Instant Learning Sound Sensor for Ubiquitous Computing

Yuya Negishi
Graduate School of Information Science
Nagoya University
1, Furo-Cho, Chikusa-ku, Nagoya, JAPAN
negishi @ el.itc.nagoya-u.ac.jp

Nobuo Kawaguchi
Graduate School of Engineering
Nagoya University
1, Furo-Cho, Chikusa-ku, Nagoya, JAPAN
kawaguti @ nagoya-u.jp

Abstract

We propose a smart sound sensor for ubiquitous computing that can learn and detect events from various sensor information with a time series signal. Using the smart sensor, a developer of a context-aware system and ubiquitous service can easily utilize a real world sound as a context information without a signal processing programming. The signal processing method can be extended for the other sensors such as accelerometer and directional sensor, pressure sensor, etc. By the experiment, we evaluate the feasibility of the instant learning sound sensor.

1. Introduction

Context aware systems are beginning to play an important role to support human activities in the real world. As the way to obtain the context information into the system, a lot of input devices such as GPS, acceleration sensor, pressure sensor, and temperature sensor are used. For example, small network devices with these sensors are developed, and used to build a smart room [1][2]. The health care system that recognizes a user's action that washing hand and having a shower, etc, in the bathroom with a microphone is proposed by Jianfeng Chen [3]. There are also some researches and systems using signal processing with a time series data from sensors to obtain context information. However, the design of the recognition algorithm of a complex pattern is not easy for the prototyping of the system because an analysis of a feature quantity requires a lot of time.

In this paper, we propose an instant learning sensor which can learn signal pattern instantly on the site. The learning sensor can record and learn the time series signal pattern that a user wants to detect as an event. Instant learning sensor has following three features.

(1)Instant Learning : In the sensor installation site, the sensor can instantly learn an signal pattern by demonstration of a target event. The system automatically analyzes the recorded signal, and chooses most appropriate algorithm to extract a feature quantity. Therefore by using the instant learning sensor, user can build context-aware systems without a signal processing programming.

(2)Smart Sensor : Sensor can process the signal by itself. It is possible to cooperate and distribute with other devices easily.

(3)Simple Device : It is possible to implement by cheap devices such a microcomputers. So, we aim at light processing as much as possible.

We think the instant learning sensor can support easy building of an advanced ubiquitous service. For example, in systems like eBlocks [4], that can easily construct smart environment by simply connecting block devices, instant learning sensors can be used as event triggers.

In the following, we focus the first feature and describe prototype of the instant learning sensor, which can recognize sounds of human's action, example for object operations. Sound has very rich context information. We designed and implemented an instant learning sound sensor using a cheap piezoelectric device as a microphone.

2. Design of Instant learning Sound Sensor

2.1. System overview

Instant learning sound sensor has two modes (Fig. 1).

(1)Event Learning Mode : Analysis of a target sound that is inputted by user's demonstration.

(2)Event Detection Mode : Monitoring a target sound and notifying if event is detected.

2.2. System architecture

We chose a piezoelectric device as a microphone device, because it is a very cheap, small and thin device, so it also possible to be attached to various objects anywhere, for example wall, desk, cup, faucet, book, telephone, trash box,

