Sound Source Separation, Localization, and Tracking of Two Moving Agents with Two Microphones

Xinhang Song and Yifan Zhu
Final Project for CS 598 PS, 2019 Fall, University of Illinois at Urbana-Champaign

Problem Statement

We would like to track the position of two moving- and talking- people using a pair of omni-directional microphones in an indoor environment.

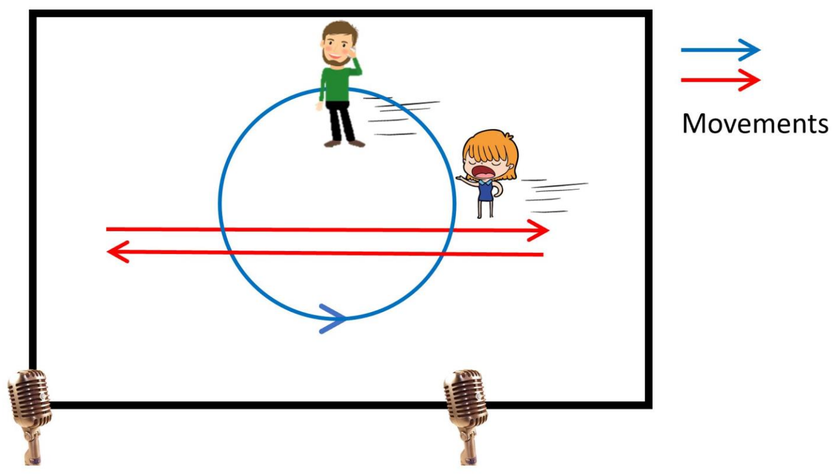

Figure 1. A male and a female talking and walking in a room, with a pair of microphones recording sound. This is also the setup of our simulator which generates delayed and mixed sounds with reflections.

Method

The overview of our method is shown in Figure 2. The localization is performed at 20Hz.

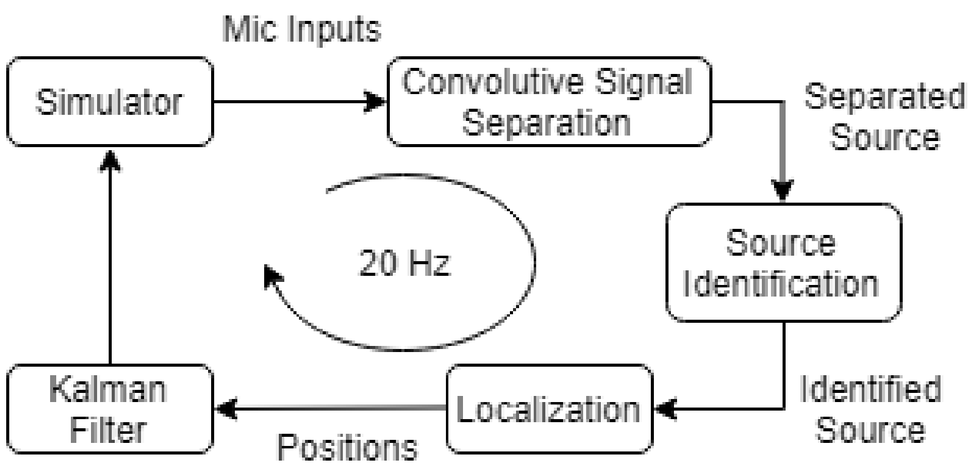

Figure 2. Overview of our method.

Simulator
- The setup of our simulator resembles that in Figure 1.
- Agents move at periodic patterns continuously and pre-recorded sounds are used.
- Sound magnitude diminishing and delay are simulated. (Reverberation not implemented yet.)

Convolutive Signal Separation
- The algorithm is based on [1], introduced in class lecture notes.
- Separate delayed and mixed signals
- FIR matrix notation

$$\underline{\hat{X}} = \underline{\hat{A}} \cdot \underline{\hat{S}}$$

$$\Delta \underline{A} \propto (I - f(\underline{A} \cdot \underline{x}) \cdot (\underline{A} \cdot \underline{x})^T) \cdot \underline{A}$$

Source Identification
- We use a pre-trained classifier for male and female voice based on data we collected.
- Frame-wise classifier in the frequency domain + voting.

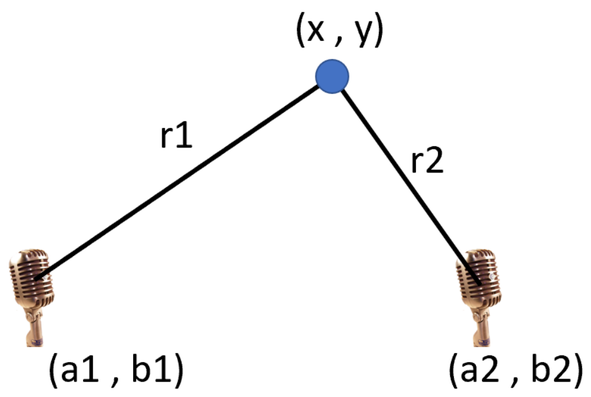

Figure 3. Localization of an agent. The difference in time-of-arrival is r1-r2, while the magnitude ratio is r1/r2, which together gives the value of r1 and r2. With the locations of the mics known, the x,y positions can be triangulated.

