DAGA 2018 MÜNCHEN

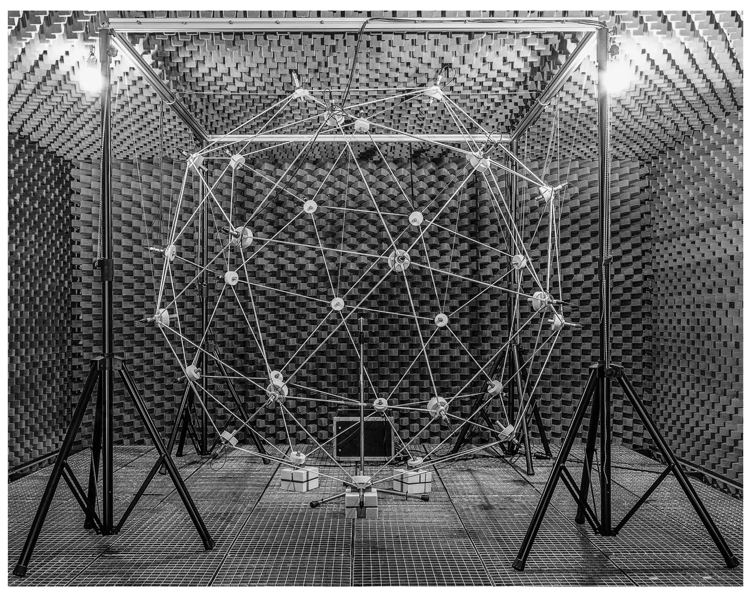
Tim Lübeck[1], Johannes M. Arend[1,2], Christoph Pörschmann[1]

Technology Arts Sciences TH Köln

A Real-Time Application for Sound Source Localization Inside a Spherical Microphone Array

