Exploiting an External Microphone to Improve Time-Difference-of-Arrival Estimates for Euclidean Distance Matrix-Based Source Localization

Klaus Brümann, Simon Doclo
Dept. of Medical Physics and Acoustics and Cluster of Excellence Hearing4all,
University of Oldenburg, Germany

Hearing4all Exzellenzcluster im Auditory Valley

Carl von Ossietzky Universität Oldenburg

Introduction

- Many source localization methods rely on estimated time-differences-of-arrival (TDOAs) between microphone pairs
- Accurate TDOA estimation in noisy and reverberant environments not straightforward
- We assume an external microphone is available in the vicinity of the source (capturing strong direct source component, relative to noise and reverberation)

What is an External Microphone?

An **additional microphone** to a main array, which is usually **spatially separated** and **has an unknown position**.

MAIN IDEAS

Exploit favourable acoustic conditions of external microphone near the source to

- improve reliability of estimated TDOAs
- improve accuracy of Euclidean distance matrix (EDM)-based source localization [1]

without requiring knowledge of external microphone position.

TDOA Estimation

Generalized Cross Correlation with Phase Transform (GCC-PHAT) Method [2]

Noisy and reverberant m-th microphone signal

$$Y_m(\omega) = \underbrace{X_{d,m}(\omega)}_{\text{direct}} + \underbrace{X_{r,m}(\omega)}_{\text{reverberation}} + \underbrace{N_m(\omega)}_{\text{noise}}$$

Frequency-domain GCC-PHAT function between i-th and j-th microphones

$$\psi_{i,j}(\omega) = \frac{\mathbb{E}\{Y_i(\omega)Y_j^*(\omega)\}}{|\mathbb{E}\{Y_i(\omega)Y_j^*(\omega)\}|}$$

Time-domain GCC-PHAT function

$$\xi_{i,j}(\tau) = \int_{-\omega_0}^{\omega_0} \psi_{i,j}(\omega)e^{j\omega\tau}d\omega$$