Methods – continued

Localization
- Separated sound signals of the 2 sources from 2 mics are used to compute the difference in time-of-arrival and magnitude.
- Maximum cross-correlation used for estimated time-of-arrival difference.
- Triangulation is performed to compute the x-y location of the agent.

Kalman Filter
- Two independent Kalman filters are used for the two agents.
- The state includes the position and velocity of the agent. The system dynamics matrix and observation matrix are listed below, the measurement noise matrix is diag(0.1).

$$state = \begin{bmatrix} x \\ y \\ \dot{x} \\ \dot{y} \end{bmatrix} \quad A = \begin{bmatrix} 1 & 0 & dt & 0 \\ 0 & 1 & 0 & dt \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Current Results

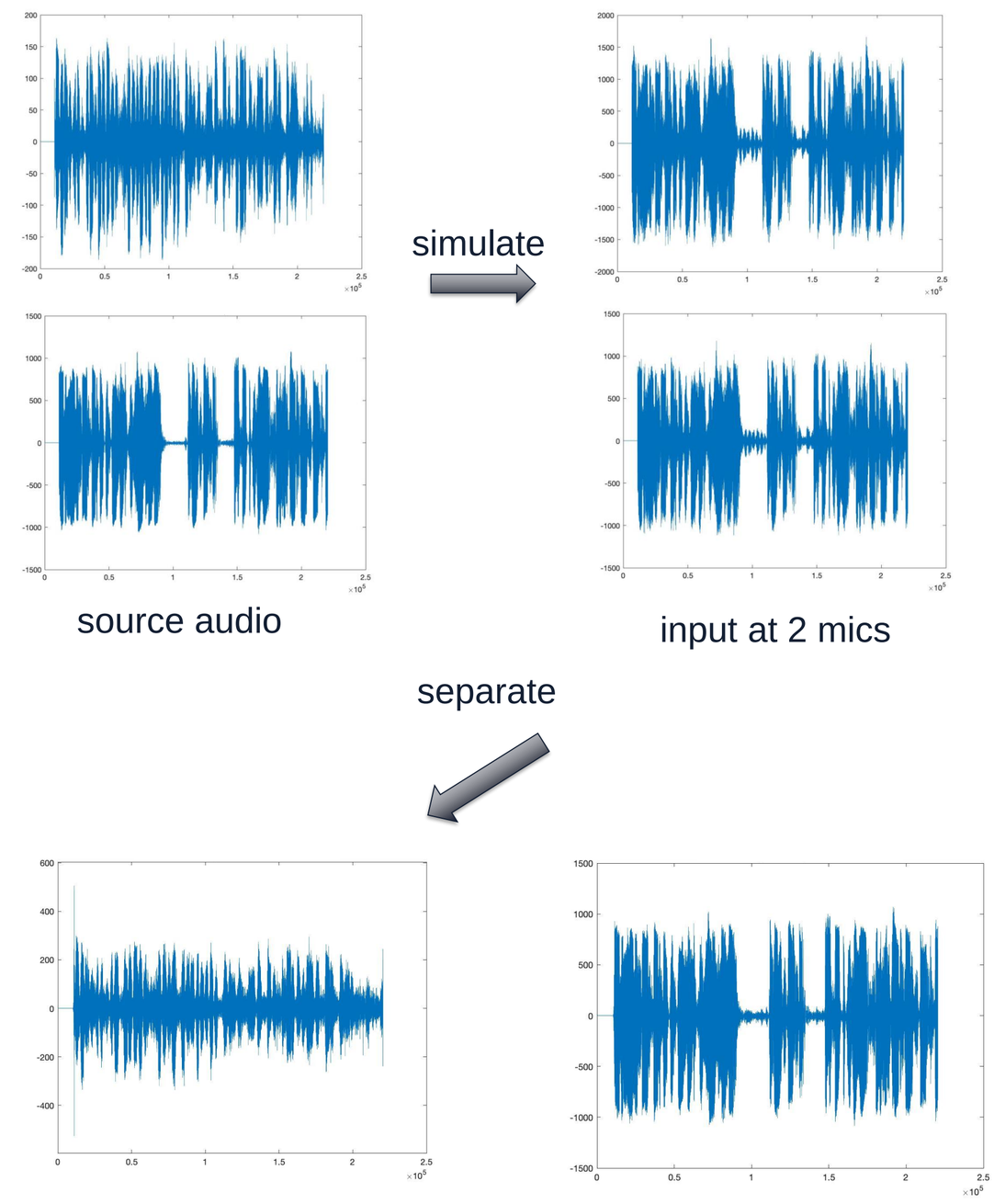

Figure 4. Ground truth, mixed and separated signals.

Conclusions and Future Work

- The current separated signals are not ideal, which makes the estimation of sound magnitude ratio and difference in time-of-arrival difficult. Therefore, we have not been able to robustly estimate the location of the moving agents and run tests
- We suspect the reason is that we have not modeled the movement of the agents
- Make FIR matrices time-dependent and re-evaluate our algorithm

Reference

[1] Smaragdis, Paris. "Blind separation of convolved mixtures in the frequency domain." *Neurocomputing* 22.1-3 (1998): 21-34.

ILLINOIS