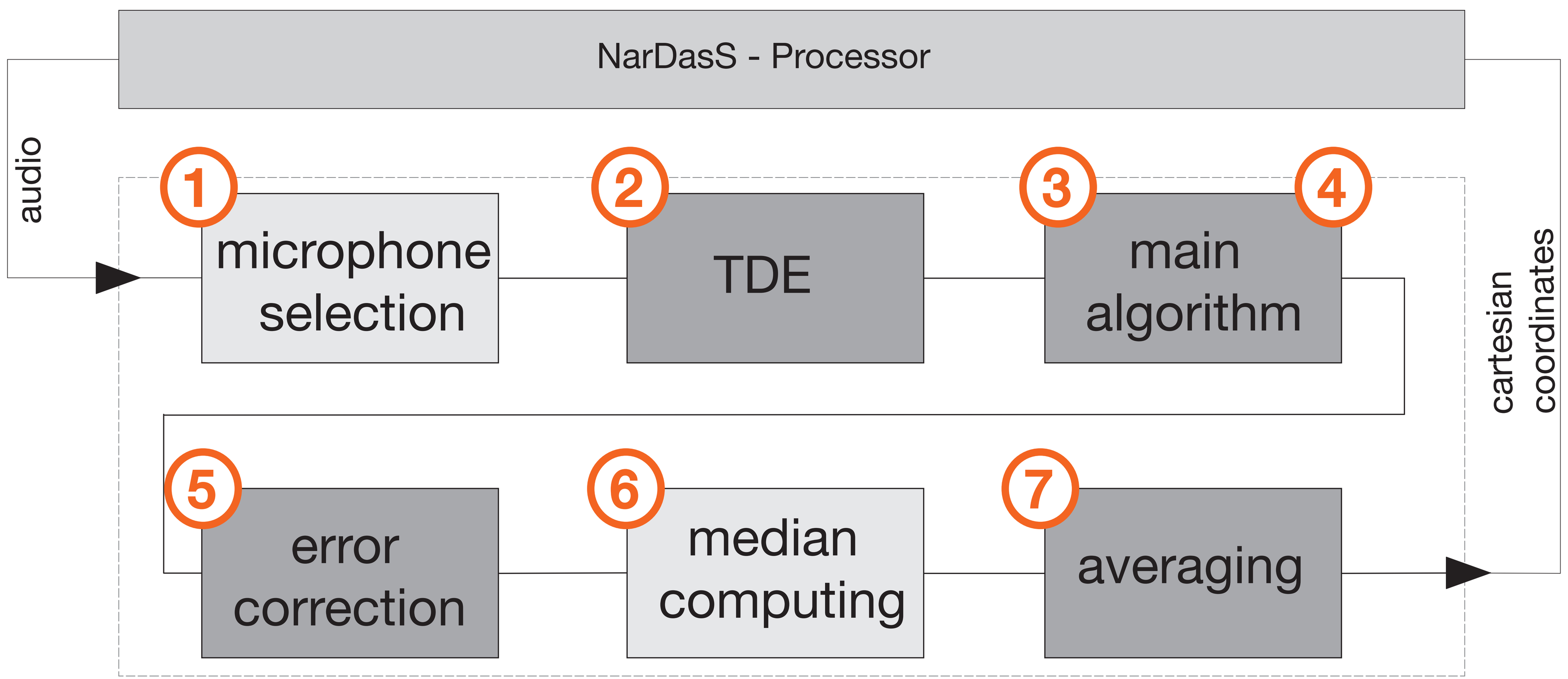
INTRODUCTION

- The **NarDasS**-project aims at a binaural reproduction of self-generated sound [1].
- The user is surrounded by a 32-channel microphone array with a diameter of 2 m.
- Off-center positions of the user require exact localization for proper microphone level adjustments.



AIM OF THIS WORK:
- Real-time sound source localization using Time Differences of Arrival (TDOA)

SYSTEM OVERVIEW



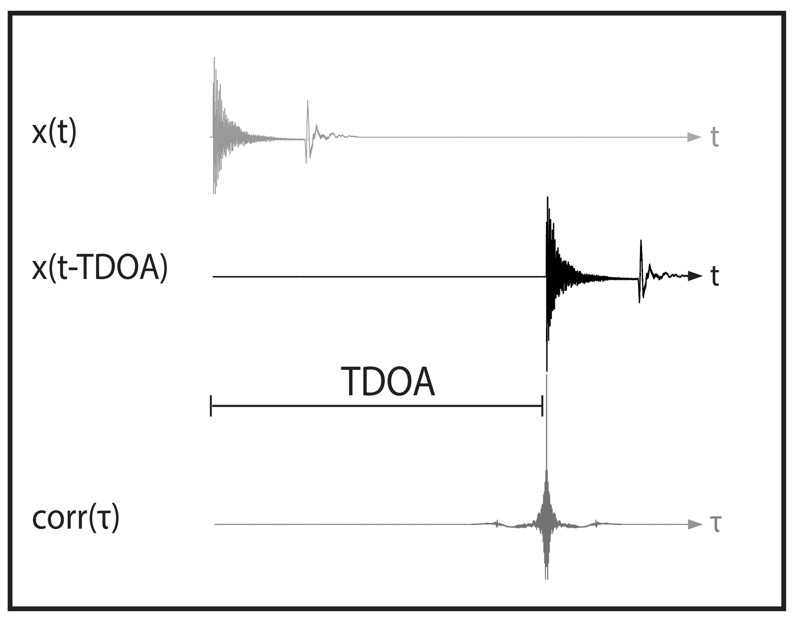
IMPLEMENTATION

1 Microphone Selection

- Determine the reference microphones.

2 Time Delay Estimation (TDE)

- Passive localization methods provide no information about the time of flight (TOF).
- Determine the TDOAs using a cross-correlation with **GCC-PHAT** technique:
  weighting of cross power spectrum density to retain the phase information only.
- Maximum of cross-correlation corresponds to the time delay.
- Cross-correlations of all microphones lead to 31 TDOAs.



