Sound Based Localization for Indoor Robots

Saurav Gyawali, Faizan Muhammad

May 7, 2019

Abstract

Sound based localization for indoor robots aims to solve the initial localization for the "Kidnapped robot" problem. This problem occurs more than often in real life and mobile robots need human help to estimate the location. Our tool provides this initial estimate to the robot in a room level. To achieve this we use a room impulse response, measured using microphone and later processed for classification. This response is usually unique to a room and enough to tell multiple rooms apart. With such data, we use a classifier to tell which room the robot is in.

1 Introduction

Current localization systems for mobile robots involves expensive sensors and computationally expensive algorithms. For example, data from stereoscopic camera is used to localize a robot against a pre-existing map. This method is very computationally heavy and the algorithm involved is complicated. Even with such complex method, the initial localization is mostly very unreliable. Hence, there is usually some human involvement to help the robot with the estimate.

The initial localization problem occurs more often than we want in real life. The human involvement is not very convenient in many situations. Therefore there is a need of reasonable tool in terms of computation and infrastructure to give a reliable prediction. Even having such tool to augment the already existing methods can prove to be a great time and effort saver.

Our project aims to solve parts of this problem by giving an initial estimate of the location on a room level. To achieve this, we are using audio based techniques. We measure the room impulse response, train a classifier on this data and finally predict which room the mobile robot is in.

Moreover, our project also investigates estimation of location inside the room. However, we are inconclusive in this aspect.

2 Background and Related Work

Many different methods and sensors are used for localization of mobile robots including 3D-cameras, Lidars and GPS however indoor environments often present a a challenge to these conventional methods. The GPS signal is often attenuated by the walls and not precise enough for navigation anyway. Whereas Lidars and other distance based measures are influenced heavily by changing indoor layouts and people or straight up lack of features (long straight corridors). Image-based localization systems are computationally very expensive to work in real time.

One of the more novel ways to approach this problem is by using sound. A team of researchers in UC Berkeley developed Soundloc [2] where they used the difference in sound distortion in different rooms to identify them. They emitted certain sounds and recorded their reflections and distortions and used that to find out Room Impulse Response (RIR).They then processed the multiple values of RIR per room to get features like standard deviation and kurtosis. Using these features, they were able to identify rooms to 97% accuracy for 10 rooms and 1000 samples. They used an ordinary laptop speaker and microphone for their experiments.

They model a room as a being a certain linear time-invariant system that takes in the sound and outputs it in a certain way. As represented by Figure 1 taken from the original paper, here h(t) represents the Room Impulse Response or RIR.

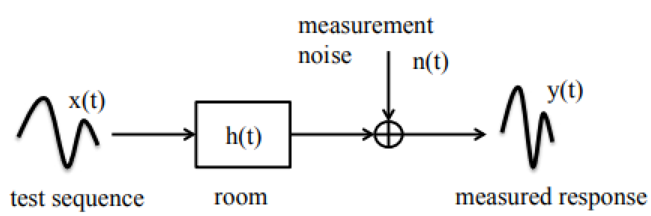


Figure 1. The room can be modeled as a linear time-invariant system and the received signal is a linear convolution of test excitation sequence and room impulse response.

Once RIR is obtained they extract several features from it including Temporal Kurtosis (Figure 2 taken from the paper) and Standard Deviation (Figure 3 taken from the paper)

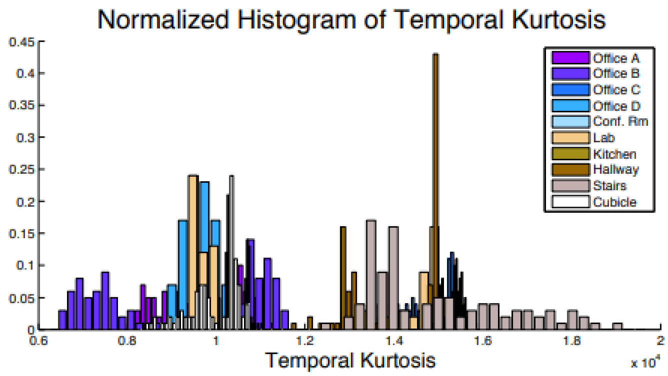


Figure 2. Normalized histogram of temporal kurtosis. Closed spaces with small volumes such as offices, conference room, and cubicle have smaller temporal kurtosis. Open spaces such as kitchen and hallway or closed spaces with large volume such as stairs have large kurtosis.

