CS545 FA21: Indoor Localization with Audio Signals

Beitong Tian
beitong2
beitong2@illinos.edu

Xiaoyang Wang
xw28
xw28@illinois.edu

January 24, 2022

Abstract

Precise indoor localization is desired in many application scenarios such as smart home, smart merchandise, and smart laboratory. In this project, we propose an innovative localization system with audio signals as input. We believe this is the first work to use both air-borne and structure-borne signals and explore their complementary to improve the localization accuracy. We design and build the real-time data acquisition system and data processing pipeline and perform various preliminary experiments on it with a one-dimensional setting. The results verify findings in the previous works and also expose additional problems and challenges which are not explicitly considered in other similar systems. Based on the results, we refine our analysis algorithms and will deploy them in our system in future work.

1 Introduction

Locating a person in an indoor environment is attractive. Examples include elder monitoring in the nursing home and operator monitoring in clean rooms. These kinds of monitoring could discover emergencies in an early stage, such as entering dangerous or prohibited areas. Many previous works have explored the possibility of using radio frequency (RF) signals, video streams, on-body devices, vibration signals, and audio signals to do indoor localization. However, these methods have their limitations and trade-offs and can only be used in a specific indoor environment. For example, the RF-based method will suffer from the non-line-of-sight (NLoS) problem and degrade its performance. The wearable device-based approach requires the user to purchase and wear the specific device, which is cumbersome. Camera-based systems are costly and are not affordable to many facilities. Compared to previous methods, we want to build an innovative system to localize a single person in a general indoor environment with high accuracy and low cost using sound and vibration signals collaboratively. Specifically, combing sound and vibration signals could mitigate the NLoS problem.

The main observation used in this project is when people are walking in an indoor environment, their shoes strike the floor and generate sounds and vibrations. The generated sound signal (also called air-borne signal) will propagate via the air and arrive at a location

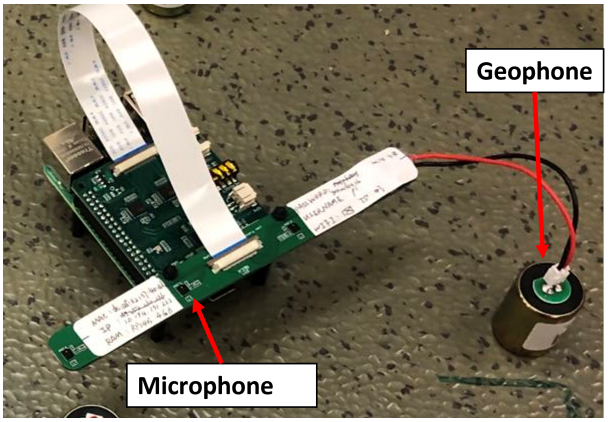

Figure 1: Microphone and Geophone hardware

after some time, which is proportional to the distance between the location of the footstep and that of the sensor. Therefore, we can deploy microphones at different locations to capture the sound and calculate its arrival time. The difference of arrival time across different microphones can be used with the famous time difference of arrival (TDoA) algorithm to calculate the location of the footstep, which also represents the person's location. The vibration signal (also called structure-borne signal) will propagate via the floor. It can be captured by the geophone, which is specially designed to capture the underground vibration, as shown in Fig.1. The vibration signal can also be used with the TDoA algorithm to localize people.

Previous works use the sound and vibration two signals separately. These resulting systems are reported with high accuracy. However, according to their settings and results, we expect that systems using only sound or vibration signals will suffer from different problems under some conditions. For example, the sound signal will suffer from a low signal-to-noise ratio (SNR) problem when there is strong background noise and an NLoS condition when there is an obstacle between the sound source and the microphone. These problems likely result in significant localization errors. On the other hand, under the conditions where the SNR and NLoS problems do not appear, we can get a very high localization accuracy because the sound signal has a stable propagation speed compared to the vibration signal.

In contrast, the vibration signal is more robust to the air-borne noise and the influence from obstacles. However, the propagation speed of the vibration signal may vary with: (1) The change of the signal frequency. (2) The change of the texture of the floor. (3) Whether there are heavy objects or vibrating objects on the floor. We need to model the propagation speed based on the testing environment to guarantee localization accuracy. However, the modeling process is labor-consuming and error-prone.

From the above discussion, we conclude that the sound signal is vulnerable to noise and obstacles but has high accuracy when the environment is ideal. In contrast, the vibration signal is robust to noise and obstacles but needs an accurate propagation speed model.

