Vol.2018-MBL-86 No.18
Vol.2018-UBI-57 No.18
2018/2/26

アクティブサウンドセンシングを用いた屋内日常物のイベント検知に関する検討

Thilina Dissanayake[1]　前川　卓也[1]　天方　大地[1]　原　隆浩[1]

Preliminary investigation of indoor event detection using active sound probing

Abstract: Event detection of indoor objects such as doors has a variety of application including intruder detection, HVAC control, surveillance of independently living elderly persons. Therefore, numerous researches have been done on this topic by the UbiComp research community. In this research, we propose a method to accurately detect events in indoor everyday objects such as doors and windows without using distributed ubiquitous sensors. In our approach, we perform event recognition conjoining the Doppler shift that occurs by the moving indoor objects and the acoustic characteristics of the environment acquired by the impulse response. Analysis of the time series data of the Doppler shift, using directional high frequency sine wave to acquire information of the direction of the occurred event, using stereo microphones to record the reflected wave can be stated as the specialties in this research. Also, by incorporating the knowledge about the state transitions of an object in to a recognition model, we will try to increase the accuracy of our recognitions.

Keywords: Indoor object, event detection, hidden Markov model, Doppler shift, impulse response

1. Introduction

Indoor context recognition is one of the most important research topics in the UbiComp research community because it is the basic technology of various applications such as lifelogging and context-aware applications. Especially, techniques to recognize events of indoor objects such as open/close events of doors and windows have a variety of applications such as intruder detection, HVAC control, surveillance of independently living elderly persons. So far, many of the UbiComp studies have employed distributed sensors to observe indoor events. In many cases, these studies employ sensors such as accelerometers, magnetometers, and state-change sensors (using, e.g., reed switches and magnets) to achieve indoor event recognition [1], [2], [3].

However, the main drawback of this approach is the large install and maintenance costs because this approach requires to attach a sensor node to each object. For example, frequent replacement of the node batteries or faulty sensor nodes places large burdens on users. When there is a large number of nodes installed in the house, the possibility of nodes getting faulty is high, hence requiring a maintenance personnel to frequently visit the home to fix them. Based on public database of IoT enabled houses, Kodeswaran et al. [4] revealed that houses equipped with 14-100 sensor nodes require a visit from a maintenance personnel every 18 days on average. Furthermore, attaching sensor nodes to indoor objects can decrease the aesthetic values of the artifacts [5]. Therefore, a reliable method for indoor event recognition that works well with a small number of devices is essential.

In our previous work, we have proposed a method of door based event detection using the Wi-Fi channel state information (CSI) [6]. However, it requires a specially modified Wi-Fi driver installed PC with a special Wi-Fi module to be set up in the environment of interest. Mahler

[1] 大阪大学大学院情報科学研究科

et al. [7] have proposed a smartphone based indoor event recognition system. At present, smartphones have become commodity devices and new models are being released every year or two. Therefore, retired smartphones have become abundant and are usually left at houses, unattended. We can employ these smartphones for indoor event recognition. In the method proposed by Mahler et al., the smartphone is either mounted on the door itself or to a wall near the latching mechanism of the door. When the smartphone is mounted on the door, magnetic sensor data from the magnetometer are used to detect the open/close events of the door. When the smartphone is mounted on the wall, 3-dimensional acceleration data from the accelerometer are used to detect impacts caused by the door open/close events. However, this method makes it hard to increase the sensing range as well as requires multiple smartphones to be mounted on/next to each door in an environment where more than one door is present.

In this research, we discuss an approach similar to Mahler et al. where we employ a commodity smartphone for door event recognition system. In contrast to their method, we employ active sound sensing which allows us to increase the sensing range of the event recognition. Therefore, we assume that we install a smartphone into one room to detect events of doors existing in the room. Indoor objects like doors have states like opened/closed states and events like open/close events where the state transitions occur. In our research, we emit an inaudible sine wave (18 kHz to 20 kHz) from the speaker of a smartphone and record the reflected wave from the same device. By this means, we capture the Doppler shift caused by the events of the indoor objects. In addition, we emit a sweep signal and from its impulse response we capture the acoustic characteristics of the environment which consist of the information related to the states of the objects in the surrounding [8]. The acoustic characteristics of the environment differs according to the opened/closed state of the doors. By employing the impulse response based on a sine sweep, we can also recognize states/events of a sliding door or window, which does not cause the Doppler shift.

To distinguish between the events of the different objects, our method makes use of the following information.

