Improving accuracy of sound reflection estimation using neural networks

TU Delft

Author: **Ekko Scholtens** (E.Scholtens-1@student.tudelft.nl)
Supervisor(s): **Jorge Martinez Castaneda** (j.a.martinezcastaneda@tudelft.nl), **Elmar Eisemann** (e.eisemann@tudelft.nl)

1) Introduction

The acoustics of a room can be divided into three main parts as shown in Figure 1:

- **Direct sound**, the sound coming directly from the source.
- **Early reflections**, which are the first reflected sounds by, for example, walls.
- And **late reflections**, the reverberation of a room affects how spacious the room feels.

Early reflections in a small room can cause unwanted effects. Such as the creation of phantom images as illustrated in Figure 2. There are problematic as there are now 3 sources instead of the one.

Filtering all reflections is not preferred. Detection where the early reflections come from can help with filtering only the problematic ones.

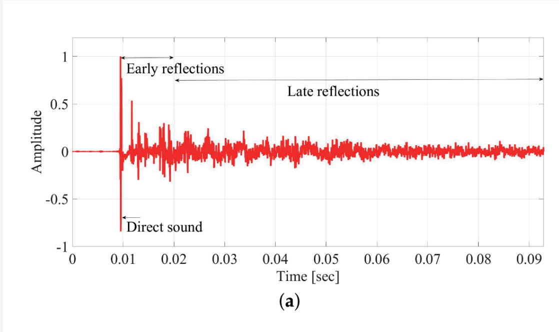

Figure 1: Real RIR behaviour [1]

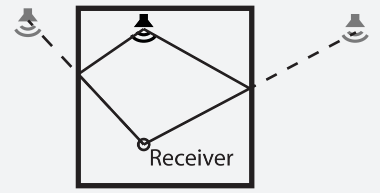

Figure 2: phantom images

2) Background

A naive approach could be to try and match reflections to each other as depicted in Figure 3.
This can be reduced to a “maximum independent set” problem, which is a NP-Hard problem.

Other solutions are in general:

- Expensive
- Designed for large music halls
- Require 10+ numbers of microphones

A neural network approach as described by Bologni [2]:

- Only used two microphones and a trained neural network to estimate direction
- But it does have issues with differentiating between front and back

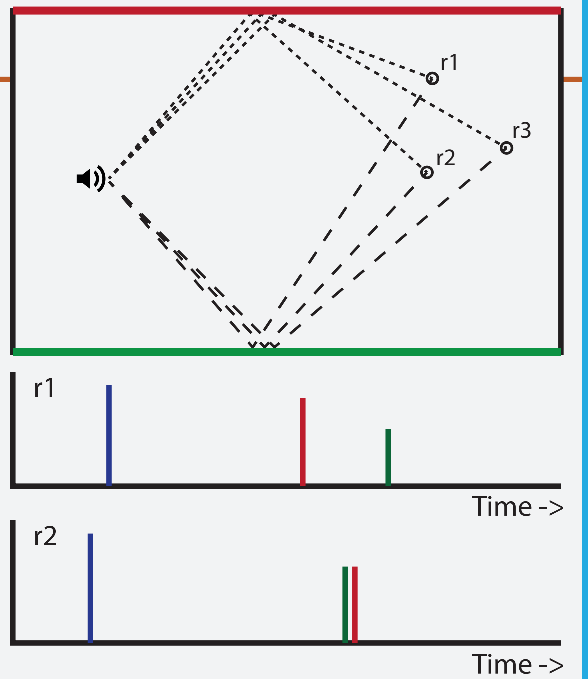

Figure 3: reflections

3) Research question

This leads us to the following questions:

- Can the addition of a third microphone to the neural network reduce the front-back ambiguity?
- Can the addition of a third microphone to the neural network improve accuracy in detecting reflections in a room?

4) Method

We generate the data using the same constraints as Bologni et al.[2] and simulated the microphone array as an array of three subcardioid microphones with an inter microphone distance of 10 cm.

The dataset is generated using the image source model and pyroomacoustics it consist of:

- 9 Rooms with sizes from 2x2 to 8x8 meters
- For each we generate a number of source and receiver location
- For each receiver position we generate 150 random rotations

We reimplemented the network as described in the paper by Bologni et al.[2].

We made two versions of the neural network:

- A the two channel version to compare against the results Bologni et al.[2] found.
- And a three channel version where we added a third microphone.

5) Results

In Table 1 shows that we increased the amount of detected sourced by adding a third microphone .

	Bologni et al.	two-channel	three-channel
Detected sources	49%	43%	58%
Front-back	–	25%	18%

Table 1: Results comparison

In Figure 4 we show a comparison of the angular error of the two(a) versus three(b) microphones. It shows that adding a third microphone, leads to a smaller standard deviation of 44.3 degrees compared against 51.5 degrees.

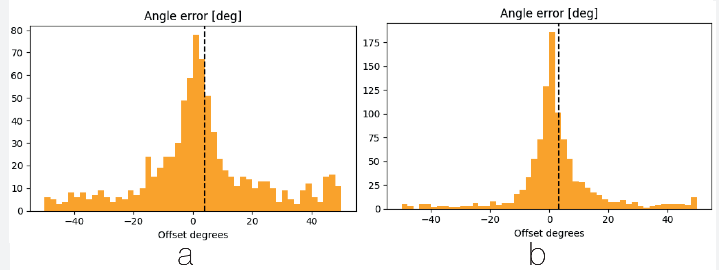

Figure 4: Angular error comparison

6) Conclusion and Future work

The results show that **adding a third microphone can significantly improve the accuracy** by detecting more sources and reduces the front-back ambiguity.

Figure 5 shows a single sample comparison between the expected and the predicted output. It shows that the predicted result could be used to find reflections in a room, and in turn find place where to place sound absorption panels.

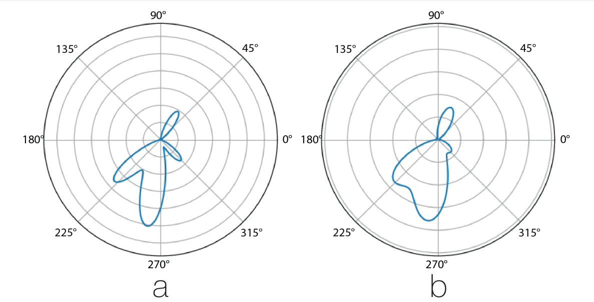

Figure 5: Expected(a) and predicted(b) angles

Future work:

- Evaluate the proposed method on real-world rooms.
- Train and evaluate on more complex rooms shapes and acoustic properties.
- Change the neural network to only output the angles of the reflections.

References

[1] Cecchi, Stefania, Alberto Carini, and Sascha Spors. "Room response equalization—A review." Applied Sciences 8.1 (2017): 16.
[2] Bologni, Giovanni. "Room geometry estimation from stereo recordings using neural networks." (2020).