The problem we want to solve in this project is: can we collaboratively use these two kinds of signals to boost indoor localization accuracy?

Our main contributions in this project are:

1. We develop and build a distributed data acquisition system to collect, process and

visualize the signals.

2. We report our findings after testing our system in a real office environment with a one-dimensional setting.

3. Based on our preliminary results, we propose a new data processing pipeline.

2 Proposed Methods

2.1 Data Acquisition System

One of the main contribution of our project is our data acquisition system. Unlike projects working onacquistion existing datasets, there is no existing dataset for our to use. Hence, we need to collect the data by ourselves.

2.1.1 Overview

Unlike previous works which connect the output of each sensor to one centralized data collector which is bulky and unpractical, we build a distributed system where each sensing module sends the data to the server via network. We have four identical modules as shown in Fig.1. These modules are embedded with high performance analog to digital converter to collect sound signals from microphones and geophones. The collected data are streamed to the a laptop via WiFi or Ethernet. The laptop, working as a server, synchronizes the data, processes the data and display useful results in an graphical user interface (GUI).

2.1.2 Time Synchronization

One problem with using a distributed sensing system is we need to synchronize the signal from each sensing module. If we have one-millisecond offset, we will have a $0.001 \times 340/2 = 17$ cm error. We have built a system to test the time synchronization error. In our experiment, we find the minimum error we can get with a wireless network environment is 2 milliseconds, and the variance is large. We finally choose to use a wired network like Ethernet and precision time protocol (PTP) to synchronize the time. With this method, the error can be kept under 0.1 milliseconds, so the introduced localization error can be ignored.

We are also thinking to use some events with known locations such as door open to synchronize the time. With the location of the door and sensing modules, we can calculate the time offset and re-synchronize the time. In this case, we don't need to use a wired network anymore where we can make our system more space-friendly.

2.1.3 Graphical User Interface

We add the GUI for helping us understand our data and algorithm outputs in real-time. During the experiment, we find we can get much more observations with the help of our GUI, and these observations can inspire us to improve our algorithm. For example, we find using different strengths to strike the floor will generate very different signals. Fig.2 shows the GUI interface.

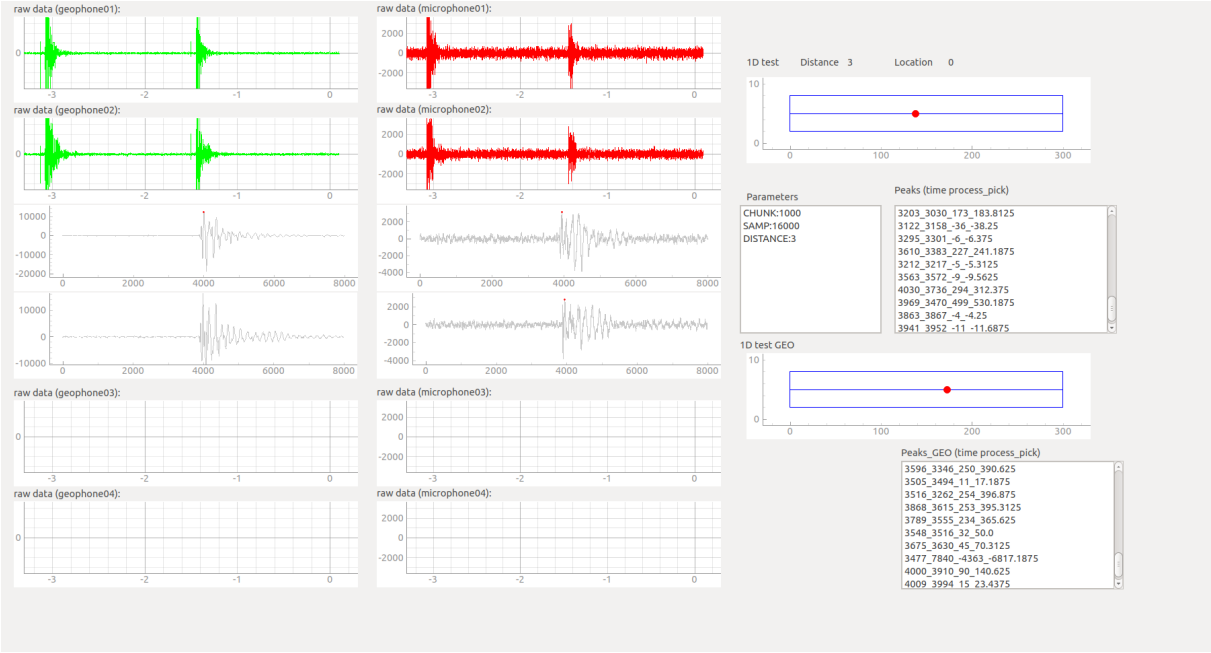

Figure 2: Indoor Localization GUI

The GUI has four regions. The top left 4 panels which show green and red signals are used for streaming real-time data from geophones (green ones) and microphones (red ones). The 4 bottom-left panels have the same functions. There is no data in these panels because we only use two sets of sensors [1] for one-dimensional setting.

