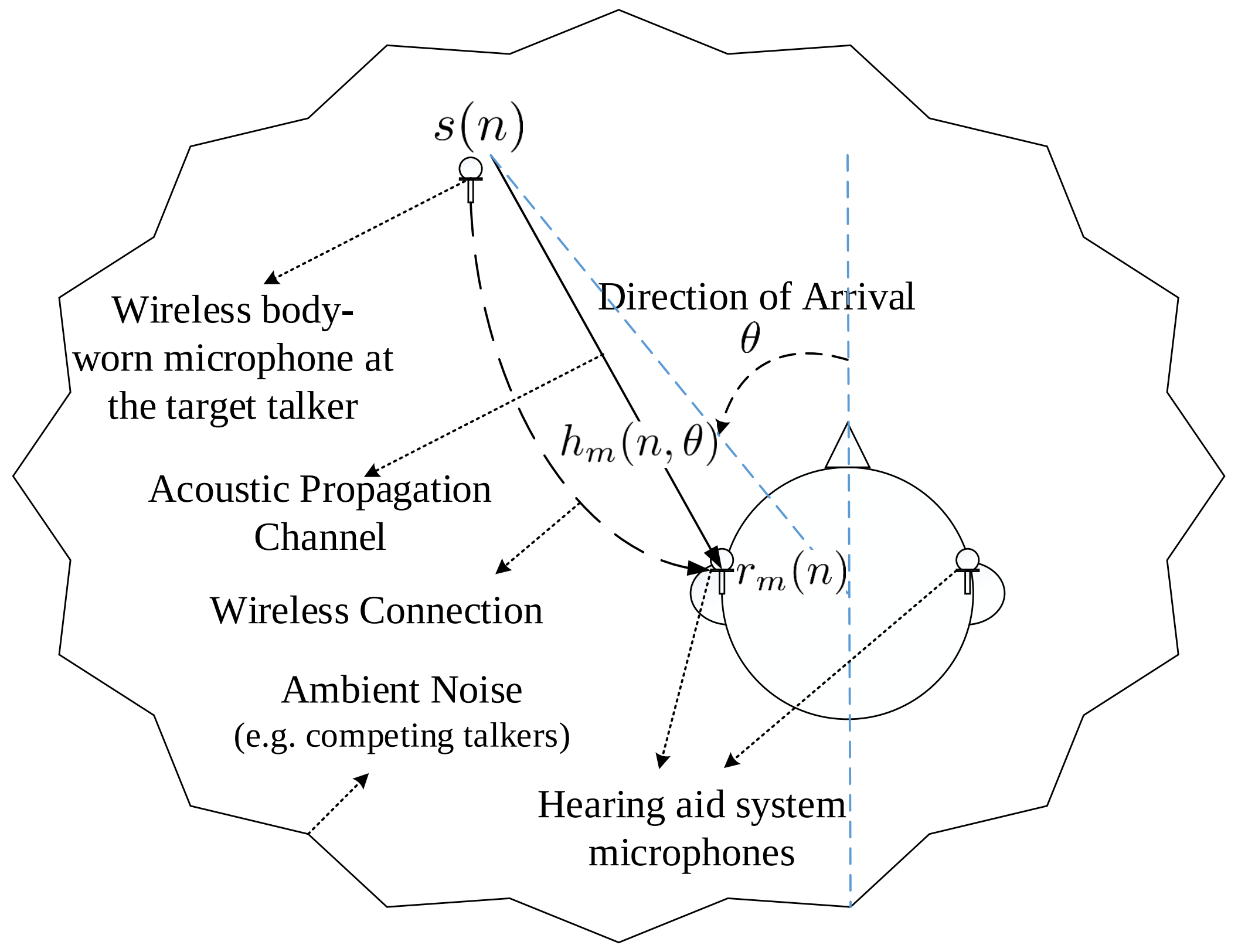
Informed Direction of Arrival Estimation using a Spherical-Head Model for Hearing Aid Applications

Mojtaba Farmani[1], Michael Syskind Pedersen[2], Zheng-Hua Tan[1], Jesper Jensen[1,2]

[1]*Aalborg University, Denmark* [2]*Oticon A/S, Denmark*

Introduction

- A direction of arrival (DOA) estimator for a binaural hearing aid system (HAS).
- The HAS can connect to a **wireless microphone.**
- The wireless microphone informs the HAS about the **noise-free** version of the target sound.
- A **spherical-head model** is used to consider the acoustic impacts of the head on the received signals at the HAS.
- The proposed DOA estimator is based on a **maximum likelihood (ML)** framework.
- To evaluate the likelihood function efficiently, **an inverse discrete Fourier transform (IDFT) technique** has been used.