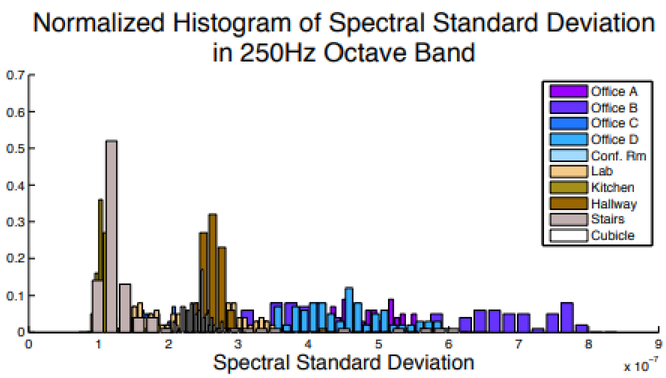


Figure 3. Normalized histogram of spectral standard deviation in octave band centered at 250 Hz. Places covered with carpet such as offices, lab, hallway and cubicle exhibit a larger spectral standard deviation. Places without special sound reduction such as stairs and kitchen shows a smaller spectral standard deviation.

The creators of the Batphone app [3] used a different approach to accomplish indoor localization. Instead of attempting to profile the room's sound reaction, they attempted to profile the room's ambient noise and use that to localize the device. While the results are impressive based on how difficult the task is, they suggest ambient noise can only be used as an extra enhancement with another method and not the main method on its own.

3 Problem Formulation and Technical Approach

Our goal was to investigate the approaches we can use to get as much meaningful information out of the RIR as possible and use them to get objective results about our approach.
We used Matlab for all data-collection, data-processing and machine learning. The code and the datasets are available at [1]. We started with first developing the RIR detection by modifying [5].
We use the digital signal processing tools in Matlab to generate the RIR. First we create a sine-sweep wave. This is a collection of sine waves starting from a low frequency and ending at a higher frequency. This wave is played through the speaker and simultaneously the audio response is recorded using the microphone. Discrete Fourier Transform is performed on this audio reponse and the original sine-sweep. The two transforms are combined with division operation and an inverse Discrete Fourier Transform is performed in the result. This result is the required RIR.
Then we used the RIR to create some simplified features. We fit an exponential curve to the amplitude of the RIR and used the coefficients of that curve as features. We also calculated a reverb time value which is the time it takes for the amplitude to decrease to a 100th of its initial value. Furthermore, we calculated the specteral deviation, which is the standard deviation of the frequency response of the room. Figure below shows a sample RIR:

