Incremental Averaging Method to Improve Graph-Based Time-Difference-of-Arrival Estimation

Klaus Brümann[1], Kouei Yamaoka[2], Nobutaka Ono[3], Simon Doclo[1]

[1]Carl von Ossietzky Universität, Oldenburg, Germany [2]University of Tokyo, Tokyo, Japan
[3]Tokyo Metropolitan University, Tokyo, Japan

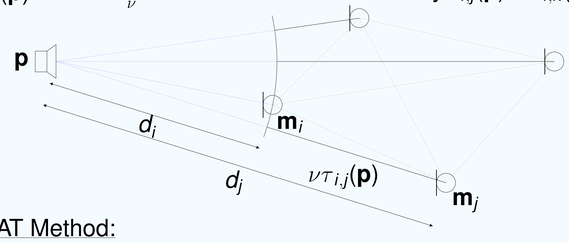
Introduction

- Several methods exist for source localization (TDOA, SRP-PHAT, DNN)
- **Focus of this paper: TDOA-based source localization**
- Noise & Reverberation negatively affect TDOA estimation accuracy
- State-of-the-art TDOA estimation methods use **just one (potentially unreliable) cross-power spectral density (CPSD)** per microphone pair

State-of-the-Art TDOA Estimation

Generalized Cross-Correlation with Phase-Transforms (GCC-PHAT) [1]

TDOA $\tau_{i,j}(\mathbf{p}) = \frac{\|\mathbf{p}-\mathbf{m}_j\|_2 - \|\mathbf{p}-\mathbf{m}_i\|_2}{\nu}$ Consistency $\tau_{i,j}(\mathbf{p}) = \tau_{i,k}(\mathbf{p}) - \tau_{j,k}(\mathbf{p})$



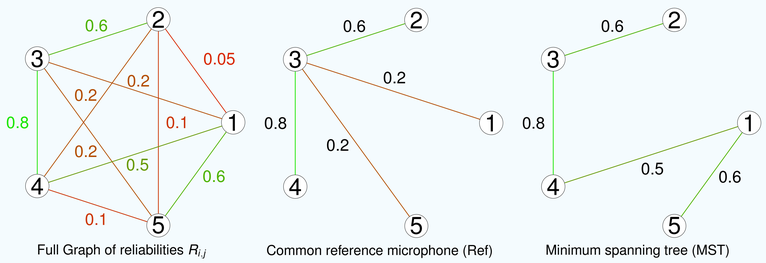
GCC-PHAT Method:
Microphone signal $Y_i(\omega)$
CPSD $\psi_{i,j}(\omega) = \mathbb{E}\{Y_i(\omega)Y_j^*(\omega)\}$
GCC-PHAT function $\xi_{i,j}(\tau) = \int_{-\infty}^{\infty} \frac{\psi_{i,j}(\omega)}{|\psi_{i,j}(\omega)|} \exp(\jmath\omega\tau) d\omega$
TDOA estimate $\tilde{\tau}_{i,j} = \underset{\tau}{\operatorname{argmax}}\ \xi_{i,j}(\tau)$

Graph-Based Minimal Set

For M-microphone array: TDOA-based source localization methods generally use **minimal set** of $M-1$ microphone pairs.

Optimal minimal set determined using **minimum spanning tree (MST)** [2] of graph of GCC-PHAT reliabilities:

- Vertices $\longleftrightarrow$ microphones
- Edges $\longleftrightarrow$ cost (negative GCC-PHAT reliability: $R_{i,j} = \xi_{i,j}(\tilde{\tau}_{i,j})$)



MAIN IDEA

We propose to incrementally re-estimate TDOAs based on average over **multiple CPSDs from multiple microphone pairs** by leveraging TDOA consistency relations

Proposed Method

Indirect CPSD Computation

Direct source CPSD $\psi_{i,j}(\omega) = \frac{1}{(4\pi)^2 d_i d_j} \exp(-\jmath\omega\tau_{i,j}(\mathbf{p}))$

Indirectly compute CPSD indirectly via k-th microphone as

$$\tilde{\psi}_{i,j}(\omega, \tau_{j,k}(\mathbf{p})) = \psi_{i,k}(\omega) \exp(\jmath\omega\tau_{j,k}(\mathbf{p}))$$

(requires phase alignment)

Incremental Method for Averaging CPSDs

We propose to incrementally re-estimate the TDOAs corresponding to **edges of the MST**. In each step H, to improve each re-estimated TDOA we compute averaged CPSD based on:

- directly computed CPSD $\psi_{i,j}(\omega)$
- $H-1$ indirectly estimated CPSDs $\tilde{\psi}_{i,j}(\omega, \tilde{\tau}'_{j,h}) = \psi_{i,j}(\omega) \exp(\jmath\omega\tilde{\tau}'_{j,h})$
  (via microphones used in previous steps h)

