Preliminary Investigation of Detecting Events of Indoor Objects with Smartphone Active Sound Sensing

Thilina Dissanayake, Takuya Maekawa, Amagata Daichi, Takahiro Hara
Osaka University
Osaka, Japan
{thilina.dissanayake, maekawa, amagata.daichi, hara}@ist.osaka-u.ac.jp

***Abstract*—Smartphone companies release new smartphones equipped with latest technologies every year. With new features being added to each new release, users tend to upgrade their smartphones frequently. After replacing the old smartphone with a new one, the old phone is then either recycled or left unused at homes. These smartphones are usually equipped with high quality speakers and high sensitive multiple microphones. We believe that we can detect indoor events by using these parts of retired smartphones. In this paper, we propose an indoor event detection method with the ability of not only detecting door events such as door openings and closings but also differentiating between the door events of two separate doors based on active sound sensing using a smartphone placed in an environment of interest.**

***Index Terms*—Smartphone, Machine learning, Event detection, Hidden Markov models**

I. Introduction

With the development of small sensors, indoor event detection has become popular in the pervasive computing research community and small sensor device with a wireless communication function has been used to detect the indoor events. However, the main drawback of the ubiquitous sensor system is that these systems come with a large installation cost and maintenance cost. In addition, it requires to attach a separate sensor node to each door to detect door events.

Several researches have been done on using the sensors of unused smartphones (accelerometer and magnetometer) for door based event detection to achieve low cost home security systems. But the main drawback of those approaches are that the smartphone has to be mounted on the door or on the wall right next to the door (few feet away). Therefore, these methods provide very small sensing range making it impractical to detect the opening and closing events of an indoor environment where multiple doors are present.

In contrast to those approaches, we attempt to utilize active sound sensing which allows us to highly increase the sensing range. Active sound sensing requires to emit a sound by a speaker and then analyze the reflected waves. In this research we use the same smartphone to emit and to record the sound.

To detect events of indoor objects such as doors, we make use of a well known phenomenon known as "Doppler shift" which is the changing of the observed frequency of a wave when the observer is moving relative to the wave source. Even when we assume that the smartphone is both the wave source and the observer, the smartphone can observe the Doppler shift caused by indoor events occurred in a room where the smartphone is placed when the emitted sound wave is reflected by a moving object such as a door. Assume that the smartphone is inside a room emitting a sound signal with a constant frequency (sinusoidal wave) and recording the wave reflected from the surrounding at the same time. When a door event occurs during sensing, the sound waves bounce back from the door with a different frequency. This frequency shift is recorded as a distortion in the frequency spectrum of the recorded sound.

The characteristics of the frequency distortion is determined by the size of the door, speed and the direction of the door movement and smartphone's location relative to the door. This distortion can be analyzed to recognize the type of the door event and to differentiate between the doors when there are multiple doors present in the room.

Moreover, animals (including humans) that have two ears use them to get a sense of direction of a sound source. As most of the smartphones at present are equipped with multiple microphones, we try to increase the accuracy of our detections by employing two microphones and recording in stereo. To the best of our knowledge, there is no study that attempts to detect events of indoor objects using active sound sensing by a smartphone.

II. Related work

A. Monitoring door events using smartphones

Here we introduce a research that uses door or wall mounted sensors to monitor door events. Mahler et al. [3] propose a home security system based on smartphone sensors where the smartphone has to be mounted on the door itself or on the wall as close as 9 inches from the latching mechanism of the door. The authors measure the unique vibrations created by the door openings and closings using the accelerometer of a smartphone mounted near the door or capture the rotation of the door using the magnetometer of a smartphone mounted on the door. In contrast, our study does not require the smartphone to be mounted anywhere near the door and it promises a higher range of sensing.

B. Indoor context recognition with Wi-Fi

In here we introduce a research that employs Wi-Fi channel state information to detect the events and states of indoor everyday objects, such as doors and windows. Ohara et al. [4] propose a method to use a commodity Wi-Fi access point and a computer inside a room to detect the changes that occur in Wi-Fi signals propagation during indoor events.

C. Context recognition using Doppler effect

In this section, we introduce the researches that use Doppler effect for context recognition. Gupta et al. [2] propose a method of recognizing hand gestures using Doppler shift where they employ a tone in range of 18 kHz as a pilot tone. Fu et al. [1] propose a method of tracking exercises using a 20 kHz sound wave. Similarly, we use 20 kHz sound wave for the purpose of detecting events of indoor objects.

