PASSIVE ONLINE GEOMETRY CALIBRATION OF ACOUSTIC SENSOR NETWORKS

Axel Plinge[tu], Gernot A. Fink[tu], and Sharon Gannot[BIU] SPL March 2017

[tu] TU Dortmund University, Department of Computer Science, Dortmund, Germany [BIU] Bar Ilan University, Faculty of Engineering, Ramant Gan, Israel

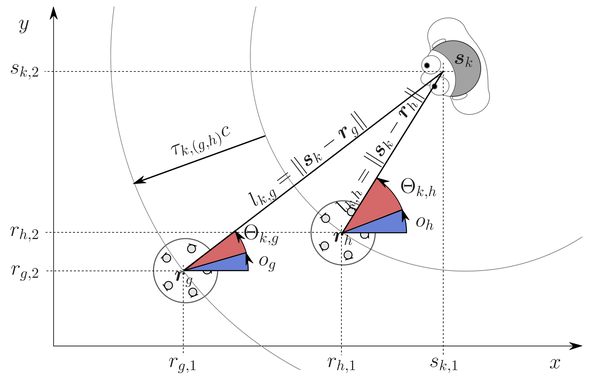
This work was supported by a fellowship within the FITweltweit program of the German Academic Exchange Service (DAAD).

Abstract

As we are surrounded by an increased number of mobile devices equipped with wireless links and multiple microphones, using them collaboratively for acoustic processing is a promising platform. These devices may constitute an acoustic sensor network (ASN). Many spatial processing algorithms require a precise knowledge of the nodes' locations and orientations. An automated real-time method is proposed to acquire this.

Problem Statement

Goal Calibrate relative geometry of several multi-microphone nodes from ambient speech by distributed online computation.



- G nodes at position $\boldsymbol{r}_g \in \mathbb{R}^d$ with orientation o_g
- Relative geometry [1], fix $\boldsymbol{r}_1 := [0,0]^T, o_1 := 0,$
- find $\boldsymbol{\gamma} = [r_{21}, r_{22} \dots r_{G2}, o_2 \dots o_G]^T$ using K speech events with directions of arrival (DoAs) $\Theta_{k,g}$ and time differences of arrival (TDoAs) $\tau_{k,(g,h)}$

1: Target Function

DoA-TDoA target function [2] for node pair measurement of speech event $\boldsymbol{m}_{k,(g,h)} = \{\Theta_{k,g}, \Theta_{k,h}, \tau_{k,(g,h)}\}$

- Triangulate source position $\hat{\boldsymbol{s}}_k$ from DoAs and geometry.
- The TDoA has to match $\rightarrow$ error function $\epsilon_{g,h}(\boldsymbol{m}_k, \boldsymbol{\gamma})$
- Penalize non-intersection with constant $\epsilon_{\max}$

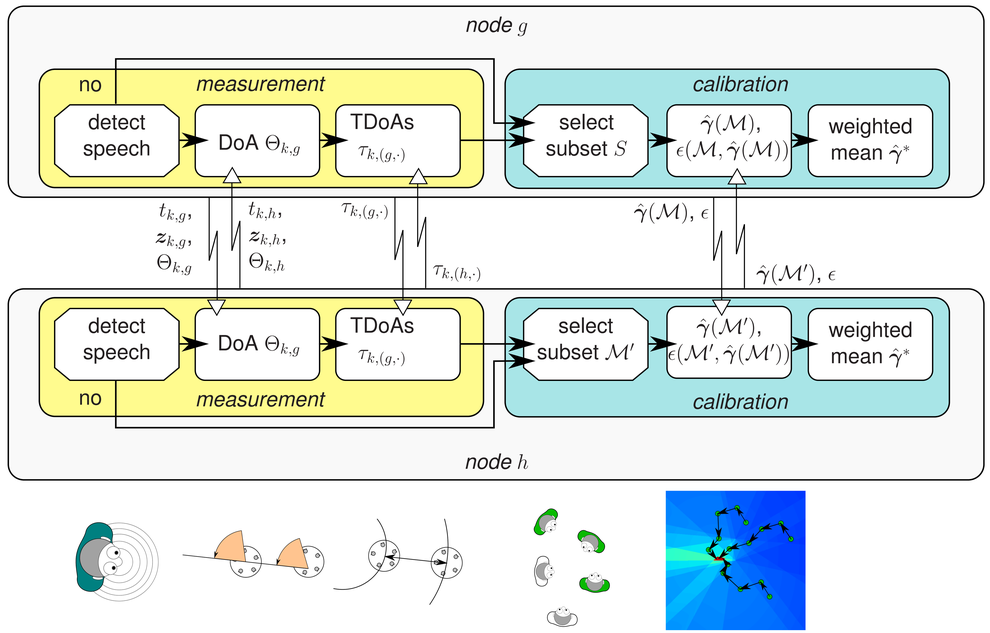
Combined over nodes and subset of all events

- Underdetermined for a single event $\rightarrow$ need min. number
- Nodes select random subsets $\mathcal{M} = \{\boldsymbol{m}_{i,(\cdot,\cdot)}, \boldsymbol{m}_{j,(\cdot,\cdot)}, \dots\}$ for distributed computation of $\hat{\boldsymbol{\gamma}}(\mathcal{M}) = \operatorname{argmin}_{\boldsymbol{\gamma}} \epsilon(\mathcal{M}, \boldsymbol{\gamma})$
- Combine these by weighted mean for overall estimate $\hat{\boldsymbol{\gamma}}^*$