- Why do we need DOA estimation?
  - **Binauralization** of the noise-free signal.

Signal Model

- **Time domain:**

$$r_m(n) = s(n) * h_m(n,\theta) + \nu_m(n)$$

- **Short-time Fourier transform domain:**

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

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

- **Simplification**

$$H_m(k,\theta) = \alpha_m(k,\theta)\mathrm{e}^{-\frac{j2\pi k}{N}D_m(k,\theta)}$$

$$\Rightarrow \tilde{H}_m(k,\theta) = \tilde{\alpha}_m(\theta)\mathrm{e}^{-\frac{j2\pi k}{N}\tilde{D}_m(\theta)}$$

- **Stack all the microphone signals:**

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

Maximum Likelihood Framework

- The additive noise signal is modeled as **a zero-mean circularly-symmetric complex Gaussian distribution**:

$$\boldsymbol{V}(l,k) \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_v(l,k)).$$

- $\mathbf{C}_v(l,k)$ can be estimated relatively easily because both the noisy signal and the noise-free signal are available.
- The likelihood function is defined by

$$p(\underline{\underline{\mathbf{R}}}(l)|\boldsymbol{S}(l), \underline{\underline{\tilde{\mathbf{H}}}}(\theta), \underline{\underline{\mathbf{C}_v}}(l)) =$$

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

  - where $\boldsymbol{Z}(l,k) = \boldsymbol{R}(l,k) - S(l,k)\tilde{\boldsymbol{H}}(k)$.

- The reduced log-likelihood function is given by

$$\tilde{\mathcal{L}} = \sum_{k=1}^{N}\{-(\boldsymbol{Z}(l,k))^{\mathrm{H}}\mathbf{C}_v^{-1}(l,k)(\boldsymbol{Z}(l,k))\}. \qquad (1)$$

Head Model

Different approaches to consider the acoustic impacts of the head:

- User-Specific (measured HRTF) [1].
- **Spherical-Head Model.**
- Free-Field [2].

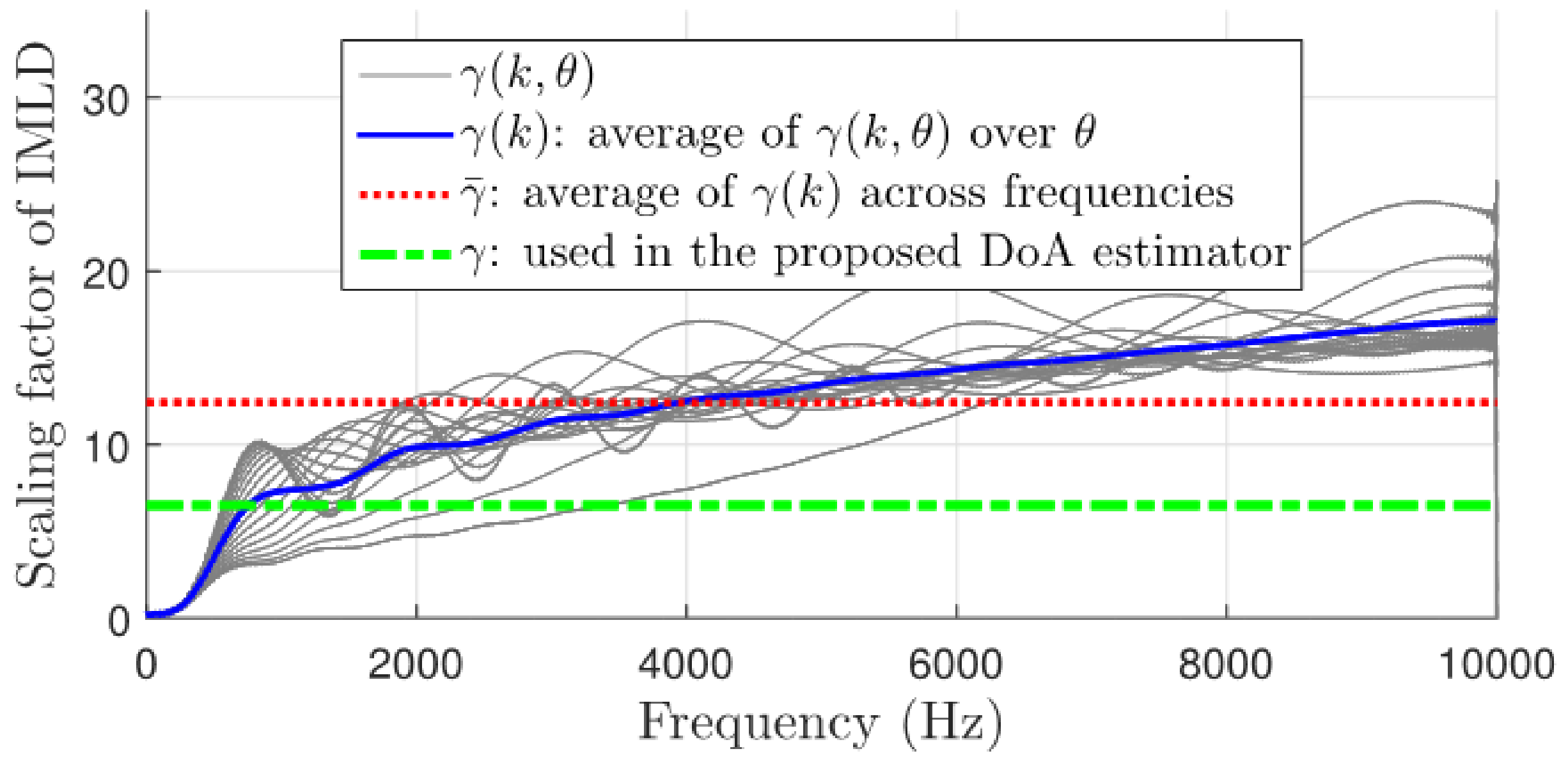
- Spherical-head model (inspired from [3]):
  - Inter-microphone time difference (IMTD) :