III. Proposed method

A. Theory of operation

The phenomenon that our study uses to detect door event is the Doppler shift as mentioned above. Since our source (speaker of the smartphone emitting a signal with constant frequency) and the observer (microphone of the smartphone) are both stationary, the Doppler shift is created by indoor objects such as doors. When a door rotates around its hinge and moves towards the smartphone, it creates positive shift causing an increment in the recorded frequency. Similarly, when the door is moving away from the smartphone creating a negative shift causing a drop in the recorded frequency. By utilizing this aspect, we can distinguish between door events such as openings and closings.

Fig. 1 shows a situation where the smartphone is placed in a room with two doors. We can see that when the Door1 is opened, there is a gradually diminishing positive velocity component of the door moving towards the smartphone until the angle between the door frame and the door ($\theta1$) becomes larger than the angle between the door frame and the line connecting the smartphone and the hinge ($\beta1$), making the observed frequency to distort to the positive direction in the frequency scale. Since then, until the door reaches the opened position, the velocity towards the smartphone remains negative so that the observed frequency distorts to the negative direction in the frequency scale. When the Door2 is opened, the velocity component of the door always stays in the direction of the smartphone until it reaches its opened position. This means that we can expect the distortion in the frequency to be positive all the way. By utilizing the above differences we can distinguish between the door events of the Door1 and Door2.

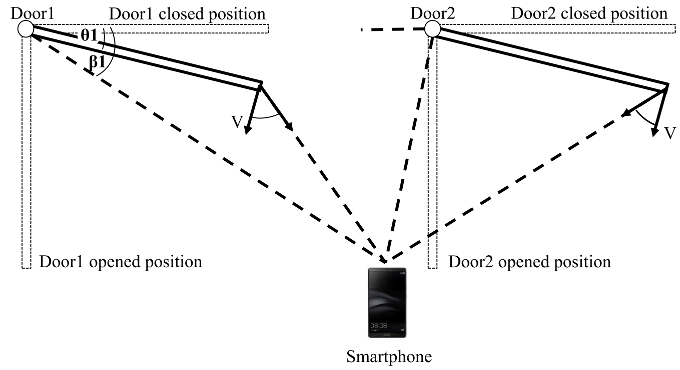


Fig. 1: Relationship between the location of the smartphone and the shape of the frequency distortion expected.

B. Sensor data

In our study, we used Nexus 6p smartphone for active sound sensing. The speaker of the smartphone emits a sinusoidal sound wave with a frequency of 20 kHz while two inbuilt microphones (one in the front and one in the back) are recording in a sample rate of 44.1 kHz. There are few benefits of using a frequency as high as 20 kHz in active sound sensing. One of them is that the emitted sound is inaudible to the human ear. Another point is that the frequency shift we can gain from movements is higher compared to the low frequencies. Finally, high frequency ranges are less affected by the low frequency noises.

We chose a meeting room as an experimental indoor environment for this preliminary study. We selected a locker door (Door1) and the main door (Door2) as our test objects. The rough sketch of the room is shown in the Fig. 2.

Here we introduce actual sensor data collected in our experimental environment illustrated in Fig. 3 and Fig. 4. Fig. 3 shows the difference in the FFT spectrogram when the distance between the door and the smartphone is increased gradually. Even when the smartphone was 5 m away from the door, we could observe the Doppler shift caused by the door.

