Informed TDoA-based Direction of Arrival Estimation for Hearing Aid Applications

Mojtaba Farmani
Michael Syskind Pedersen
Zheng-Hua Tan
Jesper Jensen

GlobalSIP 2015

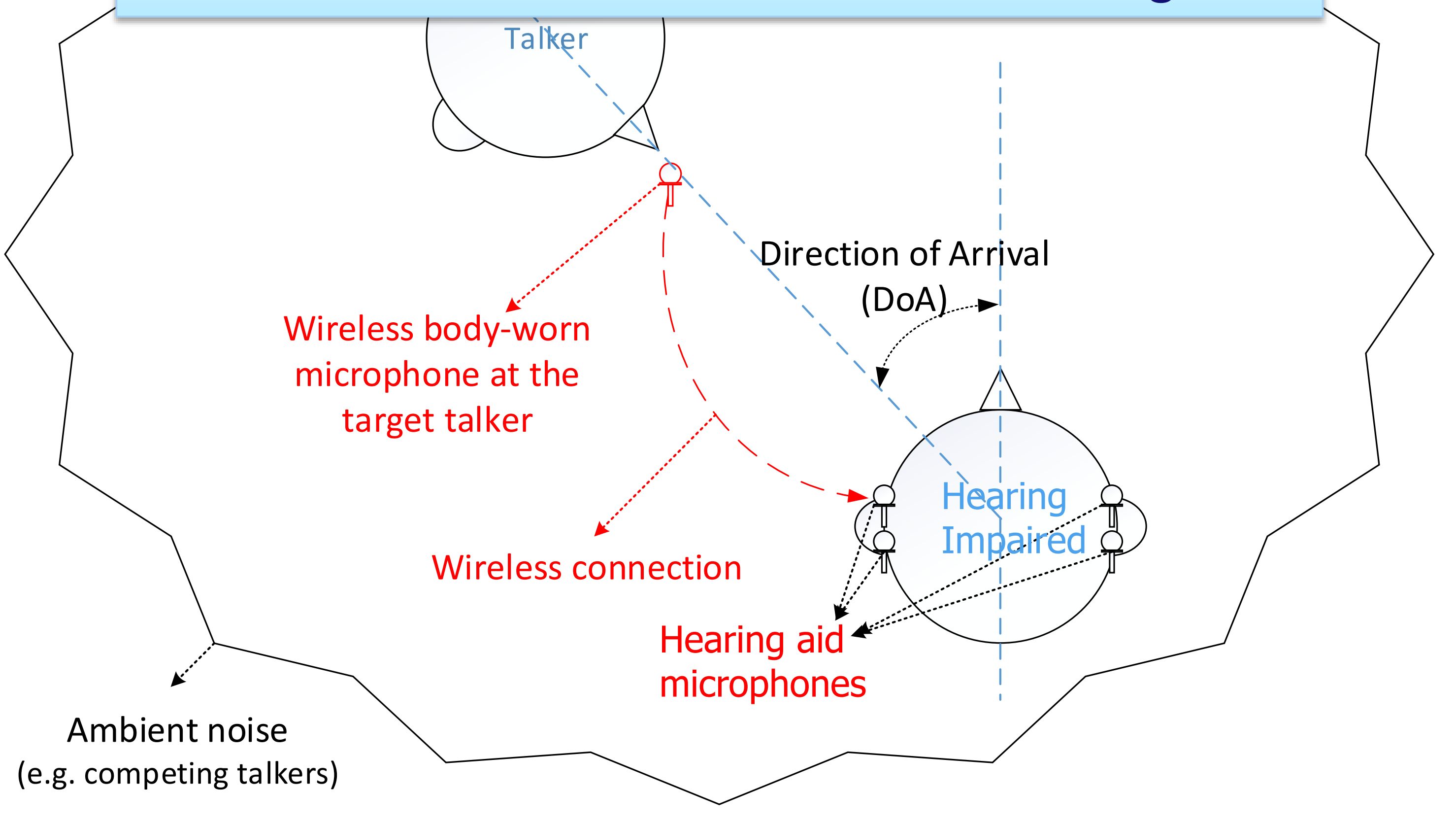
oticon PEOPLE FIRST

AALBORG UNIVERSITY DENMARK

Content

Why do we need DoA estimation if the noise-free signal is available?