The 4 panels in the middle are showing the post-processed data. Currently, we use the signal peaks to calculate the time difference of arrival. For example, in the top left one, we show the windowed (window size 0.5 second) geophone signal and use a red dot to represent its peak with a red dot.

We show the localization results in the two panels on the right side. The top one shows the microphone-only result, and the other one shows the result from geophone data. The red dot shows the output location. The blue rectangle shows the detection region. In this test, we use our heel to strike at the middle point. We can see the result calculated from microphone data is better.

We display metadata and some intermediate results in the remaining panels.

2.2 Data Process Pipeline

This section presents our system that collaboratively leverages sound and vibration signals to localize people with high accuracy. The main system design is shown in Fig.3

We will use the customized hardware platform shown in Fig.1 to capture sound and vibration signals. The signal will be processed (convert to spectrogram) and filtered (continuous wavelet transform) to remove noise afterward. We will use TDoA and angle of arrival (AoA) algorithms for the sound signals to compute the current location (x,y). During the

[1] One set of sensors contain one geophone and one microphone

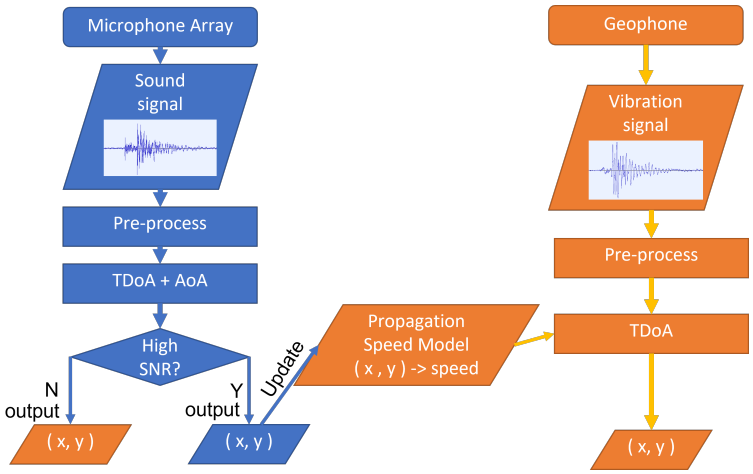

Figure 3: System Design

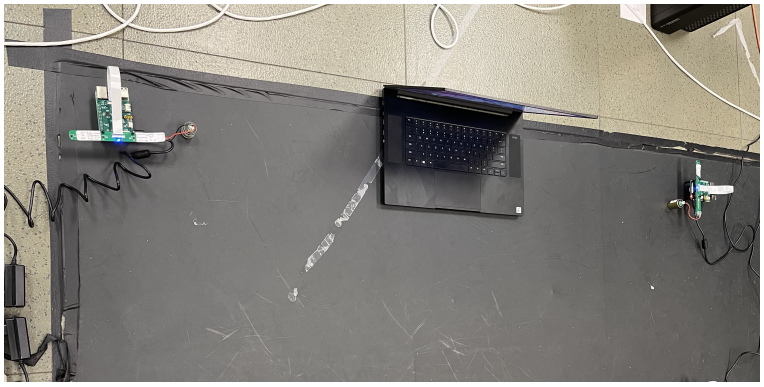

Figure 4: System Setup

computation, we will estimate a confidence score mainly based on the SNR. If the confidence score is higher than a threshold, we will use the location calculated from the sound signal as our system output and update the propagation speed model. We will run a similar process for the vibration signal and use our propagation speed model instead of a constant speed as the input of the TDoA algorithm. We will use the output location when the confidence of the sound subsystem is low, as shown in the orange block in the bottom left corner of Fig.3.

3 Preliminary Results and Findings

3.1 Experiment Setting

We set up our system as shown in Fig.4. The distance between the two sensing modules is three meters which is similar to the width of a small office room. We only test the one-dimensional setting in this project because most findings and problems are general to the two-dimensional setting.