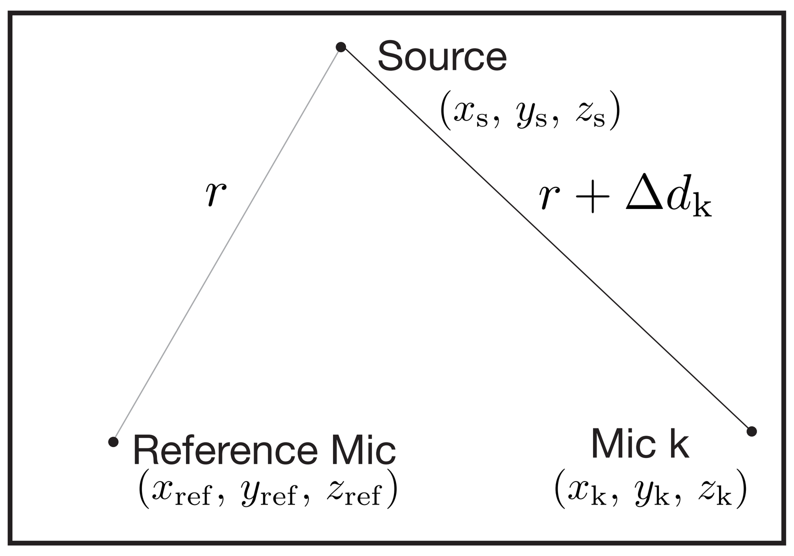
3 Setup of the Equation System

- Use Pythagoras theorem to describe distances between source and microphones.

1) $r^2 = (x_{\mathrm{ref}} - x_{\mathrm{s}})^2 + (y_{\mathrm{ref}} - y_{\mathrm{s}})^2 + (z_{\mathrm{ref}} - z_{\mathrm{s}})^2$

2) $(r + \Delta d_{\mathrm{k}})^2 = (x_{\mathrm{k}} - x_{\mathrm{s}})^2 + (y_{\mathrm{k}} - y_{\mathrm{s}})^2 + (z_{\mathrm{k}} - z_{\mathrm{s}})^2$

3) $x_{\mathrm{k}}^2 - 2x_{\mathrm{k}}x_{\mathrm{s}} + y_{\mathrm{k}}^2 - 2y_{\mathrm{k}}y_{\mathrm{s}} + z_{\mathrm{k}}^2 - 2z_{\mathrm{k}}z_{\mathrm{s}} - 2r\Delta d_{\mathrm{k}} = x_{\mathrm{ref}}^2 + y_{\mathrm{ref}}^2 + z_{\mathrm{ref}}^2 + \Delta d_{\mathrm{k}}^2$



- Substituting leads to a linear equation system with four unknown, non-squared variables: **A x = b**

$$\begin{bmatrix} 2(x_{\mathrm{ref}} - x_2) & 2(y_{ref} - y_2) & 2(z_{\mathrm{ref}} - z_2) & -2\Delta d_1 \\ 2(x_{\mathrm{ref}} - x_3) & 2(y_{ref} - y_3) & 2(z_{\mathrm{ref}} - z_3) & -2\Delta d_2 \\ 2(x_{\mathrm{ref}} - x_4) & 2(y_{ref} - y_4) & 2(z_{\mathrm{ref}} - z_4) & -2\Delta d_3 \\ 2(x_{\mathrm{ref}} - x_5) & 2(y_{ref} - y_5) & 2(z_{\mathrm{ref}} - z_5) & -2\Delta d_4 \end{bmatrix} \begin{bmatrix} x_{\mathrm{s}} \\ y_{\mathrm{s}} \\ z_{\mathrm{s}} \\ r \end{bmatrix} = \begin{bmatrix} x_{\mathrm{ref}}^2 + y_{\mathrm{ref}}^2 + z_{\mathrm{ref}}^2 + \Delta d_1^2 - x_2^2 - y_2^2 - z_2^2 \\ x_{\mathrm{ref}}^2 + y_{\mathrm{ref}}^2 + z_{\mathrm{ref}}^2 + \Delta d_2^2 - x_3^2 - y_3^2 - z_3^2 \\ x_{\mathrm{ref}}^2 + y_{\mathrm{ref}}^2 + z_{\mathrm{ref}}^2 + \Delta d_3^2 - x_4^2 - y_4^2 - z_4^2 \\ x_{\mathrm{ref}}^2 + y_{\mathrm{ref}}^2 + z_{\mathrm{ref}}^2 + \Delta d_4^2 - x_5^2 - y_5^2 - z_5^2 \end{bmatrix}$$

