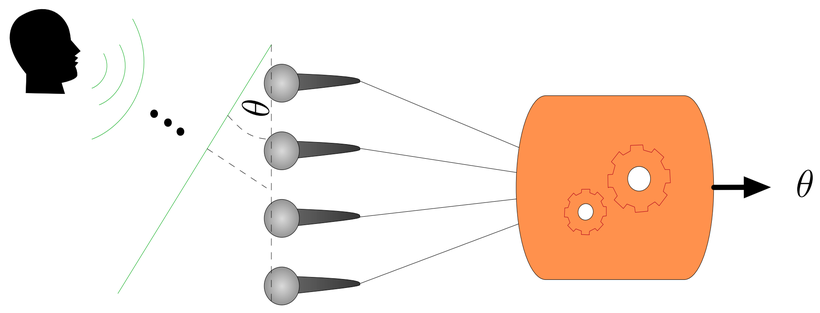
INTERNATIONAL AUDIO LABORATORIES ERLANGEN
A joint institution of Fraunhofer IIS and Universität Erlangen-Nürnberg

Python-Based Real-Time Direction-of-Arrival Estimation Using Deep Learning

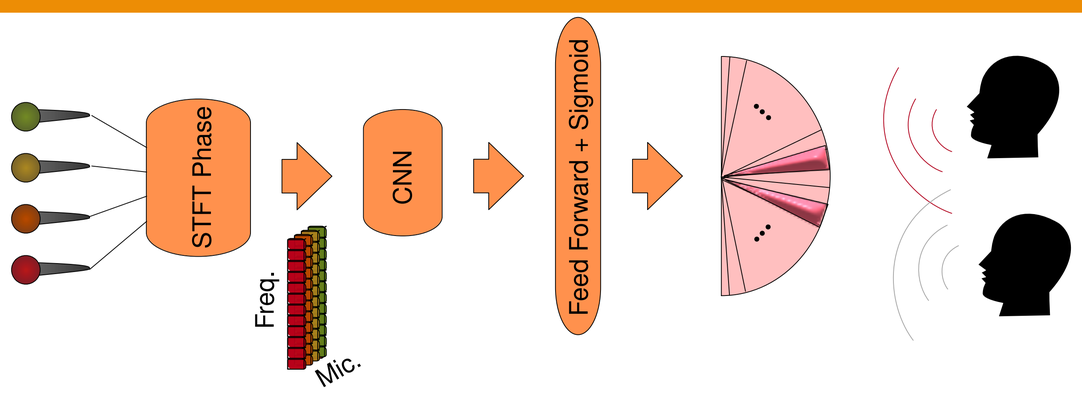
Wolfgang Mack, Soumitro Chakrabarty, Emanuël A. P. Habets

1. Introduction



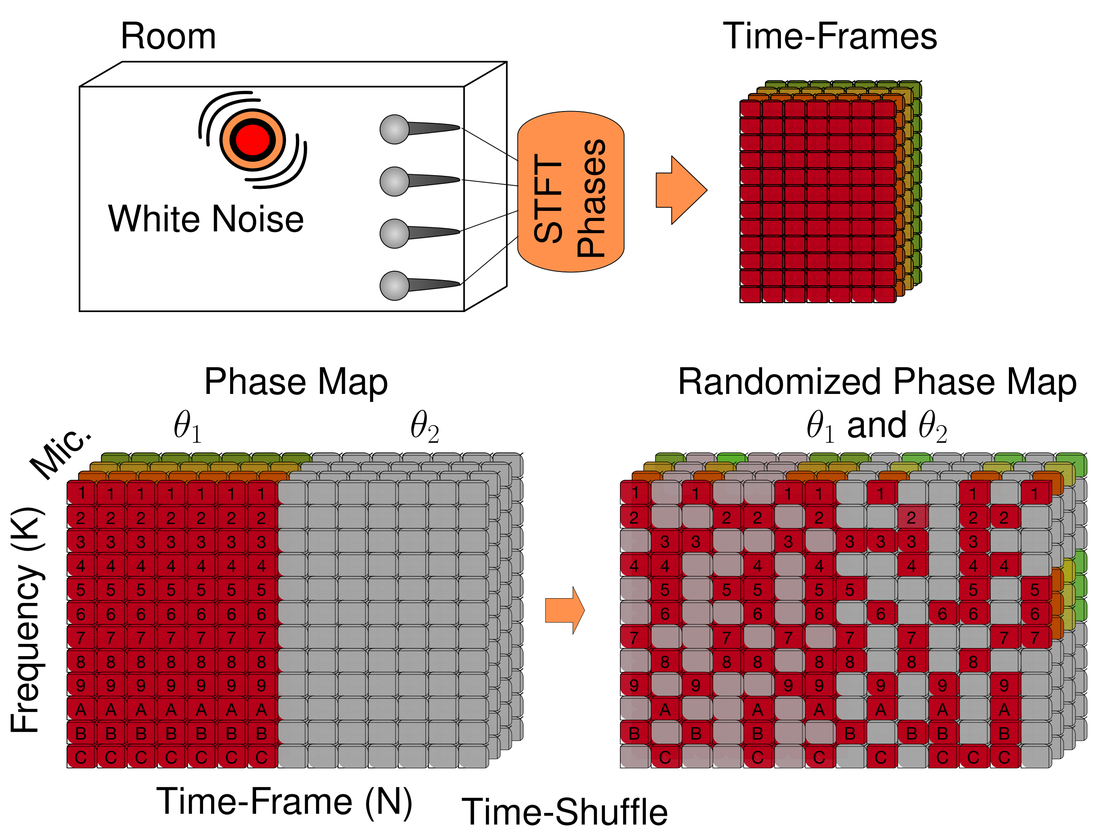
- An uniform linear array (ULA) with four microphones captures direcitonal, reverberant sound sources.
- The inter-microphone distance is 8 cm.
- The direction-of-arrival $\theta \in [0, 180]$ is to be obtained from the microphone signals (one or more sound sources).