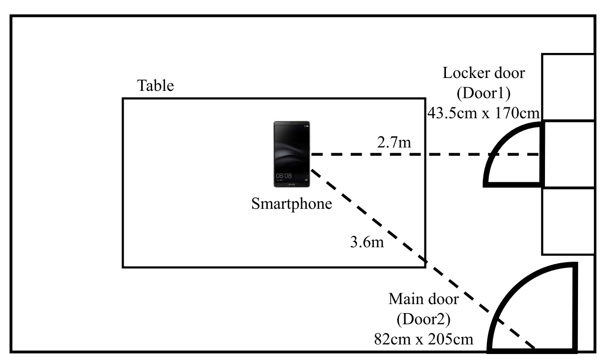
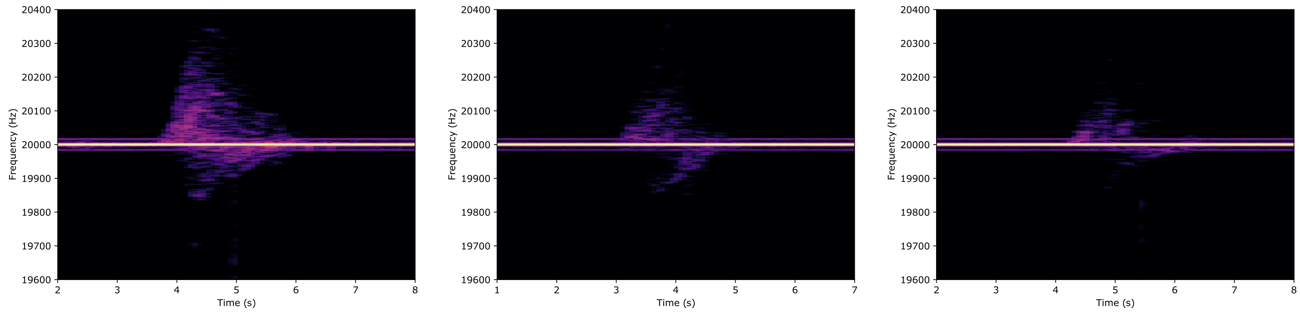


Fig. 2: Rough sketch of the testing environment.



(a) (b) (c)

Fig. 3: FFT spectrograms of open events of Door2 when the smartphone was placed (a) 1 m away from the Door2 (b) 3 m away from the Door2 (c) 5 m away from the Door2.

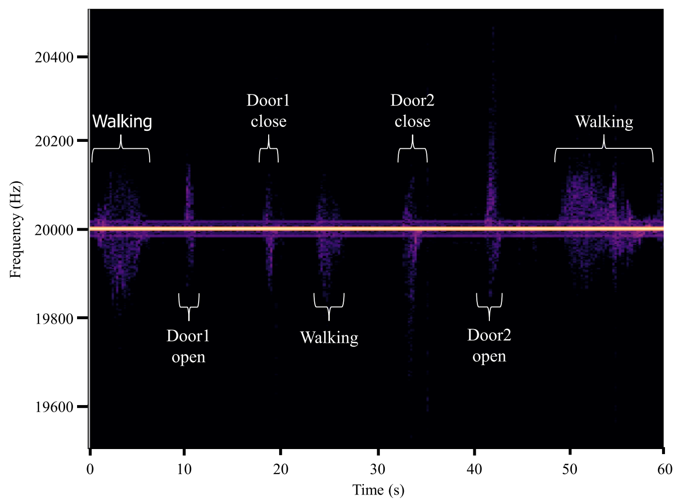


Fig. 4: FFT spectrogram of one data set (front channel) where the smartphone was placed 2.7 m away from the Door1 and 3.6 m away from the Door2.

As shown in Fig. 4, the open and close events of Door1 and Door2 are clearly distinguishable from one another.

C. Overview

1) Preprocessing: First, we calculate channelwise (stereo channels recorded from the microphone in the back and the microphone in the front) FFT components of the recorded sound, applying the Blackman window to 96% overlapping sliding windows of 22050 samples, corresponding to 0.02 sec. We then select and extract 3850 frequency bins from either sides of the bandwidth of the original signal, i.e. 20 kHz, that was emitted from the smartphone. This produces 7700 point feature vectors. Then, we reduce the dimensionality of these vectors using Principal Component Analysis (PCA). We found out that 100 dimensional vectors best fit for our purpose. Then, we normalize the data by applying

$$x_{normalized}(t) = \frac{x(t) - \bar{\mathbf{x}}}{\max(\mathbf{x}) - \min(\mathbf{x})} \tag{1}$$

2) HMM recognition: We yield our feature vectors by following the above procedure. Next, we classify those vectors into separate classes where each class represents a door event of a particular door or "Other" class. This is achieved by supervised learning approach where we label the feature vectors before the training phase. Each labels contains information about its class name as well as the start and the end times of the door event. Then we prepare a left-to-right hidden Markov model (HMM) for the each event where an output of each state is modeled with a mixture of Gaussians over observed values. We use an HMM with 10 states and 256 Gaussian mixtures for each state. We employ the Viterbi algorithm to determine the most probable state sequence of the HMMs allowing us to detect in to which class a specific feature vector at a given time is classified.