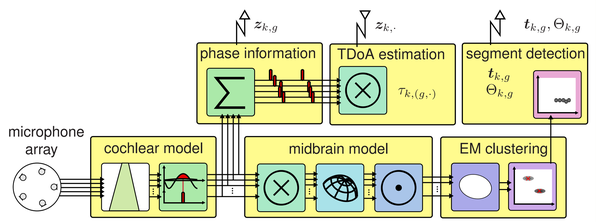
Further Materials

- paper & appendix `http://patrec.cs.tu-dortmund.de/resources`
- code `https://github.com/Plinge/audiogeocal`
- demo video `https://vimeo.com/177715229`

Method Overview



2: Neuro-Inspired SpeechLocalization

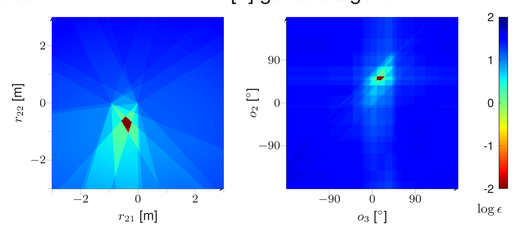


- neuro-inspired model for speaker localization [3]
- detect speech segment k with time span $\boldsymbol{t}_{k,g}$ & DoA $\Theta_{k,g}$
- also transmit sparse spike representation $\boldsymbol{z}_{k,g}$
- TDoAs are estimated by using $\boldsymbol{z}_{k,\cdot}$ from other nodes

3: Genetic Algorithm

Target Function $\epsilon(\mathcal{M}, \boldsymbol{\gamma})$ to minimize

- $\boldsymbol{\gamma}$ high-dimensional $((G-1) \times (d+1))$
- non-intersections cause plateaus
- highly non-continuous and non-differentiable
- use differential evolution [4] genetic algorithm



Cuts through the target function $\epsilon(\mathcal{M}, \boldsymbol{\gamma})$ for an example with $G = 3$

References

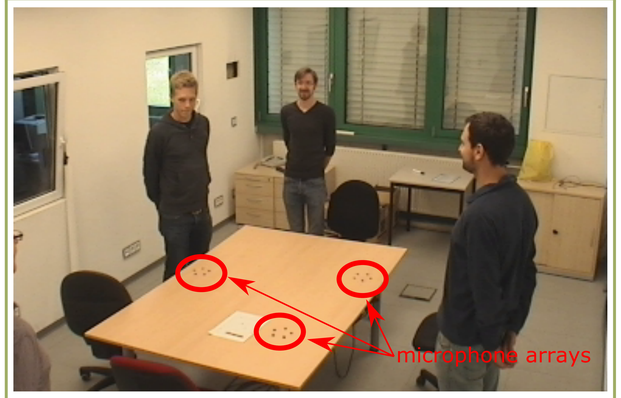
[1] A. Plinge, F. Jacob, R. Haeb-Umbach, and G. A. Fink. Acoustic microphone geometry calibration: An overview and experimental evaluation of state-of-the-art algorithms. *IEEE Signal Process. Mag.*, 33(4), July 2016

[2] A. Plinge and G. A. Fink. Geometry calibration of multiple microphone arrays in highly reverberant environments. In *Int. Workshop on Acoustic Signal Enhancement*, Antibes – Juan les Pins, France, Sept. 2014

[3] A. Plinge and G. A. Fink. Online multi-speaker tracking using multiple microphone arrays informed by auditory scene analysis. In *European Signal Process. Conf.*, Marrakesh, Morocco, Sept. 2013

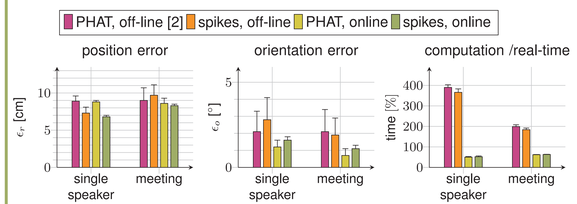
[4] R. Storn and K. Price. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal on Global Optimization*, 11:341 – 359, 1997

Evaluation

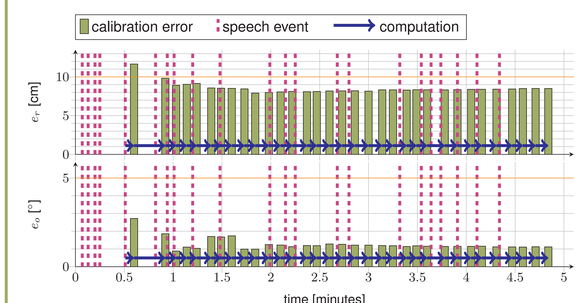


- Smart conference room ($T_{60} = 0.67\,\mathrm{s}$)
- Three 5-mic uniform circular arrays in table (48 kHz)
- Sequence #1 single speaker at $K = 19$ sitting / standing positions
- Sequence #2 four persons in a meeting ($K = 22$, picture above)
- 100 Monte-Carlo runs
- Subset size $|\mathcal{M}| = 5$
- Population size 10 for genetic algorithm

Method Comparison



Example Run (Meeting)



Conclusion

The proposed approach is able to calibrate an acoustic sensor networks' geometry online during a meeting in a reverberant conference room.