- Observed reflected sound that contains information of the Doppler shift changes over time, which differs with the relative location of the object with respect to the smartphone. (We explain it in Section 3 in detail.) Therefore, in order to model the event, we use a hidden Markov model (HMM) [9], which is used in modeling time-series data.
- The directionality of the sound waves emitted by a speaker increases with the frequency. Therefore, by emitting a composite sine wave composed of low frequency and high frequency components, we obtain the information related to the relative location of the object with respect to the smartphone. When there is a door in front of the smartphone, we can expect the Doppler shifts from the both sine waves (high frequency and low frequency) to be relatively similar in amplitude. On the other hand, when the door is situated elsewhere other than in front of the smartphone, the amplitude of the Doppler shift from the low frequency sine wave can be expected to be higher than that of the higher frequency sine wave.
- It is common for smartphones to be equipped with multiple microphones. Animals (including humans) that has two ears use them to get a sense of direction of a sound source. Similarly, we obtain information regarding indoor events using two microphones of the smartphone.

2. Related work

Many researchers make use of active sound sensing for context recognition. Gupta et al. [10] propose a method of recognizing hand gestures using the Doppler shift where they employ a tone in the range of 18kHz as a pilot tone. Fu et al. [11] propose a method of using a 20 kHz sound wave to track exercises using the Doppler shift caused by the movements of the body.

In several mobile computing studies [12], [13], sound beaconing has been used to estimate relative positions to other devices. Avtive sound sensing has also been used to locate a smartphone user. Rossi et al. [14] propose a method to locate a smartphone based on active sound fingerprinting. The authors measure impulse response at each indoor position and train a classifier that estimates a user's current position using the observed impulse response. Tung et al. [15] also make use of active sound probing for indoor location tagging. Our prior study [8] employs active sound sensing as well as passive sound sensing by smartphones to estimate location semantics such as toilet and restaurant.

Similar to the above approaches, we also employ a sound wave with a constant frequency to detect the events of indoor objects. Note that our method has several features, which are mentioned in the introduction section

such as making use of composite signals and a grammar that defines state transitions of an object, to achieve door event/state recognition.

3. Investigation of active sound sensing

Here we investigate the characteristics of signals obtained by active sound sensing, and based on the investigation we design the indoor event recognition method in the next section.

3.1 Theory of operation of event detection using Doppler shift

To detect the open/close events in an indoor objects such as doors, we make use of a well known and well understood phenomenon known as "Doppler effect" or "Doppler shift". This effect is defined as the shift of the observed frequency of a wave when the observer is moving relative to the wave source. In this research, both the observer and the wave source is the smartphone and the Doppler shift is caused by moving doors in the environment.

Imagine a scenario where the smartphone is set up inside a room and it emits a wave with a constant frequency, recording the reflected wave at the same time. When a door event occurs in the environment, it causes the reflected wave to have a different frequency than the original frequency that the smartphone has emitted. This frequency shift can be seen as a distortion in the FFT spectrum of the recorded sound.

We observed that the characteristics of this distortion changes according to the size of the door, speed and the direction of the door movement and the door's location relative to the smartphone. This distortion can then be analyzed to detect the door event and to differentiate between the door events of multiple doors. In detail, when the door rotates around it's hinge and move towards the smartphone, it causes an increment in the recorded frequency creating a positive shift in the FFT spectrogram. Similarly, when the door is moving away from the smartphone, it causes a decrement in the recorded frequency creating a negative shift in the FFT spectrogram. By utilizing this characteristic shifting of frequencies, we can distinguish between the open/close events of the door. Next we will look at how can we differentiate between door events of multiple doors. Fig. 1 shows a situation where the smartphone is placed in a room with two doors. When we look at a open event in Door1, we can see that there is a gradually diminishing positive velocity component from the door towards the smartphone since the door started

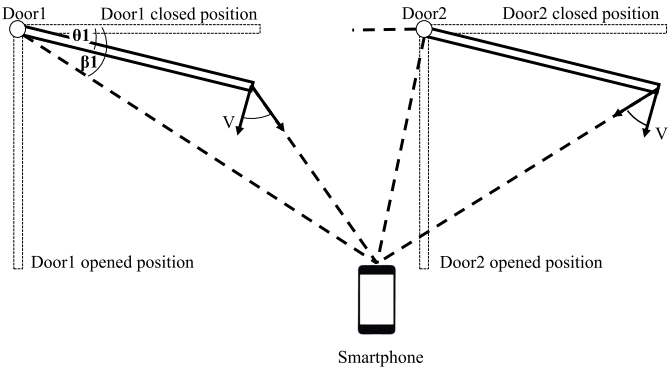

図 1 Relationship between the location of the smartphone and the characteristics of expected frequency shift

moving from the closed position until the angle between the door frame and the door (θ 1) becomes larger than the angle between door frame and a line connecting the smartphone and the hinge (β 1), resulting an increment in the observed frequency. Since then until the door reaches it' s opened position, the velocity component from the door towards the smartphone stays negative, resulting a decrement in the observed frequency.