2. DOA Estimation [1, 2]



- Compute the STFT-phases of the microphone signals.
- Feed the phases of a single time-frame into a CNN with subsequent feed-forward layers.
- The phases are mapped to 37 DOA classes, which represent an angular DOA resolution of 5 degrees.
- Multiple sources can be simultaneously localized via classification.

3.Training Data Generation [1, 2]

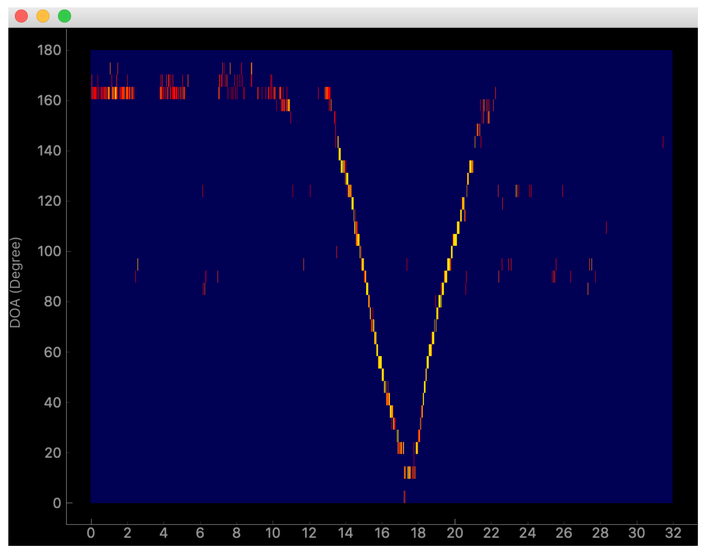


- Training data is simulated with white noise sources convolved with simulated room impulse responses.
- To simulate concurrent sound sources, a time-shuffle is applied to the phase maps of two sources.

4. Parameters

- Trained with simulated room impulse responses of 6 rooms with dimensions $(6 \times 6 \times 2.7)$, $(5 \times 4 \times 2.7)$, $(10 \times 6 \times 2.7)$, $(8 \times 3 \times 2.7)$, $(8 \times 5 \times 2.7)$,
  with reverberation-times $\in \{0.3\ s, 0.2\ s, 0.8\ s, 0.5\ s, 0.7\ s, 0.8\ s\}$ and 7 positions per room.
- STFT-parameters: 16 kHz sampling frequency, 10 ms hop-length, 32 ms window length, Hann-window.

5. Demonstration



Visualization 1

- DNN output of all 37 DOA classes over time (update-rate 10 ms).
- Output values below 0.5 are set to zero.

Visualization 2

- Polar plot of the 37 DOA classes (update-rate 10 ms).
- DNN output averaged over 5 time-frames (72 ms).

6. Ask for ...

- ... offsets of the microphones to show robustness.
- ... varying inter-microphone distances.
- ... different audio sources (e.g., clapping, clinking, speech, etc.).

[1] Soumitro Chakrabarty and Emanuël A. P. Habets, "Broadband DOA estimation using convolutional neural networks trained with noise signals," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2017, pp. 136–140.

[2] Soumitro Chakrabarty and Emanuël A. P. Habets, "Multi-speaker DOA estimation using deep convolutional networks trained with noise signals," *IEEE J. of Sel. Topics in Signal Processing*, vol. 13, pp. 8–21, Feb. 2019.

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Fraunhofer IIS