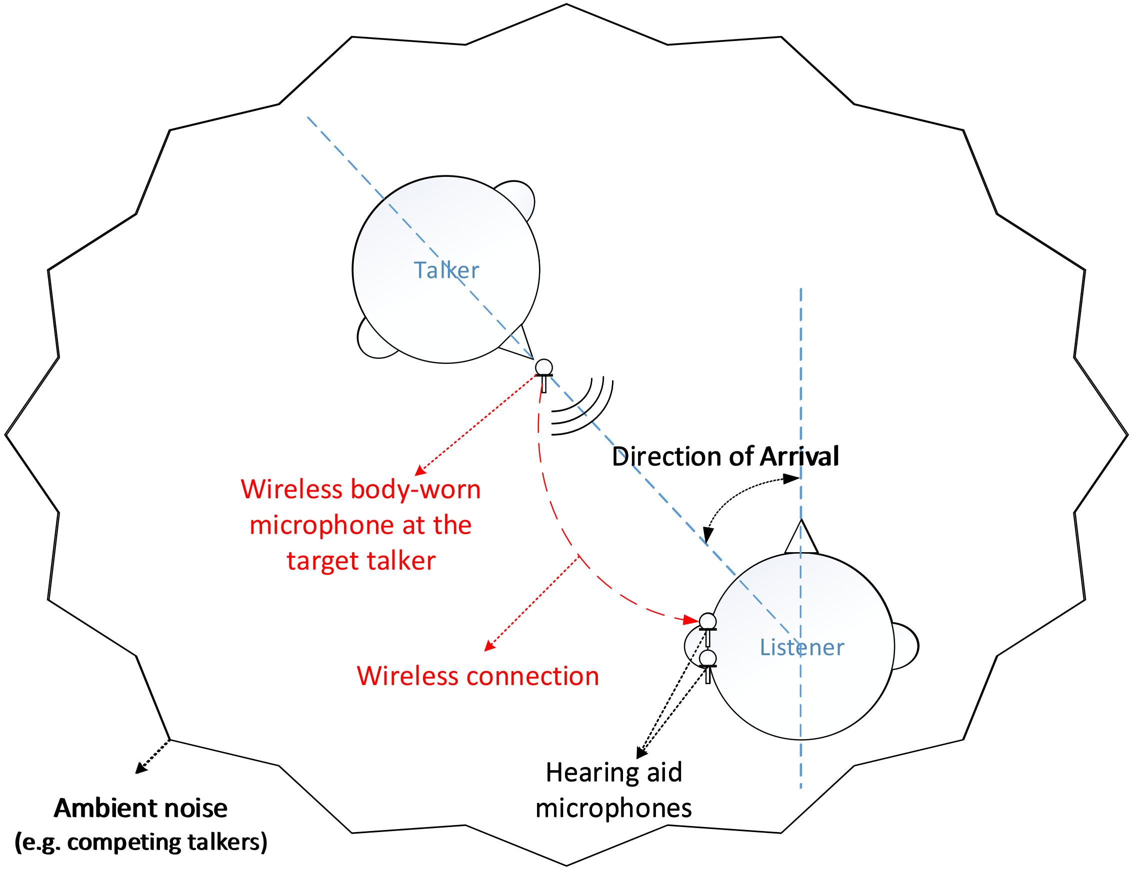
Binauralization of the noise-free signal



Introduction

DoA estimation algorithms:

- "Uninformed"
- "Informed"



Introduction

- Contribution:

Proposing a TDoA-based DoA estimator for the

”informed” Source Localization problem

via a Maximum Likelihood Approach.

Content

Signal Model (Time Domain)

$$r_m(n) = s(n) * h_m(n) + v_m(n)$$

- $r_m(n)$: noisy received signal at microphone m.
- $s(n)$: noise-free target signal emitted at the talker's position.
- $h_m(n)$: the acoustic channel impulse response from the target talker to microphone m.
- $v_m(n)$: additive noise component.

Signal Model (STFT Domain)

- Short time Fourier transform (STFT) domain:
  - Frequency dependent processing
  - Computational efficiency
  - Low latency algorithm implementation
- Time Domain:

$$r_m(n) = s(n) * h_m(n) + v_m(n)$$

- STFT Domain:

$$R_m(l,k) = S(l,k)H_m(k) + V_m(l,k)$$

  - l: frame index.
  - k: frequency bin index.