When the Door2 is opened, the velocity component from the door towards the smartphone stays positive all the way from the closed position to the opened position. This means that we can expect only an increment in the observed frequency. By utilizing this characteristic differences in the frequency shifts, we can distinguish between the door events of Door1 and Door2.

3.2 Impact of distance between smartphone and object

Open/close events of doors can be detected utilizing the Doppler shift as shown above. We first investigate the effect of the distance between the smartphone and the door on our proposed method. We recorded acoustic signals, placing the smartphone at various distances from the door. We used Google Nexus 6P smartphone for active sound sensing. The smartphone emitted a sinusoidal sound wave with a frequency of 20 kHz from its speaker while two inbuilt speakers (one in the front and one in the back of the smartphone) recorded stereo sound at a sample rate of 44.1 kHz. Figure **??** shows visualized FFT power spectrums of the recorded sound signal that records an open event when the phone was placed 1 m, 3 m, 5 m away from the door. From this spectrogram, we can confirm that the frequency shift has mostly occurred towards the positive direction in an open event of the door as well

as the Doppler shift occurred by the door event can be observed by the smartphone even if the door is 5 m away from it.

3.3 Time variance of Doppler shift

Figure 3 shows a floor plan of our primary environment of interest and Figure 4 shows the FFT power spectrum obtained when open/close events of Door1 and Door2 occurred in the same environment. During an open event of Door2, the door first moves towards smartphone and then it moves away from the smartphone. In addition, the door moved relatively long time towards the smartphone than away from it. Therefore, we can observe that the frequency shift from an open event creates a strong shift towards the positive direction (higher than 20 kHz) first and then creates a relatively weak shift towards the negative direction (lower than 20 kHz) of the FFT power spectrum. Moreover, during a close event of Door2, it moves towards the smartphone for a short time after it moves away from the smartphone for a longer time. This creates a frequency shift opposite to that of an open event creating a weak shift towards the negative side after creating a strong shift to the positive side of the FFT power spectrum. In contrast, during an open event of Door1, we observed a frequency shift mainly consisting of a strong shift towards the positive side. As above, because the observed frequency shift depend on the positions of doors, it is necessary to model the time variance of the Doppler shift for recognizing open/close events.

3.4 Stereo recording using two inbuilt microphones

Fig. 5 shows the diagrams of the situations where a door is located in front of the smartphone, left to the smartphone and behind the smartphone. Fig. 6 shows the FFT power spectrograms of an open event of the door when the data was taken from the front and back microphones. By this, we can determine that the characteristics of the acoustic signals obtained by the two microphones are different because of the difference of the microphone' s location on the smartphone.

3.5 Composite signal from two sinusoidal waves with different frequencies

The characteristics of the sound wave emitted from the speaker differs according to the frequency. As example, lower frequencies tends to have a higher amplitude than that of the higher frequencies because of the hardware configurations of the smartphone. Also, the sound waves with higher frequencies appears to be more directional than that of the sound waves with lower frequencies. Fig. 7 shows the spectrograms of the audio data of obtained from the front microphones when the composite wave is composed of 18 kHz and 20 kHz sinusoidal waves, when the smartphone is placed in three different ways as shown in the Fig. 5. By this, we can obtain the information about the relative location of an object with respective to the smartphone by employing a composite sinusoidal wave composed with the superposition of two sinusoidal waves with two frequencies.

4. Method

4.1 Overview

In our proposed method, a smartphone emits a composite sinusoidal wave and also emits a sine sweep occasionally. Next, we use the acoustic signals recorded from the front and back inbuilt microphones to recognize the events/states of the objects in the environment of interest. Figure TBA shows an overview of our recognition method. After extracting features, a feature vector sequence related to Doppler shift are fed into a discriminative classifier that recognize the events of each object in the environment. In addition, a feature vector sequence related to the impulse response are fed into a discriminative classifier that recognizes states of each object. After that, the class probabilities from the discriminative classifiers are concatenated and fed into HMMs tailored to each object.