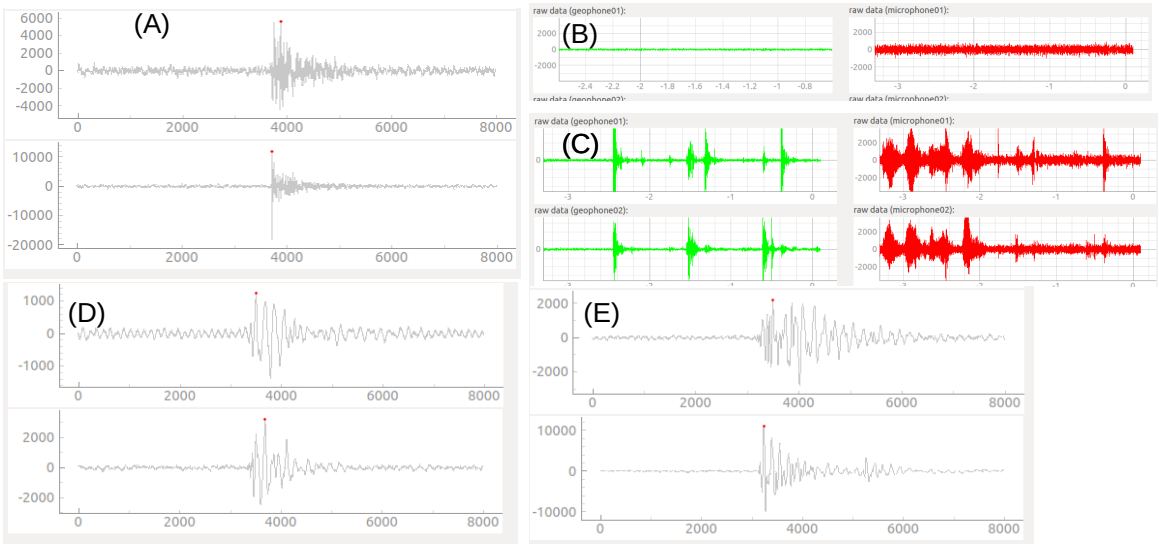

Figure 5: Results

3.2 Experiment Result and Discussion

We run some experiments to test our hypothesis and previous works results and we get the following conclusions.

1. Sound signals suffer from NLoS problems.

We block the signal between one sensor and the sound source. As shown in Fig.5(A), the top one shows the blocked signal. We can find the NLoS signal is more complex than the LoS signal due to the attenuation and reflection which makes it harder to detect the correct time difference of arrival.

2. Sound signals will be influenced by background noise.

As shown in the Fig.5(B), the background noise of the microphone signal (red) is visually larger than that of the geophone signal. We are walking and talking to simulate the normal office environment When we are capturing Fig.5(C). We can find the talking sound largely affects the SNR of the footstep sound.

3. Different footsteps (speed, strength, texture of the shoe) will influence the sound and vibration signals a lot.

When we are recording the signals in Fig.5(C)&(D), we walk in different ways. We can tell the difference between them. We need to find some features which can be used reliably as the indicator to calculate the time difference of arrival.

We focus on building the speed model for the vibration signal in our old system design as shown in Fig.3. However, after experimenting, we find getting the correct time difference of arrival is also a big challenge. Currently, we calculate the time difference of arrival with the highest peak in the signal segment. This method only works well in some cases. For example, it works when we use our heel to tap the floor gently and when we are doing finger

snapping. However, when we walk normally, the result gets worse immediately. Our next step is to find some new features to calculate the time difference of arrival.

4 New Method

We have several ideas about our new algorithms.

The first idea is background noise cancellation. We can record the background noise and subtract the noise from the input data in the frequency domain. We also find voice signals and footstep signals do not share the same band, so we can filter out voice with a low pass filter.

The second idea is we want to process our data in the frequency domain. We will use the features in the frequency domain to calculate the time difference. However, there is an uncertainty principle that makes it hard to get an accurate result in both the time domain and frequency domain. We decide to use the difference of two consecutive FFT results as our new indicator.

The remaining idea is the same as the methods we proposed in Section 2.

5 Conclusion and Future Work

In this project, we proposed a new method to do indoor localization. We built and test our system and get some interesting findings. Based on these findings, we create our new algorithms and will implement and test them in the future.

In the future, we also prepare to extend our system by adding the following features: (1) Identify people based on their footsteps. (2) Tracking people in a real-time manner. (3) Tracking multiple people at the same time. (4) Add other subsystems such as RF-based systems to see if we can further increase the accuracy.