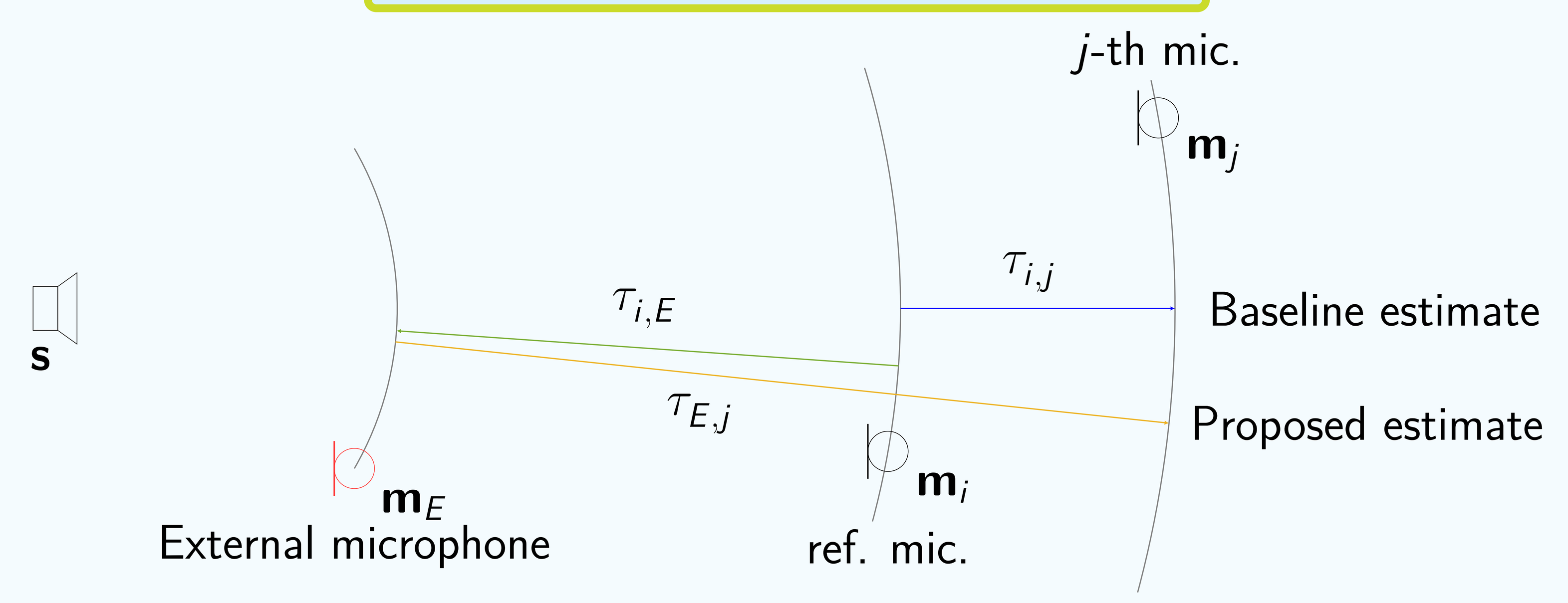
Conventional TDOA Estimation

$$\hat{\tau}_{i,j} = \underset{\tau}{\operatorname{argmax}} \; \xi_{i,j}(\tau)$$

GCC-PHAT Function Exploiting External Microphone

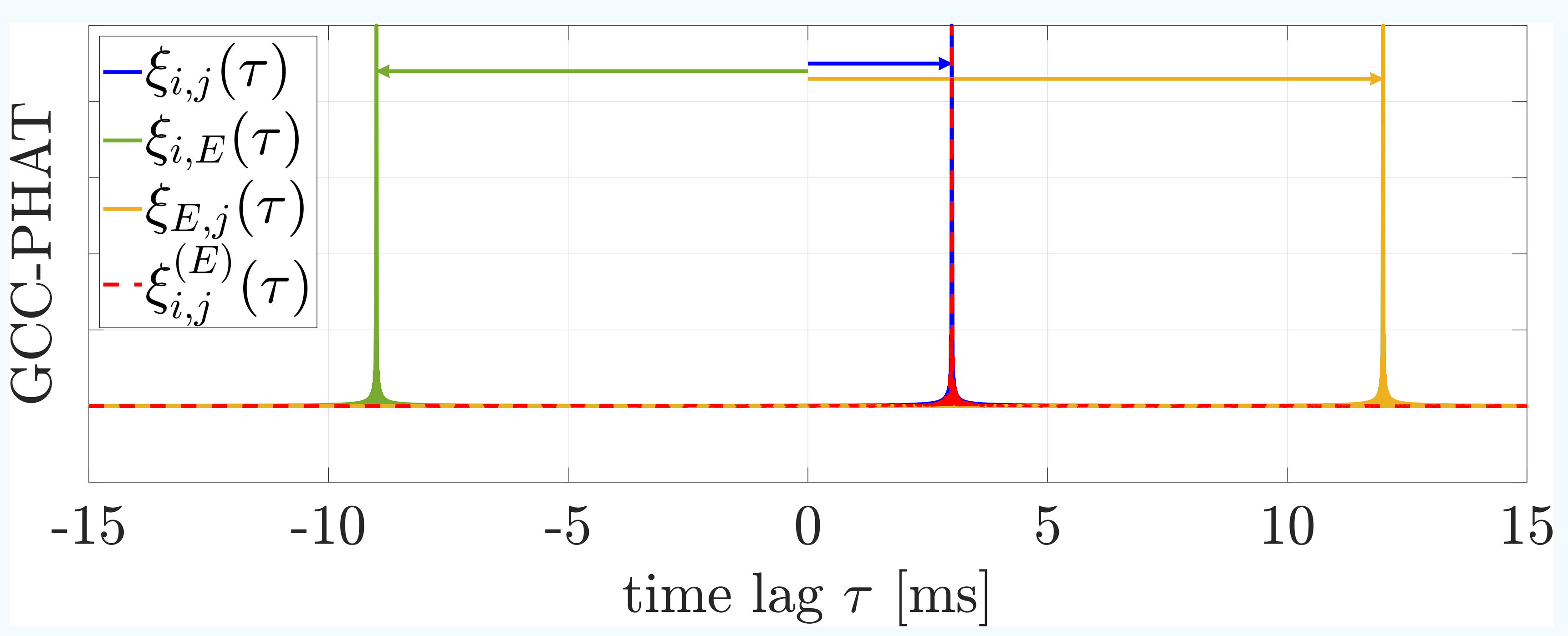
Exploit favourable acoustic conditions of external microphone by incorporating it in the GCC-PHAT function of microphones in main array

$$\xi_{i,j}^{(E)}(\tau) = \int_{-\infty}^{\infty} \xi_{i,E}(\tau')\xi_{E,j}(\tau-\tau')d\tau'$$