4.2 Feature extraction

We extract features related to the Doppler shift and impulse response form the acoustic signals obtained by both microphones.

4.2.1 Features related to Doppler shift

We extract features for event detection using the FFT power components around the 18 kHz and 20 kHz frequencies. In our proposed method, we implement a Blackman window with size 0.5 seconds to 96% overlapping sliding windows to calculate the FFT power spectrum of the audio data. We then extract 4000 frequency bins from either sides of the bandwidths of 18 kHz and 20 kHz sine waves. This produces an 8000 dimensional frequency vector sequence. After that, we reduce the dimensionality of the vectors by using time frame wise averaging. Here, we use a 100 dimensional window to a 50% overlapping sliding window along the frequency axis of each data frame tak-

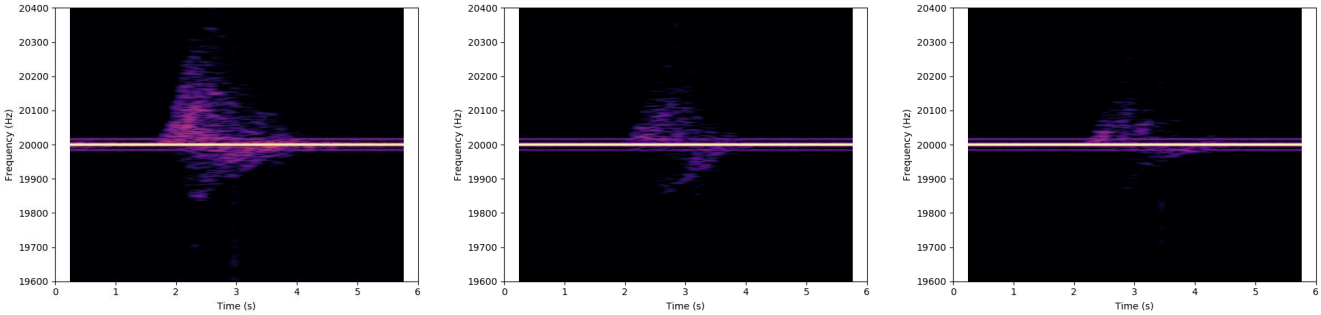

図 2 Smartphone was palced 1 m, 3 m, 5 m away from the door.

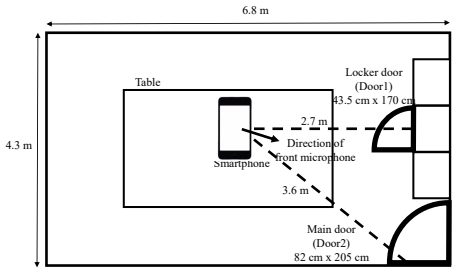

図 3 Environment of preliminary experiment (1)

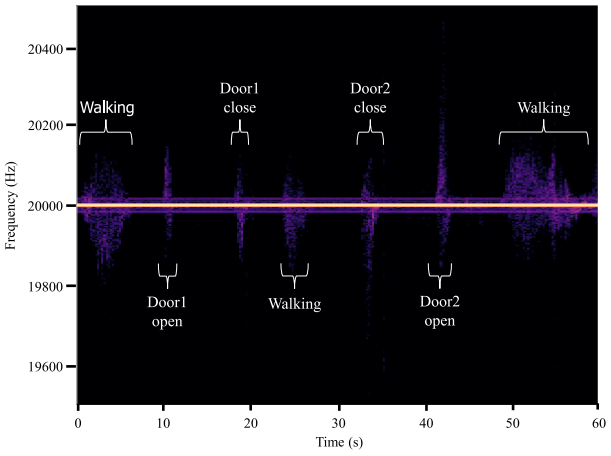

図 4 FFT spectrogram of front channel where the smartphone was placed 2.7 m away from the Door1 and 3.6 m away from the Door2 as shown in Figure 3

ing the average of the FFT power components inside the window as one dimension. With this method, we can reduce the dimensionality of the vectors from 8000 to 160 dimensions while still retaining most of the characteristics of the original feature vectors.