$$\begin{aligned}\Delta T(\theta) &= \tilde{D}_{\text{left}}(\theta) - \tilde{D}_{\text{right}}(\theta) \\ &= \frac{b}{c}\left(\sin(\theta) + \theta\right).\end{aligned}$$

    - b is the head radius, and c is the sound speed.
  - Inter-microphone level difference (IMLD) :

$$\begin{aligned}\Delta L(\theta) &= 20\log_{10}\left(\frac{\tilde{\alpha}_{\text{left}}(\theta)}{\tilde{\alpha}_{\text{right}}(\theta)}\right) \\ &= \gamma\sin(\theta)\end{aligned}$$

- Scaling factor γ using theoretical HRTFs [4]



DOA Estimator

- **Equation (1) should be expanded based on the head model.**
- **Let us denote**

$$\mathbf{C}_v^{-1}(l,k) \equiv \begin{bmatrix} C_{11}(l,k) & C_{12}(l,k) \\ C_{21}(l,k) & C_{22}(l,k) \end{bmatrix}.$$

- **The reduced log-likelihood function is given by**

$$\tilde{\mathcal{L}}(\theta, D_{\text{left}}) = \frac{f^2(\theta, D_{\text{left}})}{g(\theta)},$$

Where

$$\begin{aligned} f(\theta, D_{\text{left}}) = \sum_{k=1}^{N} \Bigg(& C_{11}(l,k)R_{\text{left}}(l,k) + \\ & C_{12}(l,k)R_{\text{right}}(l,k) + 10^{\frac{\gamma\sin(\theta)}{20}}\Big(C_{21}(l,k)R_{\text{left}}(l,k) + \\ & C_{22}(l,k)R_{\text{right}}(l,k)\Big)e^{j2\pi\frac{k}{N}[-\frac{b}{c}(\sin(\theta)+\theta)]}\Bigg) \times \\ & S^*(l,k)e^{j2\pi\frac{k}{N}D_{\text{left}}(\theta)}, \end{aligned}$$

