SOURCE COUNTING IN SPEECH MIXTURES BY NONPARAMETRIC BAYESIAN ESTIMATION OF AN INFINITE GAUSSIAN MIXTURE MODEL

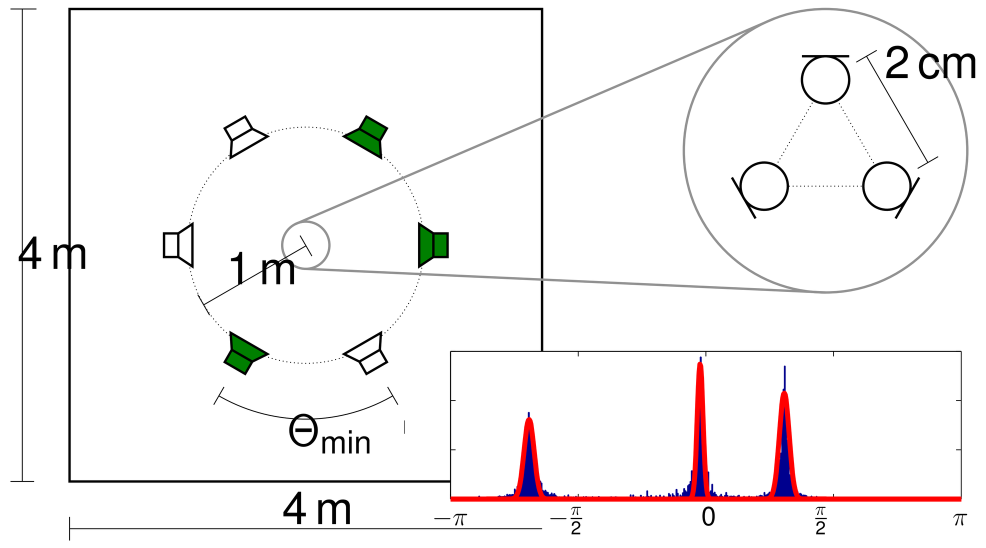
UNIVERSITÄT PADERBORN
Die Universität der Informationsgesellschaft

Computer Science, Electrical Engineering and Mathematics
NT Communications Engineering
Prof. Dr.-Ing. Reinhold Häb-Umbach

Oliver Walter, Lukas Drude and Reinhold Haeb-Umbach, University of Paderborn, Germany
{walter,drude,haeb}@nt.uni-paderborn.de, http://nt.uni-paderborn.de/

Introduction

- **Objective:** Source counting with microphone array
  - **Input:** Direction of arrival estimates (DOAs)
  - **Goal:** Automatically estimate number of speakers



- **Problem:** Model DOA histogram as mixture model
  - **Model selection:** Determine number of mixtures
  - **Circular distribution:** 2π periodicity in observations
- **Approach:** Nonparametric Bayesian modeling
  - Mixture of wrapped Gaussian distributions
    - ⇒ Each speaker represented by one mixture component
  - Dirichlet process prior over mixture components
    - ⇒ Automatically determine number of mixture components
  - Weighting of observations by their power

Mixture of wrapped Gaussians

- Shift conditional Gaussian distribution

$$p\left(d_n|\mu_l,\sigma_l^2,k_n\right)=\frac{1}{\sqrt{2\pi\sigma_l^2}}\exp\left(\frac{-(d_n+2\pi k_n-\mu_l)^2}{2\sigma_l^2}\right)$$

- Mixture of wrapped Gaussian distributions

$$p\left(d_n|\boldsymbol{\mu},\boldsymbol{\sigma}^2\right)=\sum_{l=1}^{L}P(z_n=l)\sum_{k_n=-\infty}^{\infty}p\left(d_n|\mu_l,\sigma_l^2,k_n\right)$$

- Normal gamma prior for precision σ_l^{-2} and mean μ_l

$$\boldsymbol{\Theta}_l=\{\mu_l,\sigma_l^{-2}\}\sim\mathcal{N}\left(\mu;m^{(0)},\sigma^2/\xi^{(0)}\right)\mathcal{G}\left(\sigma^{-2};\eta^{(0)},r^{(0)}\right)$$

- Hyper parameters: $\boldsymbol{\Theta}^{(0)}=\{m^{(0)},\xi^{(0)},\eta^{(0)},r^{(0)}\}$

Dirichlet process prior

- Prior over infinitely many mixture components
- Chinese restaurant process representation:
  - Probability for assignment to existing mixture component

$$P(z_n=l|\mathbf{z}_{\backslash n})\propto s_{0,l}$$

  - Probability for assignment to new mixture component

$$P(z_n=l_{\mathrm{new}}|\mathbf{z}_{\backslash n})\propto\gamma$$

- $s_{0,l}$: sum over observation weights belonging to mixture l
- γ: concentration parameter controls weight of new mixture