Signal Model (Vector Representation)

$$\boldsymbol{R}(l,k) = S(l,k)\boldsymbol{H}(k) + \boldsymbol{V}(l,k)$$

- $\boldsymbol{R}(l,k) = [R_1(l,k), R_2(l,k), \dots, R_{\mathrm{M}}(l,k)]^{\mathrm{T}}$.
- $\boldsymbol{H}(k) = [H_1(k), H_2(k), \dots, H_{\mathrm{M}}(k)]^{\mathrm{T}}$.
- $\boldsymbol{V}(l,k) = [V_1(l,k), V_2(l,k), \dots, V_{\mathrm{M}}(l,k)]^{\mathrm{T}}$.

M: # of the considered Hearing Aid Microphones ($\mathrm{M} \geq 1$)

Content

- Introduction
- Signal Model
- Head Model
- Maximum Likelihood Framework
- Proposed DoA Estimator
- Simulation Results
- Conclusion and Future work

Head Model

- **User-Specific (measured HRTF)**

"Maximum Likelihood Approach to "Informed" Sound Source Localization", ICASSP 2015.

- **Spherical-Head Model**

"Informed Direction of Arrival Estimation Using a Spherical-Head Model for Hearing Aid Applications", ICASSP 2016.

- **Free-Field**

"Informed TDoA-based Direction of Arrival Estimation for Hearing Aid Applications", GlobalSIP 2015.

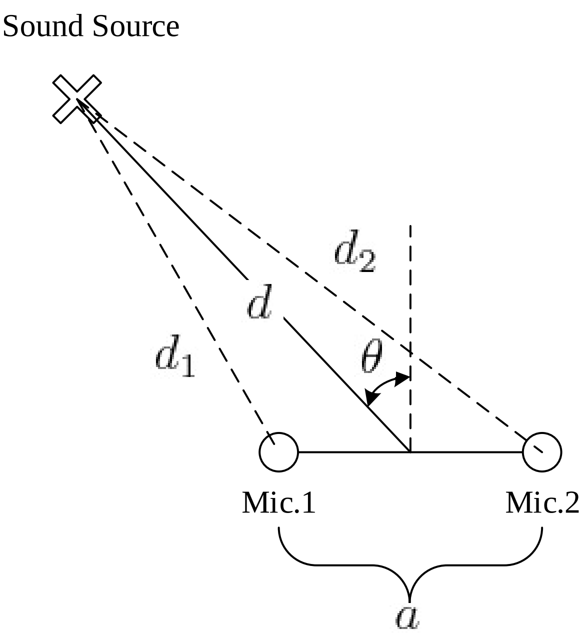
Free-Field and Far-Field Model

- Rely on minimal number of user-specific assumption.

- Acoustic channel Model:

$$H_m(k) = \sum_{n=0}^{N-1} h_m(n) e^{-\frac{j2\pi kn}{N}} = \alpha_m e^{-\frac{j2\pi k}{N} D_m}$$

- Propagation time: $D_1 = \frac{d_1}{c}$, $D_2 = \frac{d_2}{c}$



Free-Field and Far-Field Model

- Propagation time: $D_1 = \frac{d_1}{c},\ D_2 = \frac{d_2}{c}$

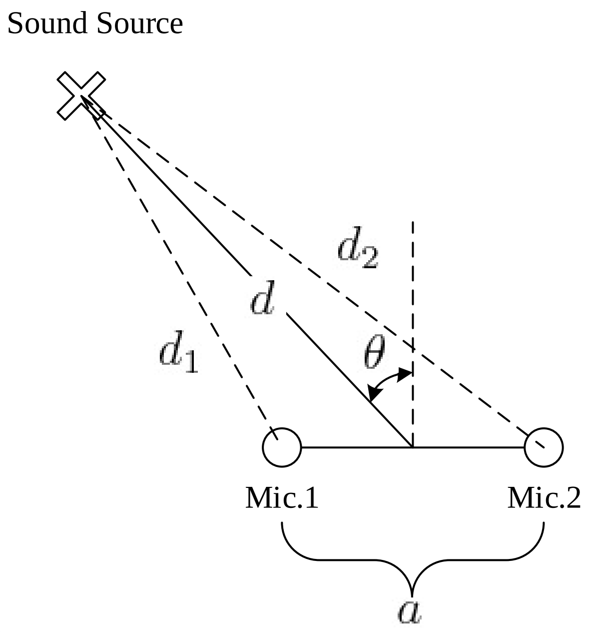
- Interaural Time Difference (ITD):

$$D_1 - D_2 = \frac{a}{c}\sin\theta$$

- Interaural Level Difference (ILD):

$$\mathrm{ILD} = 0\mathrm{dB} \Rightarrow \frac{\alpha_1}{\alpha_2} = 1$$

- DoA: $\theta = \arcsin\left((D_1 - D_2)\frac{c}{a}\right)$



Content

- Introduction
- Signal Model
- Head Model
- Maximum Likelihood Framework
- Proposed DoA Estimator
- Simulation Results
- Conclusion and Future work

