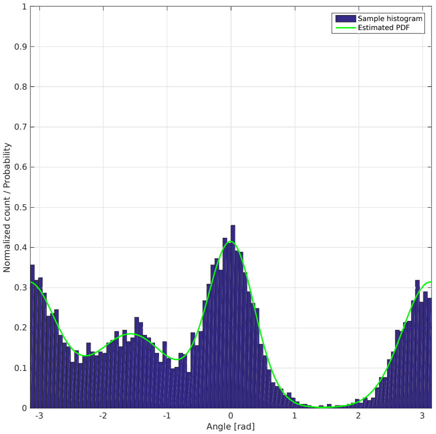


Figure: EM algorithm after ten iterations.

Probabilistic Circular Clustering

Soft-mask computation:

- The EM algorithm is used to estimate model parameters $\boldsymbol{\pi}$, $\boldsymbol{\mu}$, $\boldsymbol{\kappa}$.

Probabilistic Circular Clustering

Soft-mask computation:

- The EM algorithm is used to estimate model parameters $\boldsymbol{\pi}$, $\boldsymbol{\mu}$, $\boldsymbol{\kappa}$.
- After EM has converged, weights can be computed for each T-F unit:

$$\alpha_{kl}^{(m)} = \frac{\pi_m \mathcal{VM}(\phi_{kl} \,|\, \mu_m, \, \kappa_m)}{\sum_{i=1}^{M} \pi_i \mathcal{VM}(\phi_{kl} \,|\, \mu_i, \, \kappa_i)}$$

Probabilistic Circular Clustering

Soft-mask computation:

- The EM algorithm is used to estimate model parameters $\boldsymbol{\pi}$, $\boldsymbol{\mu}$, $\boldsymbol{\kappa}$.
- After EM has converged, weights can be computed for each T-F unit:

$$\alpha_{kl}^{(m)} = \frac{\pi_m \mathcal{VM}(\phi_{kl} \,|\, \mu_m,\, \kappa_m)}{\sum_{i=1}^{M} \pi_i \mathcal{VM}(\phi_{kl} \,|\, \mu_i,\, \kappa_i)}$$

- Hence, the soft-mask for the m-th source is specified as

$$\boldsymbol{A}_m = \begin{bmatrix} \alpha_{11}^{(m)} & \cdots & \alpha_{K1}^{(m)} \\ \vdots & \ddots & \vdots \\ \alpha_{1L}^{(m)} & \cdots & \alpha_{KL}^{(m)} \end{bmatrix}$$

Evaluation and Outlook

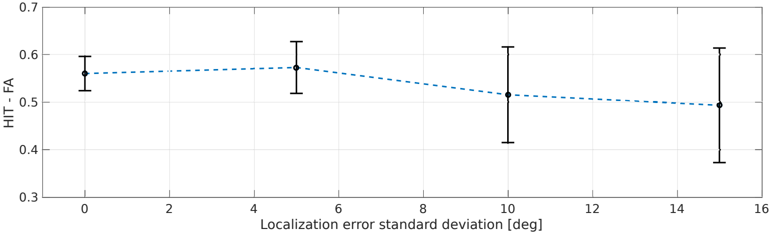
Experimental setup:

- One target source (speech) at 0° relative azimuth
- Two maskers (white noise) at -60° and 60° relative azimuth and 0dB SNR
- Anechoic conditions (KEMAR HRTF's from [Wierstorf et al. (2011)])
- Relative azimuth of the target source is assumed to be known, but localization performance is artificially degraded: $\phi_T + \mathcal{N}(0, \sigma_\phi^2)$

Evaluation and Outlook

Experimental setup:

- One target source (speech) at $0°$ relative azimuth
- Two maskers (white noise) at $-60°$ and $60°$ relative azimuth and 0dB SNR
- Anechoic conditions (KEMAR HRTF's from [Wierstorf et al. (2011)])
- Relative azimuth of the target source is assumed to be known, but localization performance is artificially degraded: $\phi_T + \mathcal{N}(0, \sigma_\phi^2)$



Evaluation and Outlook

Further steps:

- Extensive evaluation of the proposed framework with different sound types and more challenging acoustic conditions.

Evaluation and Outlook

Further steps:

- Extensive evaluation of the proposed framework with different sound types and more challenging acoustic conditions.
- Evaluation of system performance in combination with different source localization techniques.

Evaluation and Outlook

Further steps:

- Extensive evaluation of the proposed framework with different sound types and more challenging acoustic conditions.
- Evaluation of system performance in combination with different source localization techniques.
- Integration of head movements into the framework: can they help to improve segmentation performance?

Evaluation and Outlook

Further steps:

- Extensive evaluation of the proposed framework with different sound types and more challenging acoustic conditions.
- Evaluation of system performance in combination with different source localization techniques.
- Integration of head movements into the framework: can they help to improve segmentation performance?

Interested? ;-)

Get the demo: `www.twoears.eu`

Evaluation and Outlook

Further steps:

- Extensive evaluation of the proposed framework with different sound types and more challenging acoustic conditions.
- Evaluation of system performance in combination with different source localization techniques.
- Integration of head movements into the framework: can they help to improve segmentation performance?

Interested? ;-)

Get the demo: `www.twoears.eu`

This research has been supported by EU FET grant Two!Ears, ICT-618075.