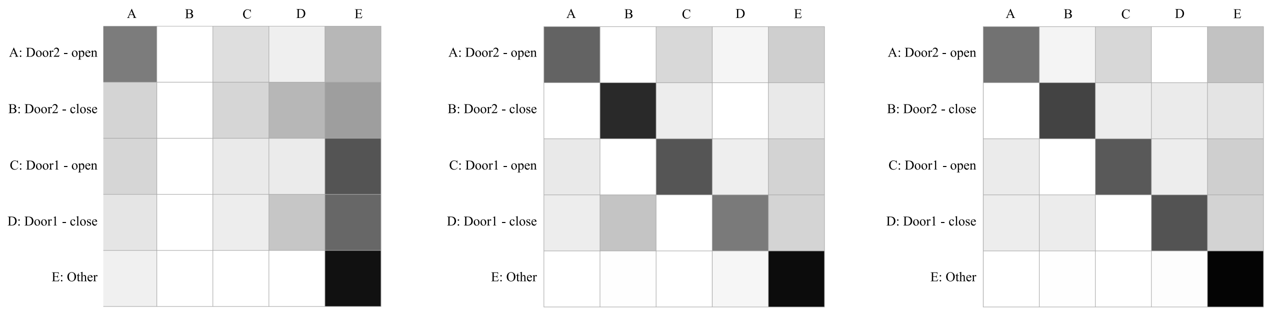
D. Evaluation

1) Data set: For this evaluation, we used data recorded from Nexus 6p smartphone in the environment illustrated in Fig. 2. We used both microphones located in the front and back for recording. The smartphone was placed vertically on the table and the locker door and the main door was opened and closed randomly such that each recording (session) is consisted of one door opening and closing from each door. We collected data from 10 such sessions and labeled the opening and closing events of the both doors. We also labeled the times where no events occur as "Other".

2) Evaluation methodology: We have two sets of feature vectors acquired from the recordings by the front microphone and the back microphone. First we perform leave-one-out cross-validation separately for each channel. Next, we use the data from both channels at the same time to perform leave-one-out cross-validation. We evaluate the performance of our approach by calculating F-measure using the precision and recall of the estimations.

TABLE I: Overall results of the proposed method.

	avg. precision	avg. recall	avg. F-measure	overall precision	overall recall	overall F-measure	error rate [%]
Back	0.327	0.353	0.335	0.806	0.806	0.806	19.419
Front	0.714	0.720	0.708	0.901	0.901	0.901	9.920
Both	0.773	0.721	0.740	0.921	0.921	0.921	7.853



(a) (b) (c)

Fig. 5: Confusion matrices of overall results when (a) only back channel is used (b) only the front channel is used (c) both channels are used.

3) Results: Table 1 shows the overall results of the proposed method. With our proposed method, we could achieve 80.58% when using the back channel, 90.08% when using the front channel and 92.15% when using both channels. The confusion matrices for overall results are shown in the Fig. 5.

IV. Conclusion

In this paper, we proposed a cost effective method of detecting events of the indoor objects using smartphone active sound sensing. With the smartphone, we emitted a sinusoidal sound wave with a frequency of 20 kHz and recorded the reflected sound using two inbuilt microphones at a sampling rate of 44.1 kHz. Then we extracted the feature vectors and used hidden Markov model to recognize the door events of a test environment. We used leave-one-out cross-validation method to evaluate the results and we were able to, not only successfully recognize door events of a particular door, but also distinguish between the door events of two doors with high precision, even when the smartphone is placed at a considerable distance from the doors. We plan to develop this method further by incorporating voice recognition techniques to distinguish between different conditions of an indoor environment.

Acknowledgment

This work is partially supported by JST CREST JPMJCR15E2, JSPS KAKENHI Grant Number JP16H06539 and JP17H04679.

References

[1] B. Fu, D. Vaithyalingam, A. Kuijper, F. Kirchbuchner, and A. Braun, "Exercise monitoring on consumer smart phones using ultrasonic sensing," 2017.

[2] S. Gupta, D. Morris, S. Patel, and D. Tan, "Soundwave: using the doppler effect to sense gestures," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2012, pp. 1911–1914.

[3] M. A. Mahler, Q. Li, and A. Li, "Securehouse: A home security system based on smartphone sensors," in *Pervasive Computing and Communications (PerCom), 2017 IEEE International Conference on*. IEEE, 2017, pp. 11–20.

[4] K. Ohara, T. Maekawa, and Y. Matsushita, "Detecting state changes of indoor everyday objects using wi-fi channel state information," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, p. 88, 2017.