Maximum Likelihood Framework

- Assume $\boldsymbol{V}(l,k) \sim \mathcal{N}\left(0, \boldsymbol{C}_v(l,k)\right)$.
- Likelihood function:

$$p(\mathbf{R}(l), \boldsymbol{S}(l), \mathbf{C}_v(l)|\mathbf{H}) = \prod_{k=1}^{K} \frac{1}{\pi^M |\boldsymbol{C}_v(l,k)|} e^{-\boldsymbol{Z}^{\mathrm{H}}(l,k)\boldsymbol{C}_v^{-1}(l,k)\boldsymbol{Z}(l,k)},$$

  - $\boldsymbol{Z}(l,k) = \boldsymbol{R}(l,k) - S(l,k)\boldsymbol{H}(k),$
  - $\boldsymbol{H}(k) = \left[\alpha_1 \mathrm{e}^{-\frac{j2\pi k}{N}D_1}, \ldots, \alpha_{\mathrm{M}} \mathrm{e}^{-\frac{j2\pi k}{N}D_{\mathrm{M}}}\right]^{\mathrm{T}}.$

- Reduced Log-Likelihood Function:

$$\mathcal{L} = \sum_{k=1}^{K} -\boldsymbol{Z}^{\mathrm{H}}(l,k)\boldsymbol{C}_v^{-1}(l,k)\boldsymbol{Z}(l,k)$$

Content

DoA Estimator

- Two Different approaches:

1. Consecutive Estimation: First estimate the ITD by estimating D_1 and D_2 independently, and then estimate the DoA.

2. Joint Estimation: Estimate the ITD and the DoA jointly.

Consecutive Estimation

- Considering the received signal of microphone m, the reduced log-likelihood:

$$\hat{\mathcal{L}}_m(\alpha_m, D_m) = -\sum_{k=1}^{N} \frac{Z_m^*(l,k)Z(l,k)}{c_v(l,k)},$$

$$Z_m(l,k) = R_m(l,k) - S(l,k)\alpha_m \mathrm{e}^{-\frac{j2\pi k}{N}D_m}$$

Consecutive Estimation

- Making $\hat{\mathcal{L}}_m(\alpha_m, D_m)$ independent of α_m:

$$\frac{\partial \hat{\mathcal{L}}_m}{\partial \alpha_m} = 0 \Rightarrow \hat{\alpha}_{\mathrm{MLE}} \rightarrow \hat{\mathcal{L}}_m(\alpha_m, D_m) \Rightarrow$$

$$\tilde{\mathcal{L}}(D_m) = \sum_{k=1}^{N} \frac{1}{c_v(l,k)} S^*(l,k) R_m(l,k) \mathrm{e}^{\frac{j2\pi k}{N} D_m}$$

- $\widehat{D}_m = \underset{D_m}{\operatorname{argmax}}\, \tilde{\mathcal{L}}(D_m)\,, \qquad m = 1,2$

- $\hat{\theta} = \arcsin\left(\left(\widehat{D}_1 - \widehat{D}_2\right) \frac{c}{a} \right)$

Generalized Cross Correlation (GCC)

$$\tilde{\mathcal{L}}(D_m) = \sum_{k=1}^{N} \frac{1}{c_v(l,k)} S^*(l,k) R_m(l,k) \mathrm{e}^{\frac{j2\pi k}{N} D_m}$$

$$\boldsymbol{\mathcal{R}}_{S,R_m}^{\mathrm{GCC}}(D_m) = \sum_{k=1}^{N} \psi(k) S^*(l,k) R_m(l,k) e^{j2\pi \frac{k}{N} D_m}$$

$$\psi(k) = \begin{cases} 1 & \text{Coventional Cross Correlation} \\ \dfrac{1}{|S^*(l,k) R_m(l,k)|} & \text{PHAT} \\ \dfrac{1}{C_v(l,k)} & \text{MaximumLikelihood} \end{cases}$$

Joint Estimation

- Let us consider the received signals of the two binaural microphones jointly:

$$\begin{cases} R_1(l,k) = S(l,k)\alpha_1 \mathrm{e}^{-\frac{j2\pi k}{N}D_1} + V_1(l,k) \\ R_2(l,k) = S(l,k)\alpha_2 \mathrm{e}^{-\frac{j2\pi k}{N}D_2} + V_2(l,k) \end{cases}$$