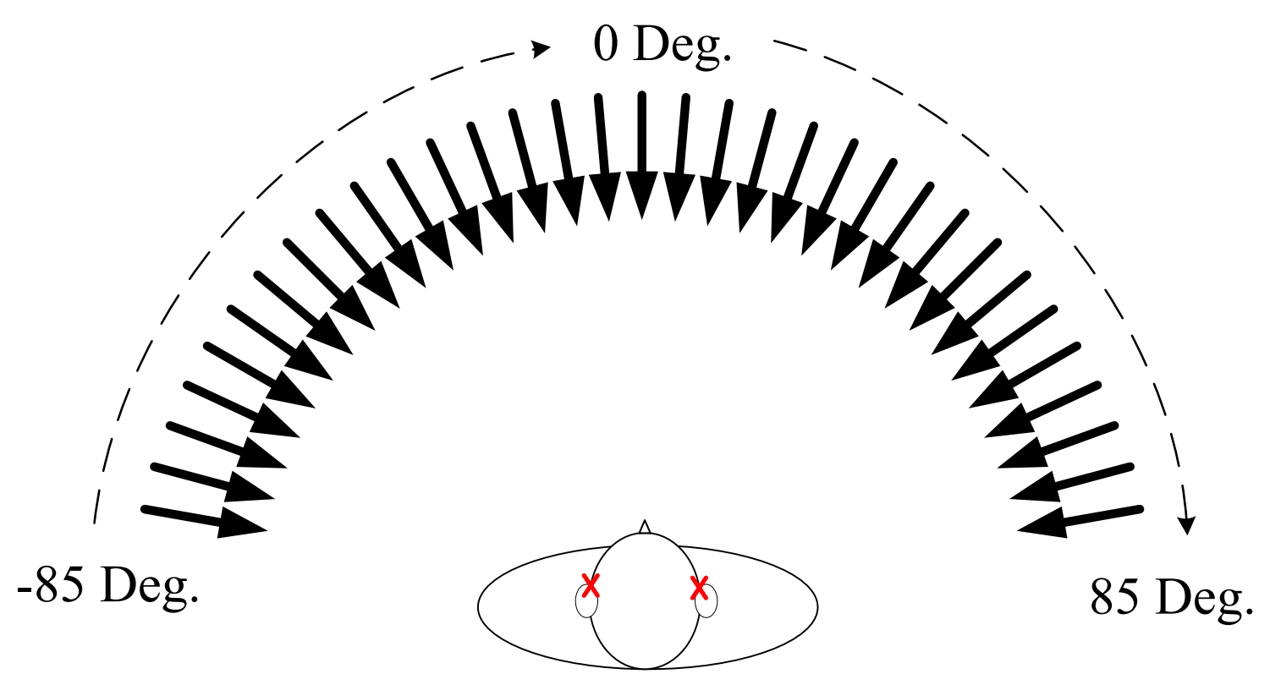
$$\begin{aligned} g(\theta) = \sum_{k=1}^{N} \Big(& C_{11}(l,k) + 2\times 10^{\frac{\gamma\sin(\theta)}{20}} C_{21} e^{j2\pi\frac{k}{N}[-\frac{b}{c}(sin(\theta)+\theta)]} + \\ & 10^{\frac{\gamma\sin(\theta)}{10}} C_{22}(l,k)\Big)|S(l,k)|^2, \end{aligned}$$

- $f(\theta, D_{\text{left}})$ **is an inverse Fourier transform with respect to** D_{left}.
- **Considering a discrete** Θ **set of different** θ**s:**