4.2.2 Features related to impulse response

We obtain acoustic characteristics of the environment by analyzing the impulse responses of the emitted sine sweeps as mentioned above. First, we calculate the FFT power spectrum of the of the audio data applying a 0.01 seconds Blackman window to a 85% overlapping sliding window. This very short window size can be justified by the fact that the sweep length is 0.05 seconds and the information about the state of the object exists in the reflected waves which last for a very short period of time. We use FFT power spectrum components corresponding to the frequency range of the sweep as features to recognize the states of the objects. Cowling et al. [16] conclude that Mel-Frequency Cepstral Coefficient (MFCC) based feature extraction technique is one of the most appropriate approaches to environmental sound recognition system achieving 70% recognition. Chen et al. [17] use MFCC features to recognize bathroom activities such as showering, flushing, and urinating with high accuracy. Similarly, we calculate 12 order MFCC features from the extracted FFT power spectrum components.

4.3 Discriminative classifier for events

For each time slice, a feature vector consisting of features related to Doppler shift is fed into a discriminative classifier for events. The discriminative classifier is a random forest [18] that is used to compute class probabilities of open/close event classes of each object in the environment of interest and a class belonging to neither of above classes. Therefore, the decision tree is $(2N+1)$-class classifier, where N is the number of door objects in the environment. With this classifier, we can separate input signals where information about events/states of multiple doors are mixed together into a time-series of class probabilities for each class. Here our method assumes that the smartphone periodically emits sine sweeps. Since the amplitudes of the sweeps are much larger those of frequency shifts caused by the Doppler shift, the small frequency shifts might be ignored when the observed signals (or features extracted from the signals that include both the frequency shifts and sine sweeps) are directly fed into a generative model such as an HMM that learns a distri-

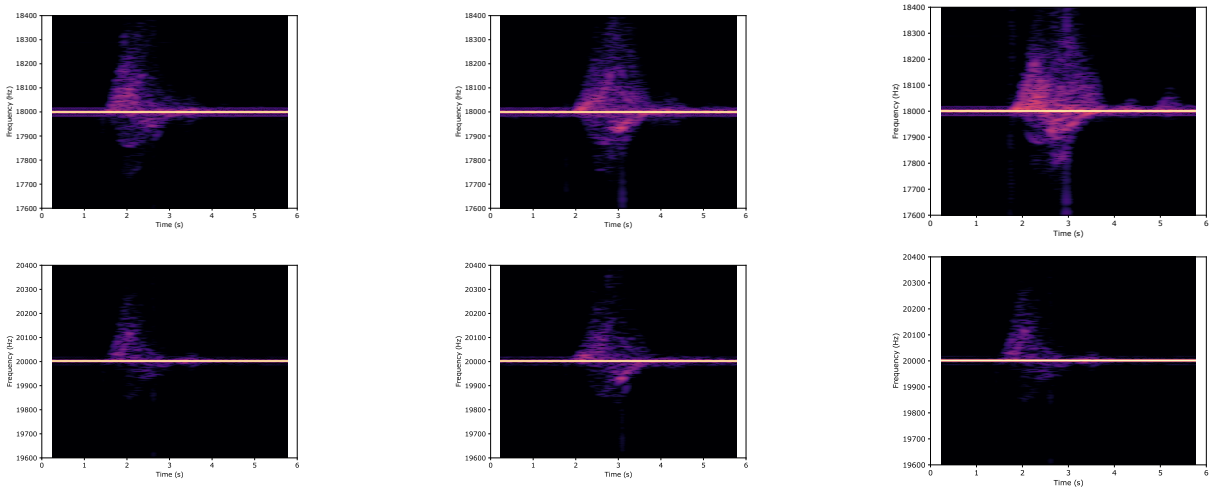

図 5 Spectrograms of a door opening where the door is situated (a) in front of the smartphone, (b) left side of the smartphone, (c) behind the smartphone. Spectrograms of the audio data from the front microphone is in the top row and the back microphone is in the bottom row.