- $\begin{cases} D_2 = \frac{a}{c} sin\,\theta - D_1 \\ \alpha_1 = \alpha_2 = \alpha \end{cases} \Rightarrow \hat{\mathcal{L}}(\theta,\alpha,D_1)$
- $\frac{\partial \hat{\mathcal{L}}(\theta,\alpha,D_1)}{\partial \alpha} = 0 \Rightarrow \ \hat{\alpha}_{\mathrm{MLE}} \rightarrow \hat{\mathcal{L}}(\theta,\alpha,D_1) \Rightarrow \hat{\mathcal{L}}(\theta,D_1)$
- $\left[\hat{\theta},\widehat{D}_1\right] = \underset{\theta,D_1}{\operatorname{argmax}}\ \hat{\mathcal{L}}(\theta,D_1)$

Joint Estimation

- $\boldsymbol{C}_v^{-1}(l,k) = \begin{bmatrix} C_{11}(l,k) & C_{12}(l,k) \\ C_{21}(l,k) & C_{22}(l,k) \end{bmatrix}$.
- $\hat{\alpha}_{\mathrm{MLE}} = \frac{f(\theta, D_1)}{g(\theta)}$,

$$f(\theta, D_1) = \sum_{k=1}^{N} p(\theta)\, S^*(l,k) \mathrm{e}^{j2\pi \frac{k}{N} D_1},$$

$$p(\theta) = C_{11}R_1 + C_{12}R_2 + (C_{21}R_1 + C_{22}R_2)\mathrm{e}^{\frac{j2\pi k}{N}\left[-\sin(\theta)\frac{a}{c}\right]}.$$

$$g(\theta) = \sum_{k=1}^{N} \left(C_{11} + 2C_{21}\mathrm{e}^{\frac{j2\pi k}{N}\left[-\sin(\theta)\frac{a}{c}\right]} + C_{22} \right) |S(l,k)|^2$$

Joint Estimation

- $\hat{\alpha}_{\text{MLE}} \rightarrow \hat{\mathcal{L}}(\theta, \alpha, D_1) \Rightarrow \hat{\mathcal{L}}(\theta, D_1) = \frac{f^2(\theta, D_1)}{g(\theta)}$.
- For a given θ, computing $\hat{\mathcal{L}}(\theta, D_1)$ results in a discrete-time sequence, where the MLE of D_1 is the time index of the maximum of the sequence.
- θ is unknown ->

  let us consider a discrete set Θ of different θs.

- $[\hat{\theta}, \widehat{D}_1] = \arg\max_{\theta \in \Theta, D_1} \hat{\mathcal{L}}(\theta, D_1)$