Bayesian inference: Gibbs sampling

- Sequential sampling from conditional distributions
- **Step 1**: Sample mixture indicator from

$$P(z_n=l|d_n,\mathbf{d}_{\backslash n},\mathbf{z}_{\backslash n},\mathbf{k}_{\backslash n},\boldsymbol{\Theta}^{(0)})$$
$$\propto P(z_n=l|\mathbf{z}_{\backslash n})\quad\underbrace{p(d_n|\mathbf{d}_{\backslash n},z_n=l,\mathbf{z}_{\backslash n},\mathbf{k}_{\backslash n},\boldsymbol{\Theta}^{(0)})}_{\text{Collapsed Gibbs sampling - Integrate out prior:}}$$

$$\sum_{k_n=-K}^{K}\int p(d_n|\boldsymbol{\Theta}_l,k_n)p(\boldsymbol{\Theta}_l|\mathbf{d}_{\backslash n},z_n=l,\mathbf{z}_{\backslash n},\mathbf{k}_{\backslash n},\boldsymbol{\Theta}^{(0)})\mathrm{d}\boldsymbol{\Theta}_l$$

$$\Rightarrow\text{Student's t-distributions}\sum_{k_n=-K}^{K}\mathcal{T}(d_n+2\pi k_n;m_l,\xi_l,\eta_l,r_l)$$

  - For $z_n=l_{\mathrm{new}}$ use $m^{(0)},\xi^{(0)},\eta^{(0)},r^{(0)}$
- **Step 2**: Decide for that shift k_n where summand is maximal
- **Step 3**: Weighted update of parameters $m_l,\xi_l,\eta_l,r_l,s_{0,l}$