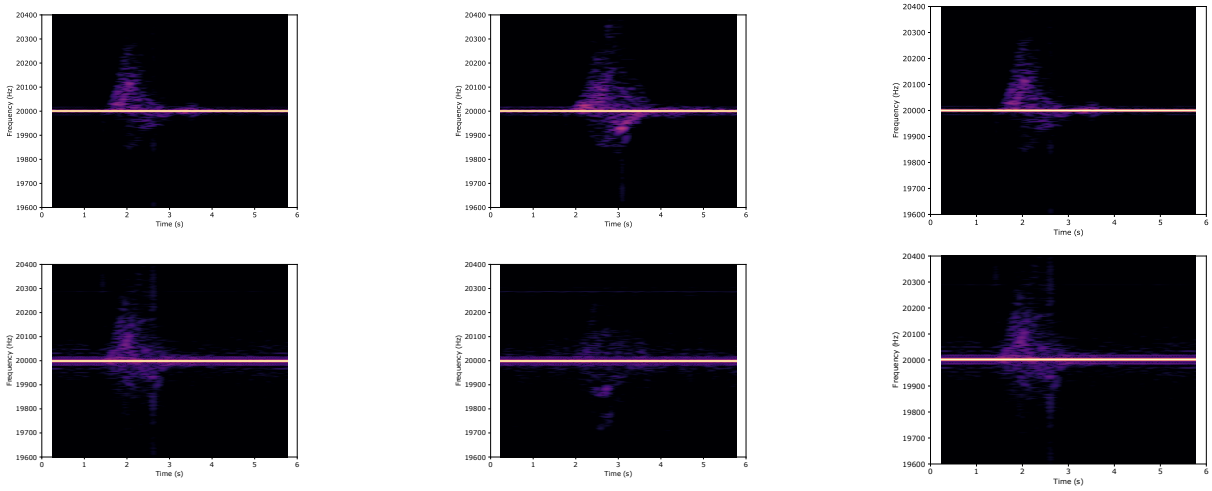

図 6 Door is situated (a) in front of the samartphone, (b) left side of the smartphone, (c) behind the smartphone.

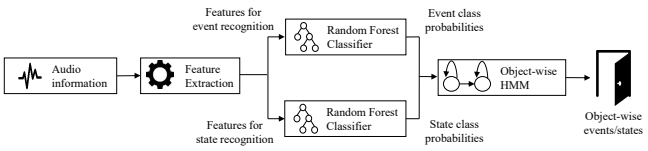

図 7 Relationship between the location of the smartphone and the characteristics of expected frequency shift

bution of input signals. In contrast, since a decision tree, which is used in our method, performs classification based on thresholds, it is not affected by scales of input features.

4.4 Discriminative classifier for states

We prepare a discriminative classifier tailored to each door object that recognizes states of the objects using information obtained by sine sweeps. For each time slice, a feature vector consisting of features related to impulse response is fed into the discriminative classifier (random forest). As mentioned in the introduction section, the classifier is trained as a five-class classifier. Here we explain how to prepare a label for a feature vector at time t for the classifier. The feature vector is labeled as "open@sweep" when the door was opened and a sine sweep was emitted at time t. When the door was closed and a sine sweep was emitted at time t, the feature vector is labeled as "close@sweep." In contrast, when the door was opened and a sine sweep was not emitted at time t, the feature vector is just labeled as "open." When the door was closed and a sine sweep was not emitted at time t, the feature vector is just labeled as "close." Otherwise, the feature

vector is just labeled as “others.” The outputs of the classifier are a class probability for each class for each time slice.

4.5 Object-wise HMM

We prepare a set of HMMs tailored to each object. The set of HMM consists of a left-to-right HMM prepared for each event/state of the object, i.e., “open,” “close,” “opened,” and “closed.” A sequence of class probability vectors obtained by the discriminative classifier for events and a sequence of class probability vectors obtained by a discriminative classifier for states tailored for the object are concatenated and then fed into the HMMs. Therefore, the values of the observed variables of the HMMs correspond to the concatenated probability vectors, and we represent its output distributions using Gaussian mixture densities (Gaussian mixture model: GMM). The hidden state of the HMM corresponds to the internal state of the event. In the model, the hidden state at time t depends only on the previous hidden state at time $t-1$, i.e., a Markov process. In addition, the observed variable at time t depends only on the hidden state at time t. Therefore, using the HMMs enables us to capture the temporal regularity of events. The Baum-Welch algorithm [19] is used to estimate the HMM parameters. When we decode test data (probability vector sequence) using the trained HMM set, we use the Viterbi algorithm to find the most probable state sequence in/across the HMMs [19]. With this information, we can know into which HMM (event/state) a probability vector at time t is classified.

4.6 Decoding with grammar

Left-to-right HMM is prepared for each event/state. This makes possible to employ the Viterbi algorithm to determine the state transitions across the HMMs where state transitions happen from the last state of a HMM modeled after a particular door to the first state of another HMM modeled after the same door. In reality, this corresponds to a door transitioning from the “closed” state to the “opened” state due to an “open” event. By taking above state transition into consideration, we can represent the state transitions of the doors using HMMs. Here we can determine the state transitions among the HMMs employing a hand crafted grammar where we train the model with the prior knowledge about the connection between the states and the events of the object of interest. As an example, we can specify that an “close” event can only occur when the door is in “opened” state.