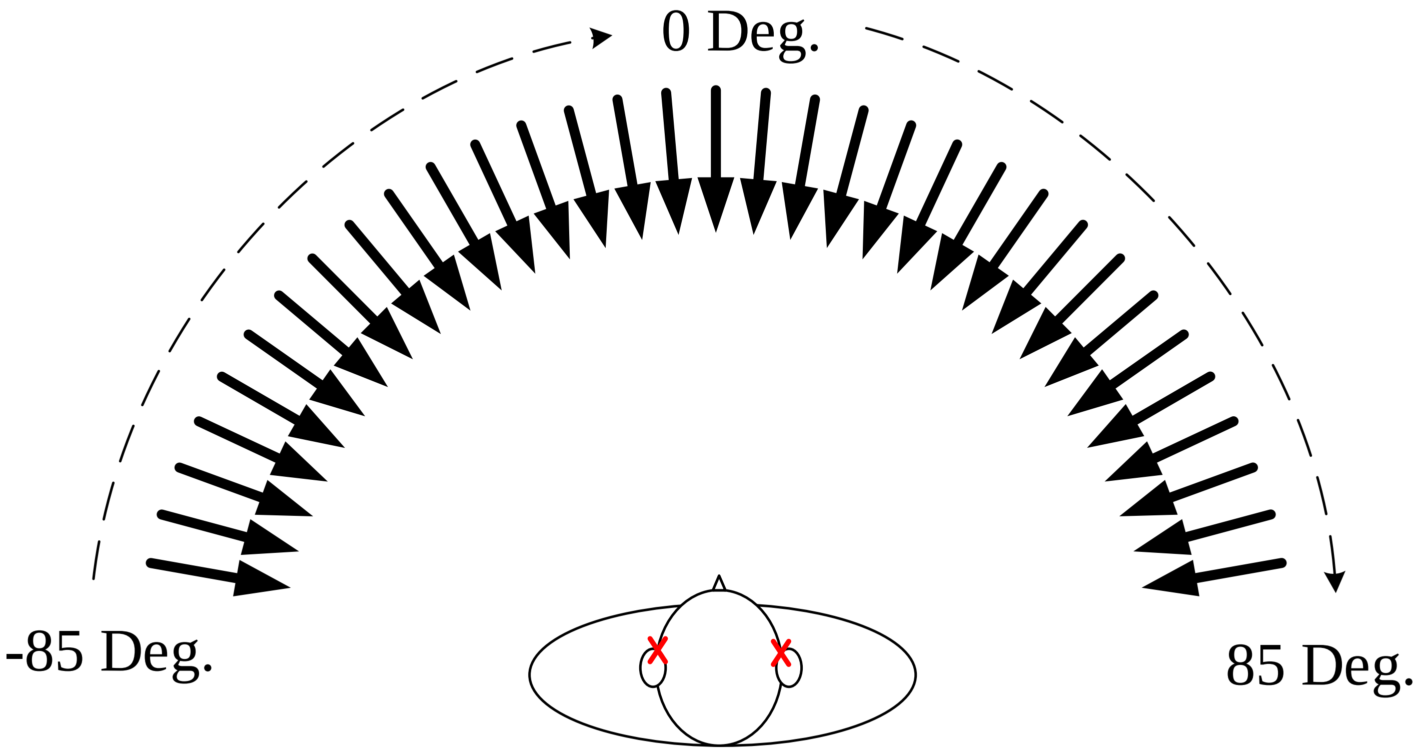
Decrease Computation Overhead

- Let us assume $V_1(l,k)$ and $V_2(l,k)$ are uncorrelated.
- $\boldsymbol{C}_v^{-1}(l,k) = \begin{bmatrix} C_{11}(l,k) & 0 \\ 0 & C_{22}(l,k) \end{bmatrix}$.
- $\hat{\mathcal{L}}(\theta, D_1) =$
$$\sum_{k=1}^{N} \left(C_{11} R_1 + C_{22} R_2 \mathrm{e}^{\frac{j2\pi k}{N}\left[-\sin(\theta)\frac{a}{c}\right]} \right) S^*(l,k) \mathrm{e}^{j2\pi \frac{k}{N} D_1}$$

Content

- Introduction
- Signal Model
- Head Model
- Maximum Likelihood Framework
- Proposed DoA Estimator
- Simulation Results
- Conclusion and Future work