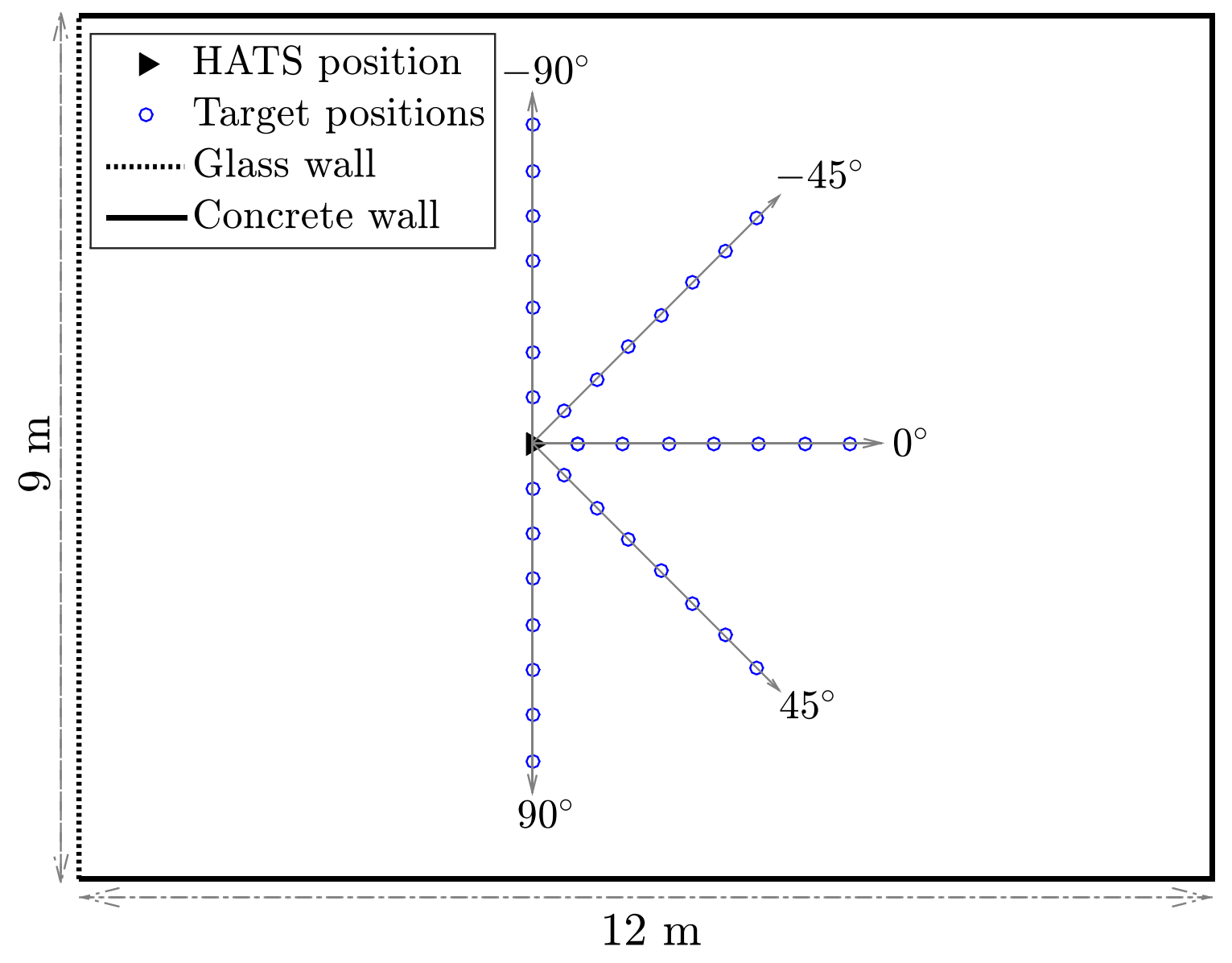
$$[\hat{\theta}, \hat{D}_{\text{left}}] = \arg\,\max_{\theta\in\Theta, D_{\text{left}}} \tilde{\mathcal{L}}(\theta, D_{\text{left}}).$$