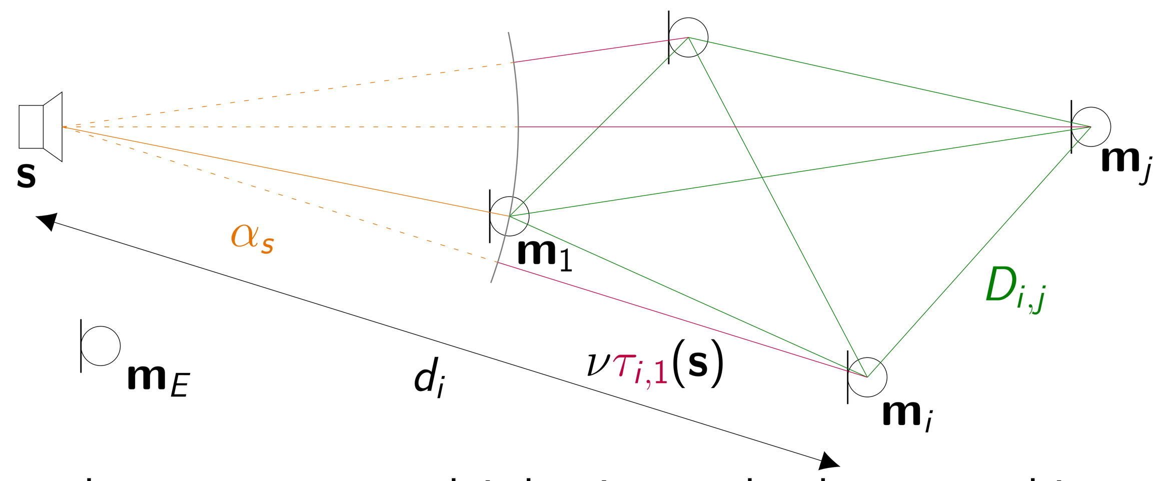
Anechoic case:

$$\xi_{i,j}^{(E)}(\tau) = \int_{-\infty}^{\infty} \delta\left(\tau' - \frac{\alpha_i - \alpha_E}{\nu}\right)\delta\left(\tau - \frac{\alpha_E - \alpha_j}{\nu} - \tau'\right) d\tau' = \delta(\tau - \tau_{i,j})$$



EDM-Based Source Position Estimation [1]

Given a distributed microphone array (DMA) with a known geometry



unknown distance between source and i-th mic. can be decomposed into two components

$$d_i = \underbrace{\alpha_s}_{\text{source to ref. mic.}} + \underbrace{\nu\,\tau_{i,1}(\mathbf{s})}_{\text{ref. mic. to } i\text{-th mic}}$$

Theory: Assuming that TDOAs $\tau_{i,1}(\mathbf{s})$ are available and defining

$$d_i(\alpha) = \alpha + \nu\,\tau_{i,1}(\mathbf{s})$$

with variable α, an EDM-based cost function $J(\alpha)$ can be used to compute α_s

$$\alpha_s = \underset{\alpha}{\operatorname{argmin}} \; J(\alpha) = \underset{\alpha}{\operatorname{argmin}} \sum_{i=3+1}^{M+1} |\lambda_i(\alpha)|$$

Practice: consider multiple TDOA estimates per microphone pair

$$\hat{\alpha}_s = \underset{\alpha, c_2, \ldots, c_M}{\operatorname{argmin}} \; J(\alpha, \hat{\tau}_{2,1}^{c_2}, \ldots, \hat{\tau}_{M,1}^{c_M})$$

Experimental Evaluation

Comparison of 3D source position estimation error $\varepsilon_s = ||\mathbf{s} - \hat{\mathbf{s}}||_2$ of EDM-based source position estimation method using TDOA estimates from:

- GCC-PHAT function $\xi_{i,j}(\tau)$ without external microphone
- External microphone-based GCC-PHAT function $\xi_{i,j}^{(E)}(\tau)$