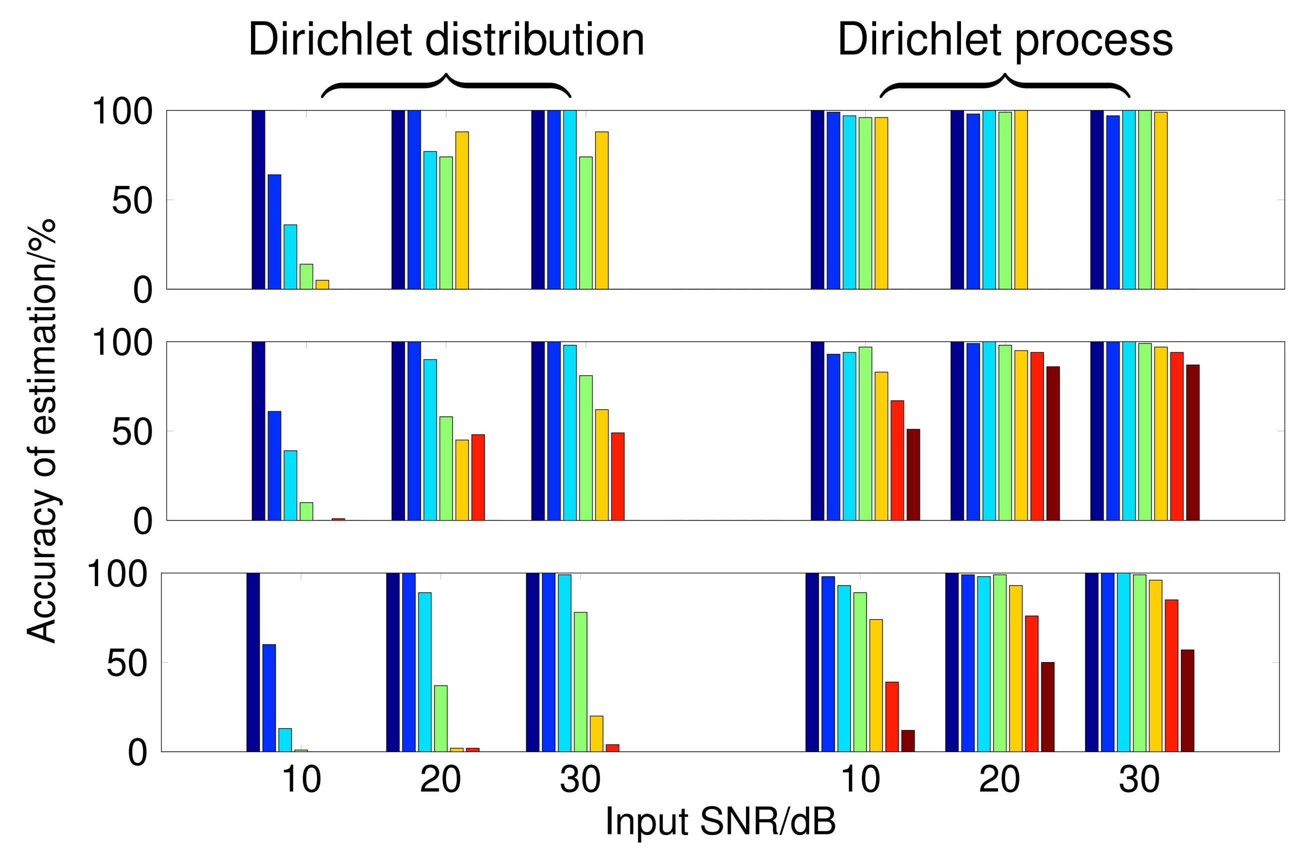
Source counting

- Consolidate mixture components:
  - Reduce γ after burn-in period
  - Remove, if probability of mean higher in other mixture
  - Remove, if variance 10x higher than lowest variance

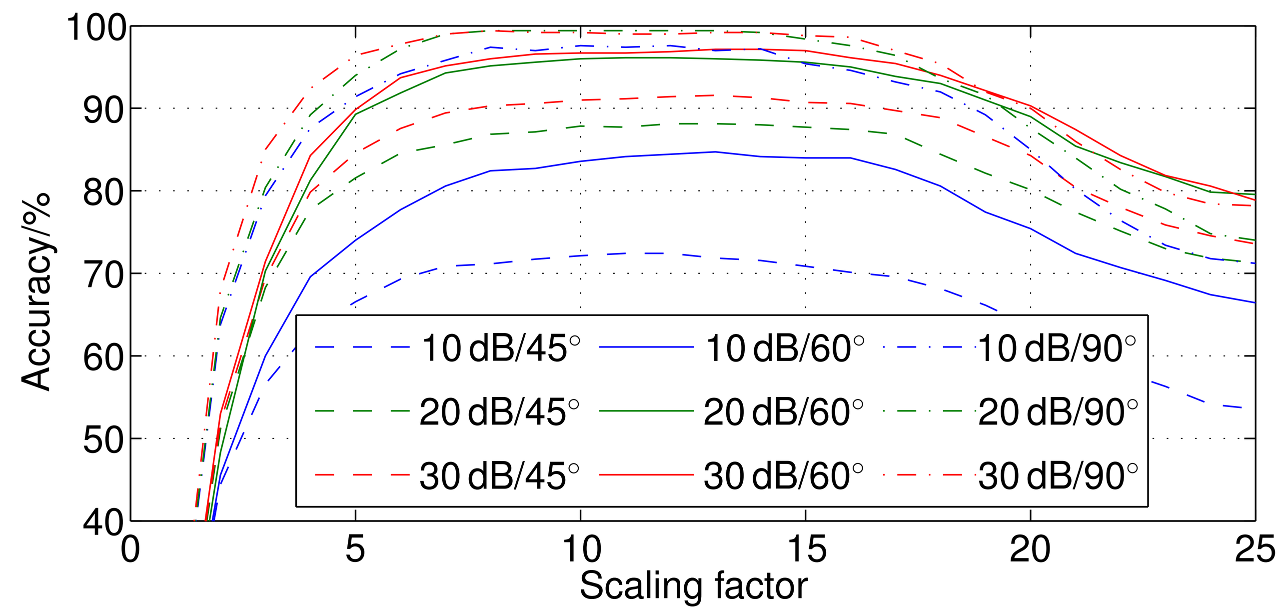
⇒ Remaining number of mixtures is number of speakers

Experimental results

- Simulated anechoic speech mixtures of 5 s at 16 kHz
- Maximum number of speakers/minimal speaker spacing
  - Top to bottom: 4/90°, 6/60°, 6/45°
- Number of active speakers
  - Left to right or blue to red: 0, 1, 2, 3, 4, 5, 6
- Input SNRs 10 dB, 20 dB, 30 dB
- Comparison with Dirichlet distribution prior



- Low sensitivity over wide range of variance floor scaling



Conclusions

- Dirichlet process prior delivers better result
- Adaptive variance thresholding
- Dirichlet process prior for other mixture models
  - Infinite complex Watson mixture model?

Work supported by DFG under contract no. Ha3455/9-1, Priority Program SPP1527 "Autonomous Learning"