HRTF Measurements



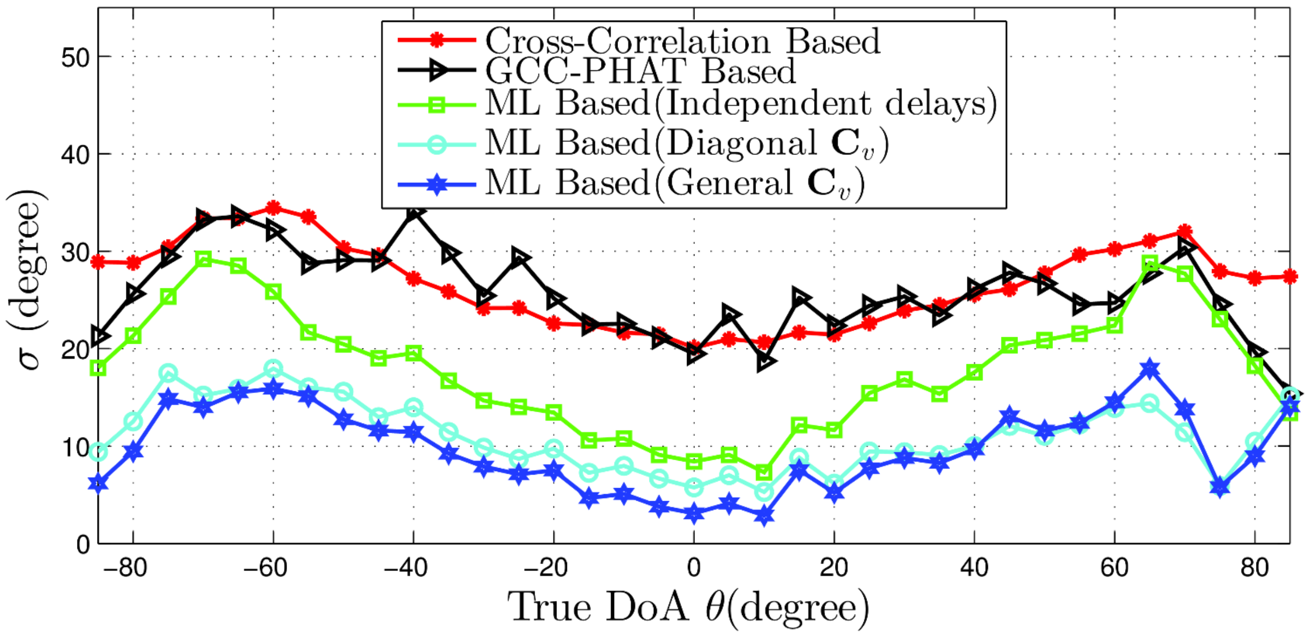
Experiment Parameters

- Sampling frequency: 20 kHz
- Frame length: N = 2048 samples
- Overlapping: A = 1024 samples
- Target Signal: 10-second sample of the ISTS signal

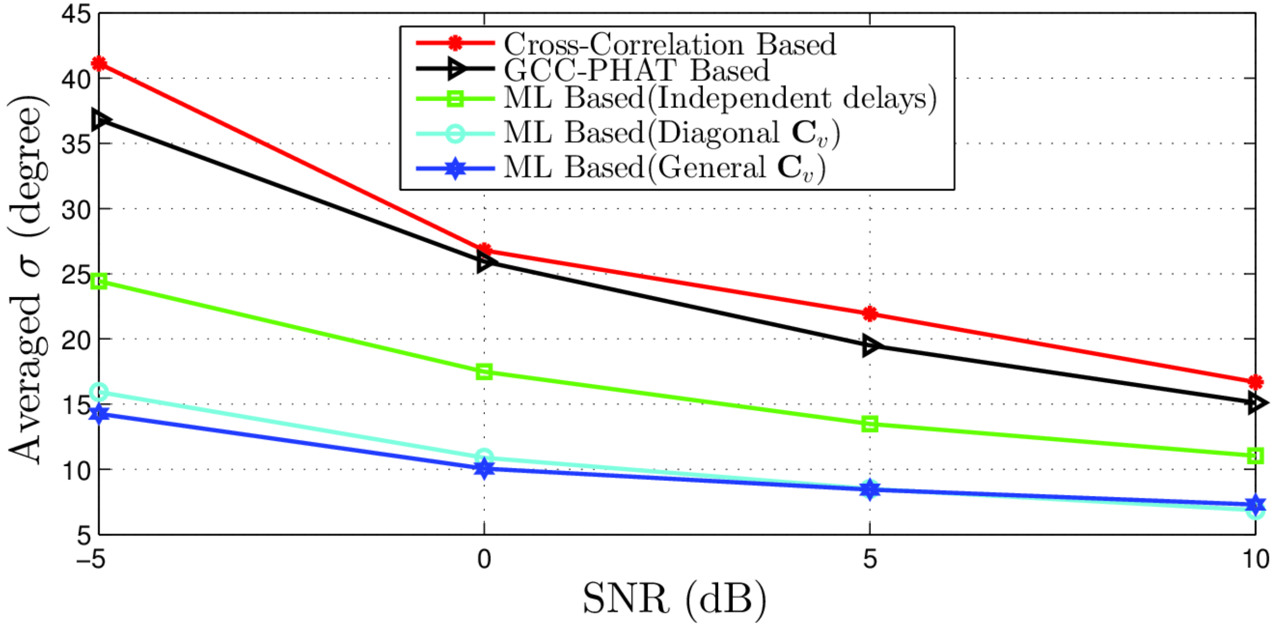
  (21 female voices in 6 different languages)
- Noise type: Large-crowd noise

  (Play back different speech signals of different men and women from each of the target positions simultaneously)
- The mean absolute error (MAE): $\sigma = \frac{1}{L}\sum_{j=1}^{L} |\theta - \hat{\theta}_j|$