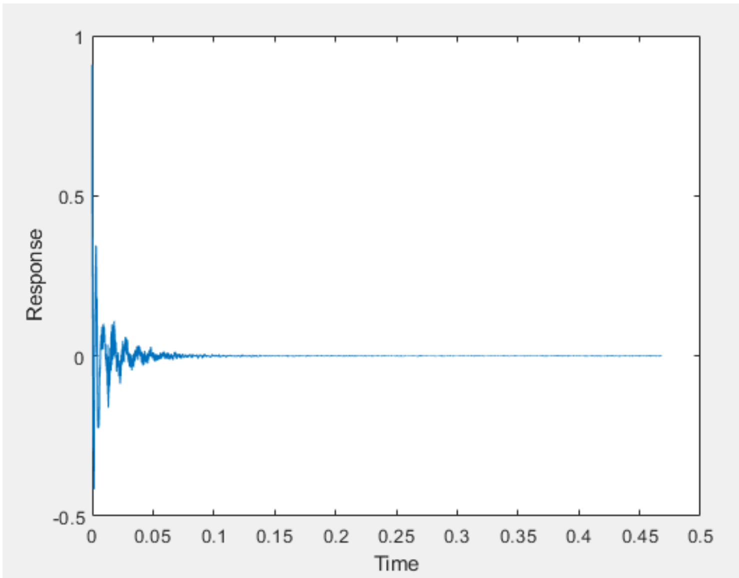
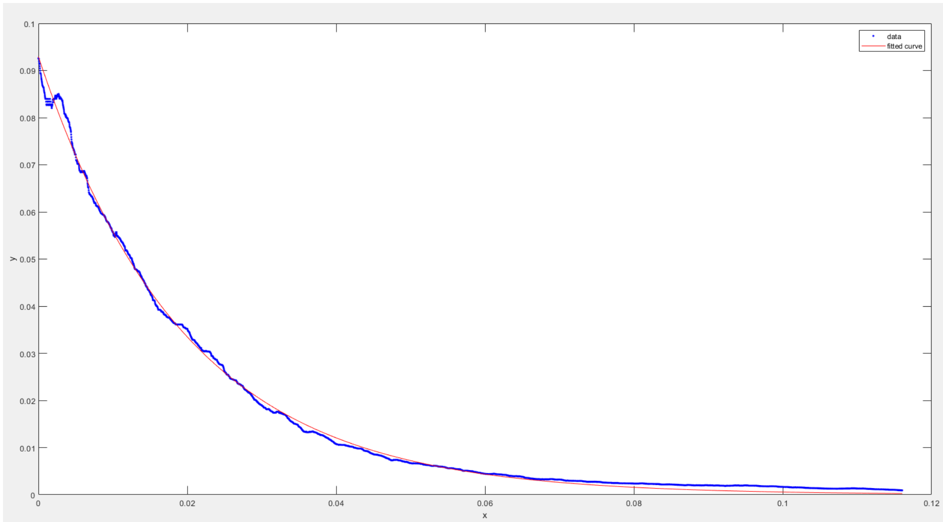
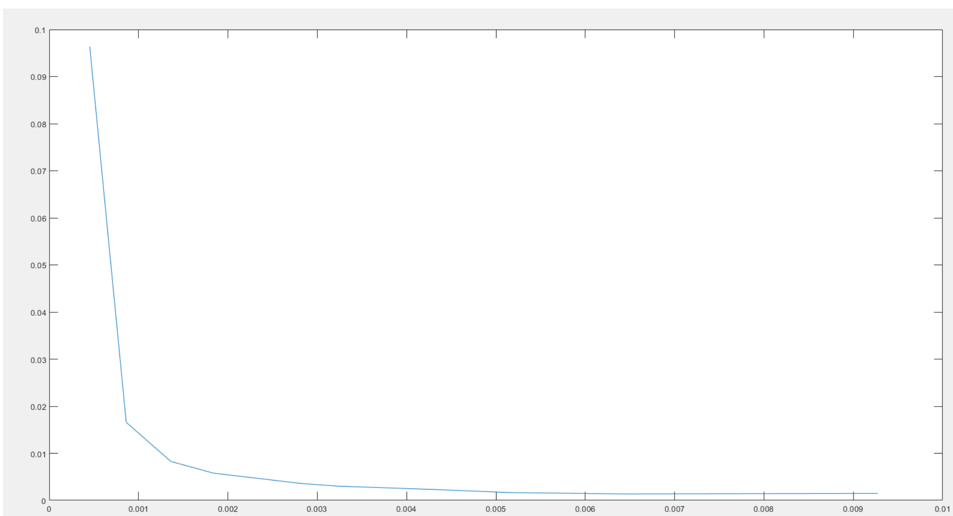


Figure below shows the amplitude of the RIR that has been smoothed and then an exponential curve fit to it:



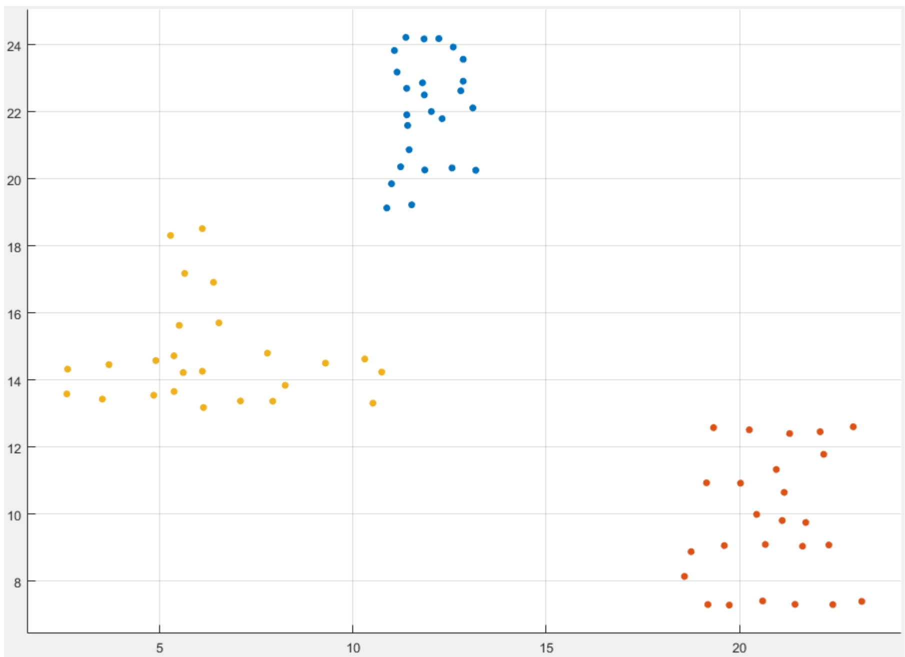
Another alternative approach we developed was to simplify the amplitude of response itself into just 20 points and use that these values as our features as shown below:



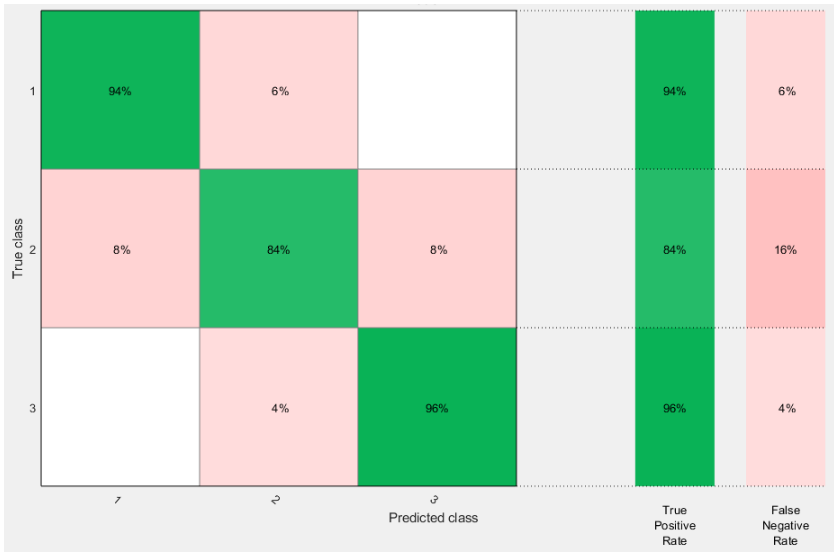
4 Experimental Results

We used the simplified features in our initial dataset collected in a house with three distinct rooms (10 samples each) and found that it gave us 100% accuracy using an SVM with 5-fold cross-validation. This assured that our setup was functional and there was meaningful information within our data.