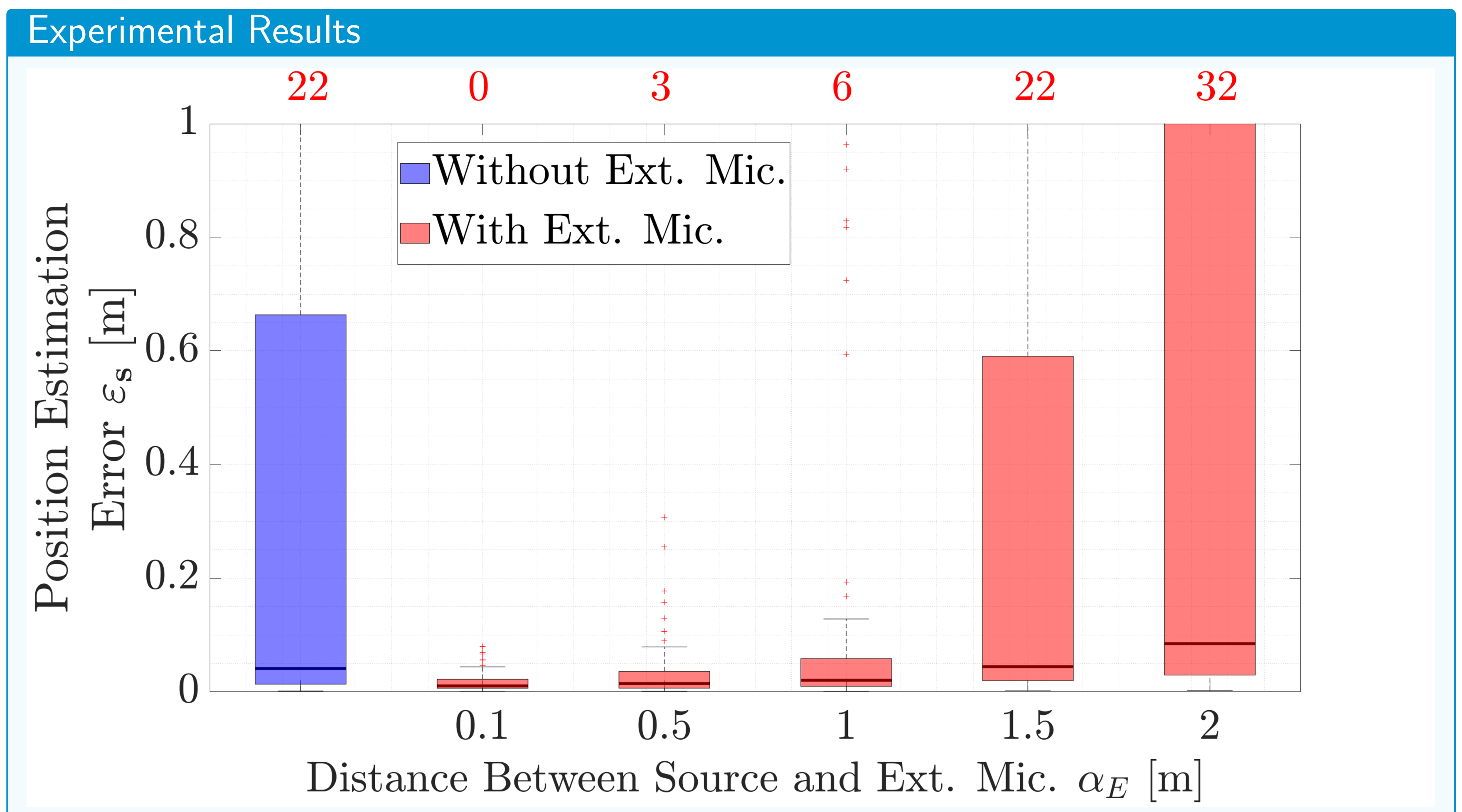
Framework and Acoustical Parameters

- $6 \times 6 \times 2.4$ m room simulated with RIR generator [3]
- $M = 6$ microphones randomly positioned within cube with cube length 2 m
- 100 acoustic scenarios (random 5 s speech signal, microphone array position & geometry, and speech source position (3 m away from centroid of DMA))
- Babble noise at average SNR = 0 dB across the microphones of the DMA
- Average direct-to-reverberant ratio (DRR) = 0 dB across the microphones of the DMA
- f_s = 16 kHz, 512 sample (32 ms) frame length, 50% overlap between frames, 1024 sample FFT-length

SNR and DRR for different external microphone positions between source and centroid of DMA

Distance between source and external microphone [m]	0.1	0.5	1	1.5	2
SNR in external microphone [dB]	23	9	5	2	1
DRR in external microphone [dB]	24	11	6	3	2

Experimental Results



3D position estimation error using EDM-based source position estimation method, where the TDOAs were estimated **without** and **with** an external microphone. The red numbers at the top indicate the number of results outside of the plotted range.

Conclusions and Outlook

- ✓ Incorporating **external microphone** for TDOA estimation **considerably improves accuracy of source position estimate**, particularly when the external microphone is located close to source.
- Further analysis of influencing parameters needed (SNR, DRR, external microphone position)

References

[1] K. Brümann and S. Doclo, "3D single source localization based on Euclidean distance matrices," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*. Bamberg, Germany: IEEE, 2022, pp. 1–5.

[2] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. on Audio, Speech, Language Processing*, vol. 24, no. 4, pp. 320–327, 1976.

[3] E. A. P. Habets, *RIR-Generator*, available at "https://github.com/ehabets/RIR-Generator".

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2177/1 - Project ID 390895286 and Project ID 352015383 - SFB 1330 B2.