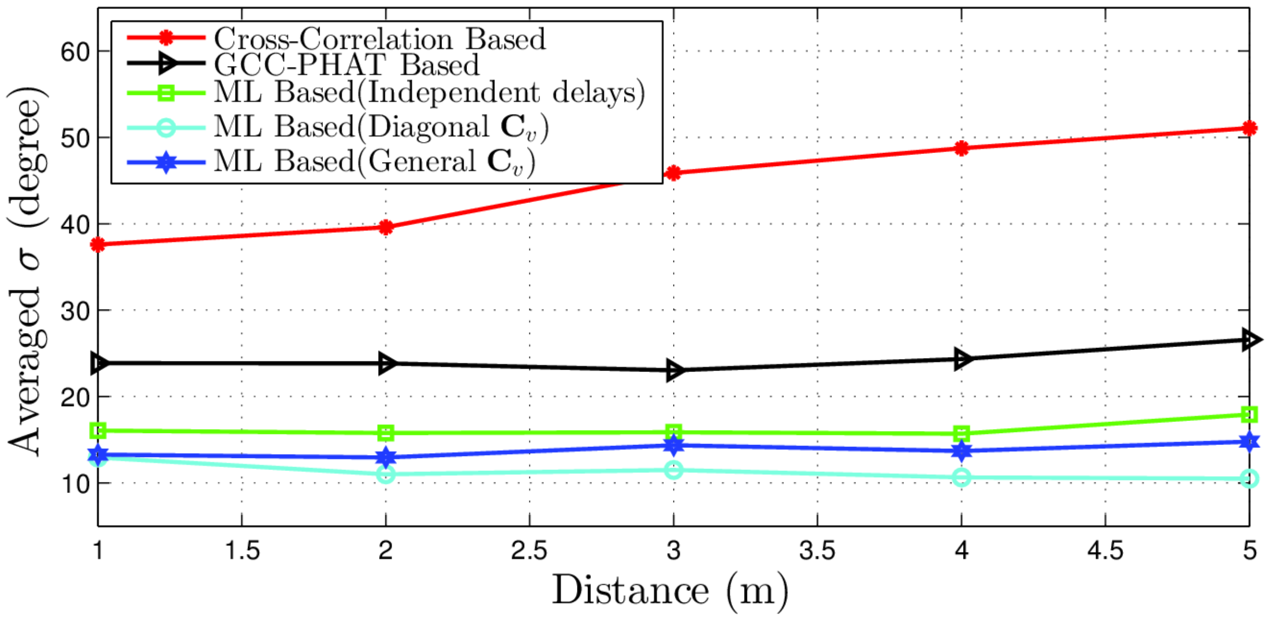
Performance as a function of θ at 0 dB SNR



Performance as a function of SNR



Performance as a function of Distance at 0 dB SNR



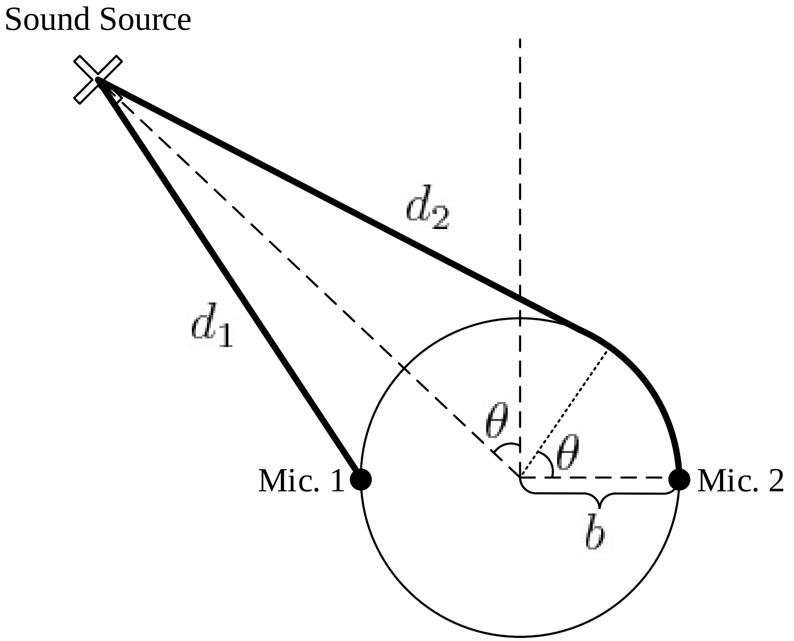
Content

Conclusion

- We proposed an “informed” TDoA-based DoA estimator via a maximum likelihood approach.
- We considered a free-field and far-field model to rely on minimal number of user-specific assumption.
- We showed that the likelihood function be calculated efficiently via Inverse Discrete Fourier Transform.

Future work

- Sphere Head Model
  - ITD:
    - $D_1 - D_2 \approx \left[\frac{b}{c}(\sin\theta + \theta)\right]$
  - ILD:
    - $20\log_{10}(\frac{\alpha_1}{\alpha_2}) \approx \gamma(k)\sin\theta$



Thank you!