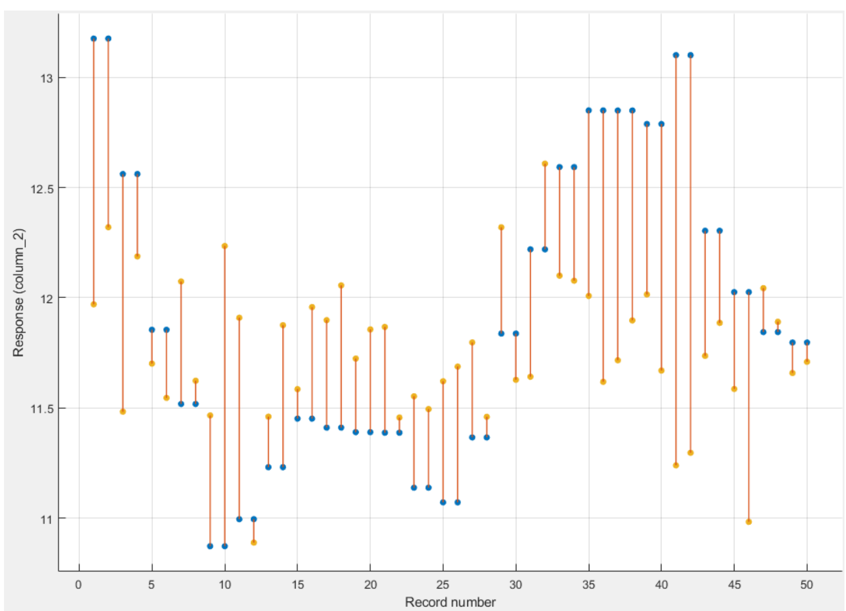
We then began collecting samples in Halligan for the main dataset. We picked three locations: a small lab room, an open lounge space and a corridor intersection. We took 50 samples in each location in 25 different position within the location and 2 samples per position. The figure below shows the positions we sampled. Yellow representing the corridor (Class 3), blue representing the small lab room (Class 1) and red representing the open lounge space (Class 2).



Using this dataset, we first tried our previous SVM approach with simplified features and we only got 76% accuracy. So we used our alternative approach with using the simplified response itself and that gave us 91.3% accuracy. Figure below shows the confusion matrix obtained with this classifier. As it can be seen, the main source of error is Class 2, which was the open lounge space. Because it is open on two sides and merges directly with it seems that it's acoustics have a very wide variation that resembles both the corridor intersection and the lab room. Furthermore, it can also be seen that there is no confusion between Classes 1 and 3.



The next step was to investigate if given the room impulse response and the room label, if it is possible to localize within the room based on the room impulse response. To test this idea, we used regression and provided the room label, y position and the features to train for the x position. The best results were obtained with a Linear SVM however, they were not good enough to claim that these RIR features contain room position information. In the figure below, blue values represent true values and yellow are the predicted values.



5 Conclusion and Future Work

From our investigation so far, it is fair to hypothesize that room labels can be accurately identified based on the RIR and other acoustic information gathered by the robot. And this information can be used as a starting point for other localization techniques such as particle filters for the Kidnapped Robot problem.

We were unable to find any evidence that this acoustic information can also predict location within the room.This might be improved by a multi-microphone setup that can take into account directional echoes and thereby use that information to localize within the room.

Overall, we believe the accuracy of this system can be improved much more and scaled to multiple rooms in the future. This can be done by identifying more features within the data, noise removal, different audio sequences and incorporating with other sensory information.

References

[1] https://github.com/gyawalisaurav/AudioLocalization

[2] Jia, Ruoxi, Ming Jin, and Costas J. Spanos. "Soundloc: Acoustic method for indoor localization without infrastructure." arXiv preprint arXiv:1407.4409 (2014).

[3] Tarzia, Stephen P., et al. "Indoor localization without infrastructure using the acoustic background spectrum." Proceedings of the 9th international conference on Mobile systems, applications, and services. ACM, 2011.

[4] https://commons.wikimedia.org/wiki/File:Acoustic_room_impulse_response.jpeg

[5] https://www.mathworks.com/matlabcentral/fileexchange/28288-archivox-room-impulse-response-measurement