4 Solution of the Equation System

- Considering all 32 microphones leads to an overdetermined equation system.
- Solving by **least square fitting**
- More stable results by use of **QR decomposition**

5 Error Correction

- In case of unsuitable position estimations use the last position.
- The algorithm checks:
  - if position differs too much from preceding estimation (improbable velocity)
  - if estimated position is located outside the array

6 Median Computing

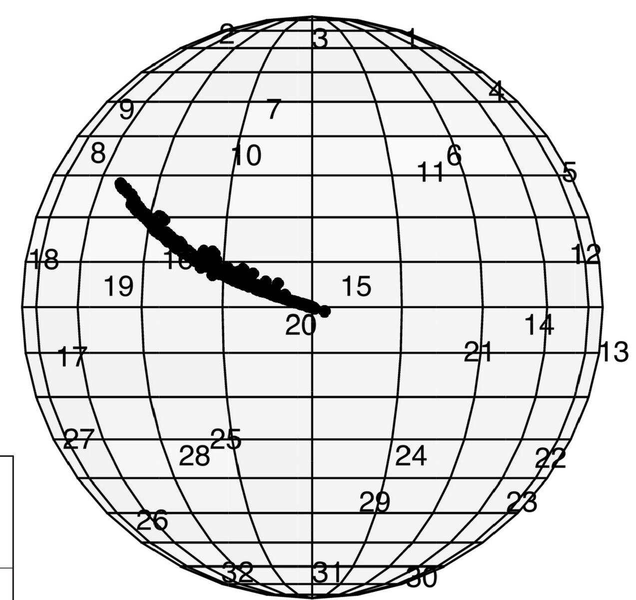
- Estimate one source position based on each reference microphone and calculate the median of all estimations (currently implemented: 9 references).

7 Averaging

- Weaken outliers by averaging over preceding estimations.

TECHNICAL EVALUATION

- Four test positions with four test signals
- Comparing results of different averaging methods.
- Evaluation of the algorithm for moving sources



	avg error [cm] a_{none}	avg error [cm] a_2	avg error [cm] a_4
Flamenco	3.9523	3.8947	3.8713
Speech	10.4693	9.927	9.7203
Noise	4.8062	4.6885	4.6558
Drums	0.1078	0.1060	0.1045
MAE	4.8339	4.6541	4.5879

RESULTS:
- Accuracy of approximately 5 cm for non-moving sources.
- Moving sources can be tracked.
- Up to now multiple sources cannot be localized.

REFERENCES

[1] Arend, J. M., Stade, P., and Pörschmann, C., „Binaural reproduction of self-generated sound in virtual acoustic environments," Proceedings of Meetings on Acoustics, 30(1), pp. 1-14, 2017.
[2] Brandstein, M. S., A Framework for Speech Source Localization Using Sensor Arrays, Ph.D. thesis, Massachusetts Institute of Technology, 1990.
[3] Knapp, C. H. and Carter, G. C., "The Generalized Correlation Method for Estimation of Time Delay," IEEE Transactions on Acoustics, Speech, and Signal Processing, 24(4), pp. 320–327, 1976.
[4] Mahajan,A. and Walworth,M.,"3-D position sensing using the differences in the time-of-flights from a wave source to various receivers," IEEE Transactions on Robotics and Automation, 17(1), pp. 91–94, 2001.
[5] Swartling, M., Direction of Arrival Estimation and Localization of Multiple Speech Sources in Enclosed Environments Mikael Swartling, 2012.

1 TH Köln, Institute of Communications Engineering, Cologne
2 TU Berlin, Audio Communication Group, Berlin
contact: tim.luebeck@th-koeln.de