We provide a grammar example written in extended BNF. The grammar for a door is described as follows.

Grammar for door

[opened close] {closed open opened close} closed [open opened]

5. Evaluation

5.1 Data set

We collected sensor data in real two environments. Figure **??** shows our experimental environments and their settings. We installed Google Nexus 6 smartphone as shown in the figure. We recorded 44.1 kHz stereo audio using the front and back microphones of the smartphone. The smartphone emitted sine sound waves and sine sweeps during the experiment as described in the proposed method section. The smartphone also recorded time stamps when it emitted sine sweeps.

Environment 1 is a storage room with two doors and a cabinet. There are several discarded desktop PCs and shelves in the room. Environment 2 is a meeting room with one door. The distance between the door and the smartphone is approximately 5 meters. In this room, we verify the performance of our method when the distance between a door and the smartphone is long. There are four tables and 11 chairs in the room.

In each environment, a participant conducted 8 sessions of data collection. To obtain ground truth, we recorded the sessions with a web camera.

Each object has two events and two states, i.e., “open” and “close” events and “opened” and “closed” states. Throughout a session, the participant used all objects so that each event of the objects occurred twice in an arbitrary order. That is, in a session, for example, a door can be opened while a window can be closed. In another session, the door can be closed while the window is opened. As above, each object is used under different conditions in our experiment. Because the participant walked at random in the room and used the objects in a random sequence, the objects such as doors were used from different sides and in different positions. The average duration of a session was approximately 300 seconds.

5.2 Evaluation methodology

We evaluated the performance of our method for each environment with leave-one-session-out cross validation. We prepared the following methods to investigate the

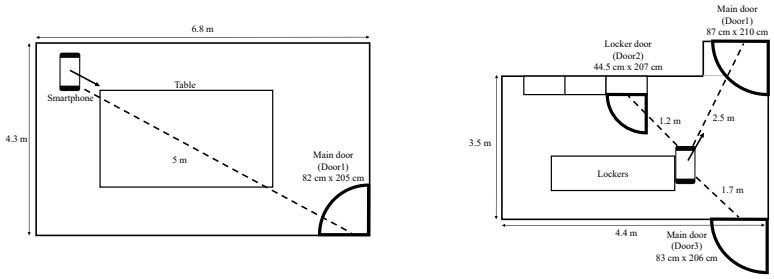

図 8 Door is situated (a) in front of the samartphone, (b) left side of the smartphone, (c) behind the smartphone.

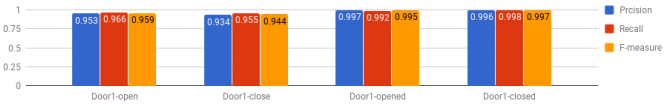

図 9 Relationship between the location of the smartphone and the characteristics of expected frequency shift

effectiveness of the composite sine waves, two microphones, our two-tier architecture with discriminative classifiers, and handcrafted grammars. Recognition accuracy for each of the above methods is evaluated using the macro-averaged precision, recall, and F-measure, calculated based on the recognition results per window of data.

5.3 Results

Fig. 12 and Fig. 13 show the detailed results of the events/states recognition of Door1 in Environment 1. Fig. 13 shows the confusion matrix.

Fig. 14 shows the detailed results of the events/states recognition of each door in Environment 2. Fig. 15 shows the confusion matrices.

6. Limitation

As we use active sound sensing to detect events and recognize states in doors, the sensing range limits to the room where the smartphone was placed. But, in contrast to the camera based event detection methods, our approach has a much higher sensing ranged and has no effect from the viewing angle of the smartphone. In compared to the event recognition methods based on Wi-Fi CSI, our method can be considered to have similar sensing range. Even though Wi-Fi signals can penetrate walls and reach objects in other rooms as well, if the event detection is based on the differences in the propagation path due to open/close events, it is difficult to detect door events in the other rooms.

謝辞 本研究の一部は JST CREST JPMJCR15E2, JSPS 科研費 JP16H06539，JP17H04679 の助成を受けて行われたものです．

参考文献

[1] Van Kasteren, T., Noulas, A., Englebienne, G. and Kröse, B.: Accurate activity recognition in a home setting, *the 10th International Conference on Ubiquitous Computing (Ubicomp 2008)*, pp. 1–9 (2008).