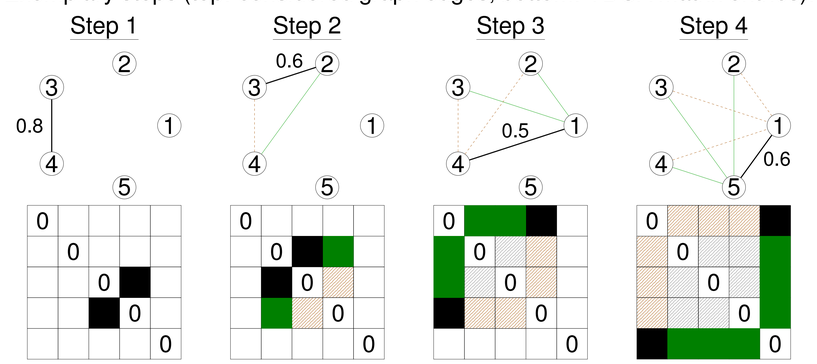
Averaged CPSD $$\psi'_{i,j}(\omega) = \frac{1}{H}\left(\psi_{i,j}(\omega) + \sum_{h\in\gamma} \tilde{\psi}_{i,j}(\omega, \tilde{\tau}'_{j,h})\right)$$

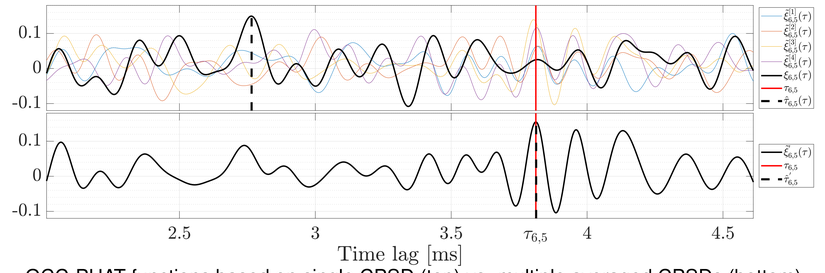
γ: set of microphones used in previous steps

Remaining TDOAs re-estimated as $\tilde{\tau}'_{h,j} = \tilde{\tau}'_{i,j} - \tilde{\tau}'_{i,h}$

- **Step 1**: no indirectly estimated CPSDs available!
  $\Rightarrow$ begin with microphone pair corresponding to MST edge with highest $R_{i,j}$
- **Subsequent steps**: consider **neighbouring** MST edge with next-highest $R_{i,j}$

Exemplary steps (top: considered graph edges, bottom: TDOA matrix entries):





GCC-PHAT functions based on single CPSD (top) vs. multiple averaged CPSDs (bottom)

Experimental Evaluation

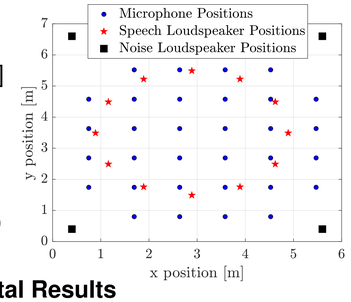
We compared following methods:

- SRP-PHAT: Steered-response power with phase transforms
- Ref-A: Randomly chosen reference microphone
- Ref-R: Reference microphone determined based on GCC-PHAT reliability
- MST: MST-based minimal set
- **MST+**: Proposed method

in terms of mean TDOA and source position estimation performance.

Framework and Acoustical Parameters

- Each result based on 100 random scenarios from BRUDEX database [3]
- 10 s signals, $M = 6$ microphones
- Babble noise @ SNR = 5 dB.
  3 reverberation levels:
  low ($T_{60} \approx 310$ ms), medium ($T_{60} \approx 510$ ms), high ($T_{60} \approx 1300$ ms)



Experimental Results

Reverb.	Mean Error Metric	TDOA Estimation Method SRP-PHAT	Ref-A	Ref-R	MST	MST+
Low	TDOA error [ms]	0.18	0.49	0.04	0.04	**0.01**
	Position error [cm]	20.0	28.1	2.9	2.9	**0.7**
	Position accuracy$_{\leq 10\text{ cm}}$ [%]	69.5	81.1	98.2	98.3	**99.8**
Med.	TDOA error [ms]	0.28	1.34	0.14	0.12	**0.08**
	Position error [cm]	29.3	70.9	9.0	8.9	**5.0**
	Position accuracy$_{\leq 10\text{ cm}}$ [%]	63.0	54.3	92.9	92.8	**95.8**
High	TDOA error [ms]	0.44	1.77	0.44	0.40	**0.28**
	Position error [cm]	37.9	90.1	24.3	20.3	**10.5**
	Position accuracy$_{\leq 10\text{ cm}}$ [%]	59.6	44.2	82.2	85.0	**92.4**

✓ Incrementally averaging **multiple CPSDs** improves TDOA and source position estimation accuracy vs. state-of-the-art **single CPSD** methods

References

[1] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, 1976.

[2] K. Yamaoka, T. Nakashima, Y. Wakabayashi, and N. Ono, "Minimum-spanning-tree-based time delay estimation robust to outliers," *IEEE Access*, vol. 11, pp. 121 284–121 294, 2023.

[3] D. Fejgin, W. Middelberg, and S. Doclo, "BRUDEX database: Binaural room impulse responses with uniformly distributed external microphones," in *Proc. ITG Conference on Speech Communication*, Aachen, Germany, 2023, pp. 126–130.

This work was funded by the German Academic Exchange Service (DAAD), the Federal Ministry of Education and Research (BMBF) - Project ID 57711070, and the German Research Foundation (DFG) under Germany's Excellence Strategy - EXC 2177/1 - Project ID 390895286.