HRIR Measurements for simulations

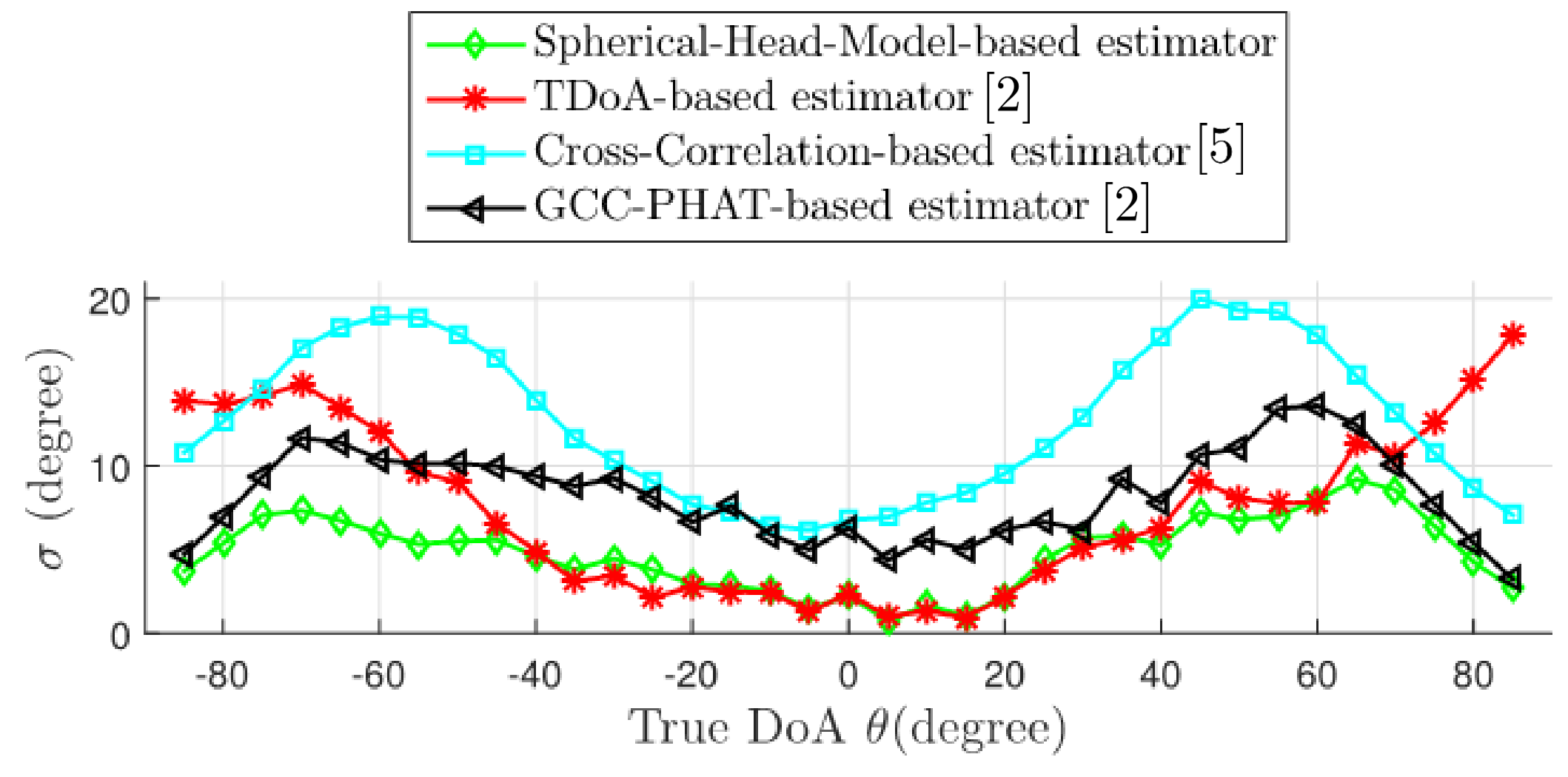
Anechoic measurements



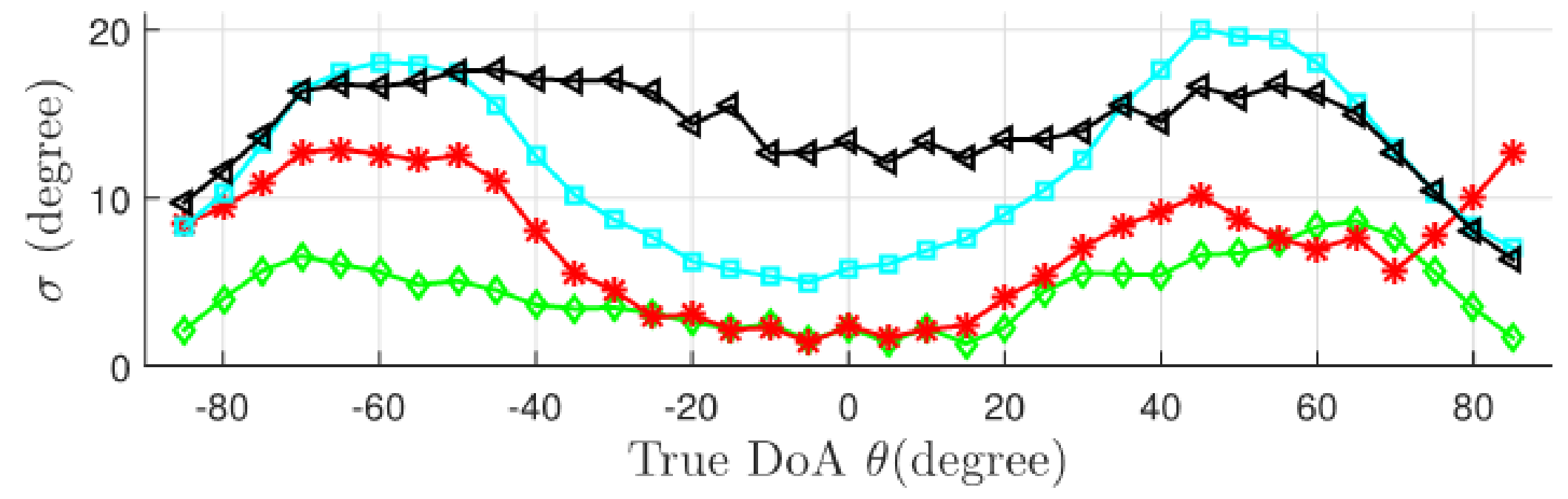
An actual reverberant room



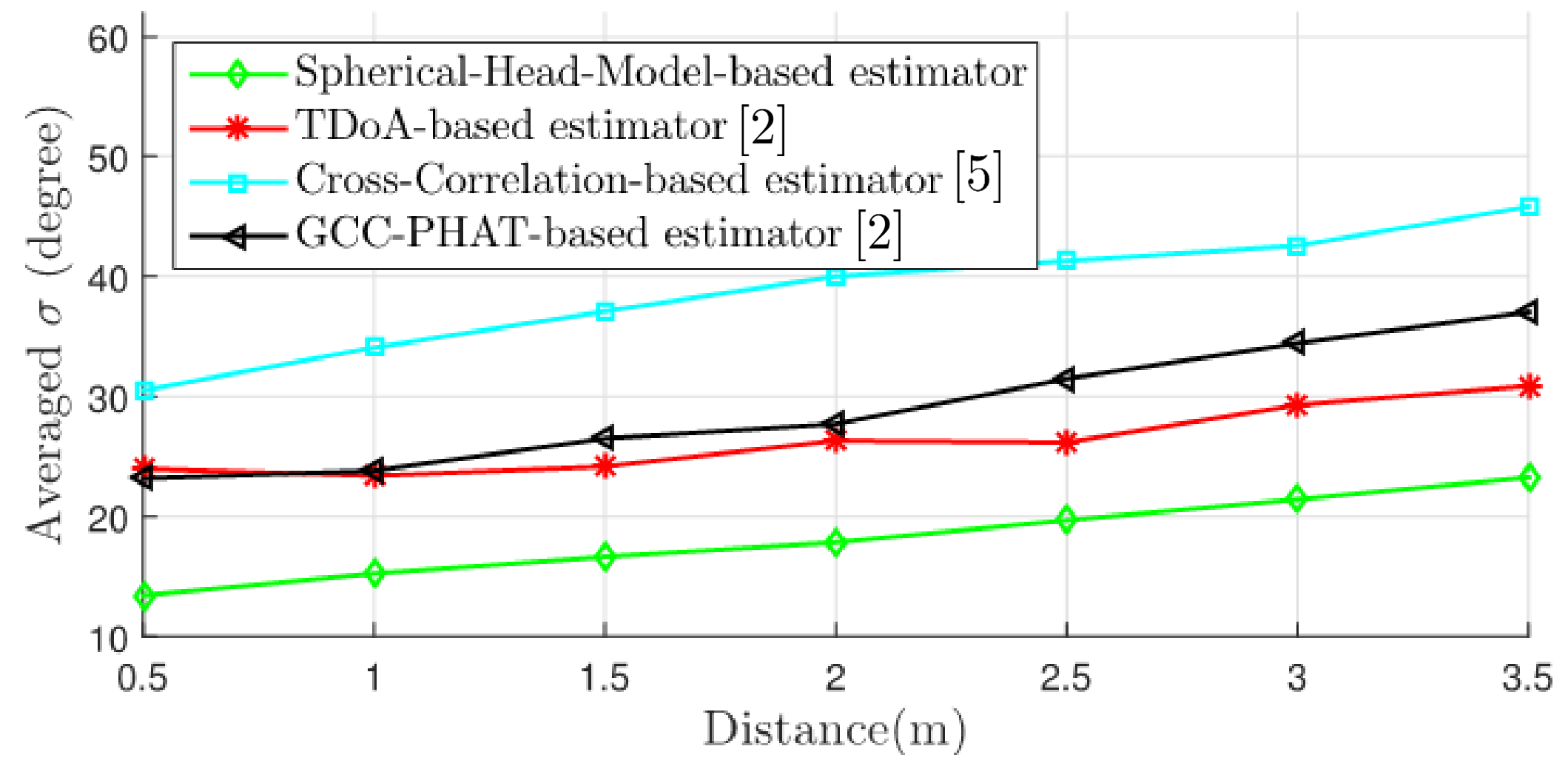
Simulation Results

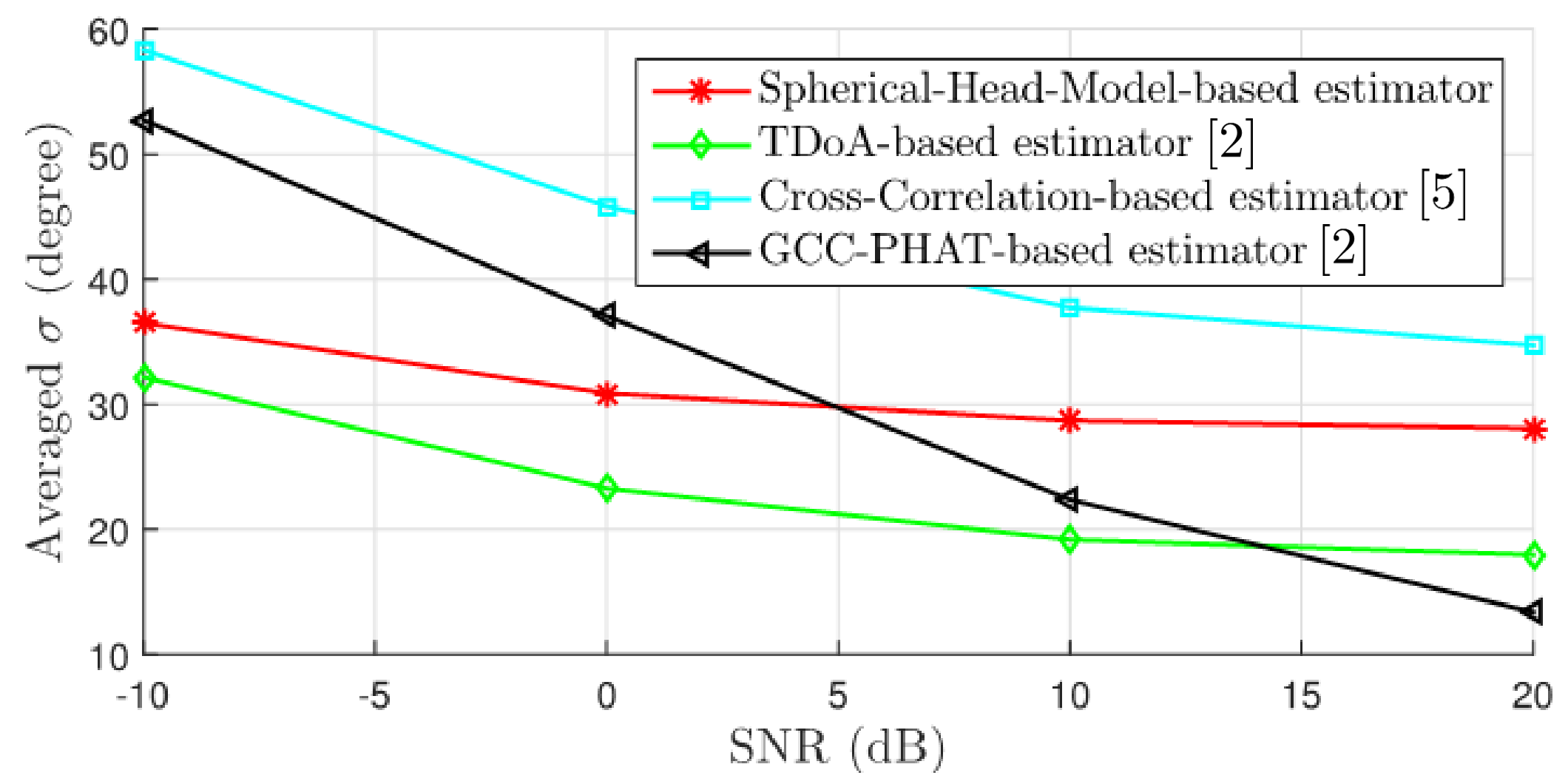


Babble-speech noise (SNR = 0 dB)



Bottling-factory noise (SNR=0 dB)





References

[1] M. Farmani et al, “Maximum likelihood approach to “informed” Sound Source Localization for Hearing Aid applications,” *in Proc. IEEE ICASSP 2015.*

[2] M. Farmani et al, “Informed TDoA-based direction of arrival estimation for hearing aid applications,” in Proc. of IEEE GlobalSIP, December 2015.

[3] M. Raspaud et al, “Binaural source localization by joint estimation of ILD and ITD,” IEEE Trans. Audio, Speech, and Lang. Process., vol. 18, no. 1, pp. 68–77, 2010.

[4] R. Duda and W.L. Martens, “Range dependence of the response of a spherical head model,” The Journal of the Acoustical Society of America, vol. 104, no. 5, pp. 3048–3058, 1998.

[5] G. Courtois et al., “Implementation of a binaural localization algorithm in hearing aids: specifications and achievable solutions,” in Audio Engineering Society Convention 136, April 2014, p. 9034.