[2] Philipose, M., Fishkin, K. P., Perkowitz, M., Patterson, D. J., Fox, D., Kautz, H. and Hähnel, D.: Inferring activities from interactions with objects, *IEEE Pervasive Computing*, Vol. 3, No. 4, pp. 50–57 (2004).

[3] Tapia, E. M., Intille, S. S. and Larson, K.: Activity recognition in the home using simple and ubiquitous sensors, *International Conference on Pervasive Computing (Pervasive 2004)*, pp. 158–175 (2004).

[4] Kodeswaran, P. A., Kokku, R., Sen, S. and Srivatsa, M.: Idea: A system for efficient failure management in smart IoT environments, *the 14th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys 2016)*, pp. 43–56 (2016).

[5] Beckmann, C., Consolvo, S. and LaMarca, A.: Some assembly required: Supporting end-user sensor installation in domestic ubiquitous computing environments, *International Conference on Ubiquitous Computing (UbiComp 2004)*, pp. 107–124 (2004).

[6] Ohara, K., Maekawa, T. and Matsushita, Y.: Detecting State Changes of Indoor Everyday Objects using Wi-Fi Channel State Information, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, Vol. 1, No. 3, p. 88 (2017).

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

[8] Tachikawa, M., Maekawa, T. and Matsushita, Y.: Predicting location semantics combining active and passive sensing with environment-independent classifier, *UbiComp 2016*, pp. 220–231 (2016).

[9] Welch, L. R.: Hidden Markov models and the Baum-Welch algorithm, *IEEE Information Theory Society Newsletter*, Vol. 53, No. 4, pp. 10–13 (2003).

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

[11] Fu, B., Vaithyalingam, D., Kuijper, A., Kirchbuchner, F. and Braun, A.: Exercise Monitoring On Consumer Smart Phones Using Ultrasonic Sensing, *4th international Workshop on Sensor-based Activity Recognition and Interaction (iWOAR 17)* (2017).

[12] Zhang, Z., Chu, D., Chen, X. and Moscibroda, T.: Swordfight: Enabling a new class of phone-to-phone ac-

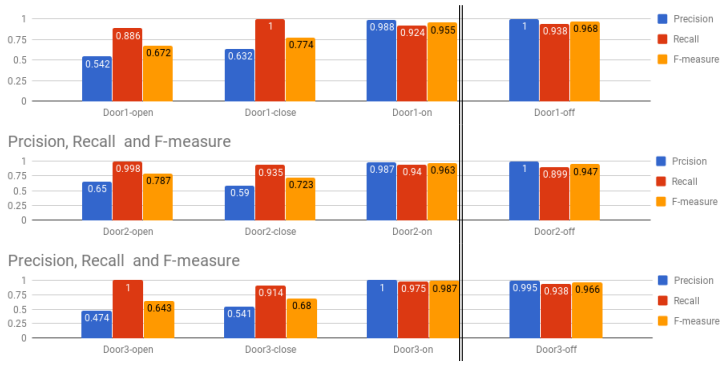

図 10 Door is situated (a) in front of the samartphone, (b) left side of the smartphone, (c) behind the smartphone.

tion games on commodity phones, *MobiSys 2012*, pp. 1–14 (2012).

[13] Constandache, I., Bao, X., Azizyan, M. and Choudhury, R. R.: Did you see Bob?: human localization using mobile phones, *MobiCom 2010*, pp. 149–160 (2010).

[14] Rossi, M., Seiter, J., Amft, O., Buchmeier, S. and Tröster, G.: RoomSense: an indoor positioning system for smartphones using active sound probing, *the 4th Augmented Human International Conference*, pp. 89–95 (2013).

[15] Tung, Y.-C. and Shin, K. G.: EchoTag: accurate infrastructure-free indoor location tagging with smartphones, *MobiCom 2015*, pp. 525–536 (2015).

[16] Cowling, M.: Non-speech environmental sound recognition system for autonomous surveillance, PhD Thesis, Griffith University (2004).

[17] Chen, J., Kam, A. H., Zhang, J., Liu, N. and Shue, L.: Bathroom activity monitoring based on sound, *Pervasive 2005*, pp. 47–61 (2005).

[18] Breiman, L.: Random forests, *Machine learning*, Vol. 45, No. 1, pp. 5–32 (2001).

[19] Rabiner, L. R.: A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE*, Vol. 77, No. 2, pp. 257–286 (1989).