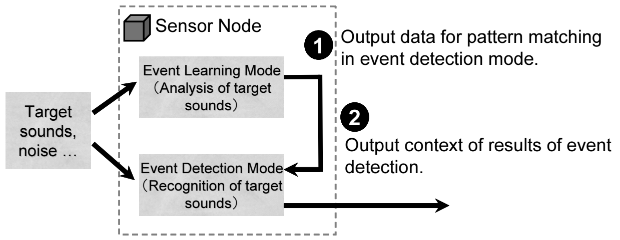

Figure 1. System Overview

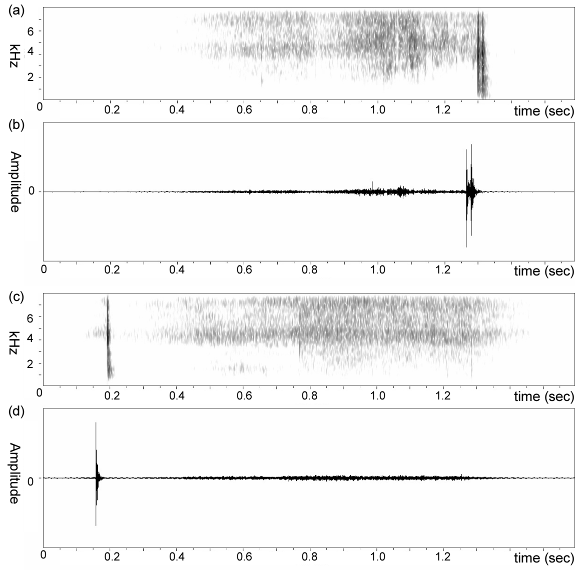

Figure 2. (a,b)Spectrogram and Waveform of opening a drawer, (c,d)closing.

and so on. Generally, it is said that sound sensing using a microphone has a susceptibility to environmental noise and human's speech voice, but we are confirmed a piezoelectric device attached to a object has a robustness to those noise. In the prototype system, we connect that device to a general PC's mic-in.

2.3. Algorithm

2.3.1 Signal Processing

In this section, we describe a process of event learning mode and detection mode in Fig. 1.

In a lot of speech recognition system, they use speech spectrum such as mell-cepstrum or liner prediction coefficients (LPC) to extract feature quantities. However, it is an unknown whether these feature quantities can be also applied to the instant learning sound sensor. So, we design a flexible configuration of the feature quantities to select a proper set for the target sound.

To discuss about the signal processing, we use the case study. Fig. 2 shows an example of target sounds. In Fig. 2, (a) and (b) show a spectrogram and waveform in opening a drawer out of the cabinet. (b) and (c) show in the opposite move of closing drawer. Comparing these two sounds, there is no distinguished difference about a spectrogram, but amplitude change patterns are obviously different. So, we think a feature quantity should include not only frequency characteristic but also amplitude characteristic.

In the prototyping system, we use frequency and amplitude characteristic by the following procedure. In one frame (512points), it calculates power spectrum using fast fourier transform (FFT), and divides it equally to 16 sections, and calculates the average of each section, and normalize by max power. The range of each element is $0.0 \sim 1.0$.

In addition to these 16 elements, 17th element value is added as an amount of the change of power by the following expression. L is constant number. This expression calculates an amount of change of the maximum value of power spectrum of current frame and previous frame, and, the result value is made to the quantum in 5 stages (-A2, -A1, A0, + A1, +A2). The range of this element is $0.0 \sim 1.0$, too. By some target sounds, it is possible to weight in either frequency or amplitude characteristic.

$$Vec_{17} = \frac{Power_{max}(n) - Power_{max}(n-1)}{L} \quad (1)$$

Input signal is mono 16 bit, sampling frequency is 16 kHz, window length is 512 points, shift length is 120 points, and we use blackman window function.

2.3.2 Instant learning

In event learning mode, it outputs a code book used at vector quantization and a code pattern dictionary used at dynamic time matching (DP matching, DTW) of event detection mode, and makes threshold determination for DP matching. When a developer installs the sensor, it is only three steps to do, shown in Fig. 3.

In first step, on piezoelectric attached site, the system records stationary sounds for a while, and makes a code book of feature quantity vectors of these sounds using LBG algorithm. In the prototype system, the code book size is 4.

In second step, it records a target sound by a developer's event demonstration and extracts non-stationary frames where the quantization error is higher than threshold at vector quantization with first step code book. Next, it makes code book of the target sound from the extracted frames. In the prototype system, this code book size is 16. Finally, it outputs a code pattern dictionary of target using the two code books.

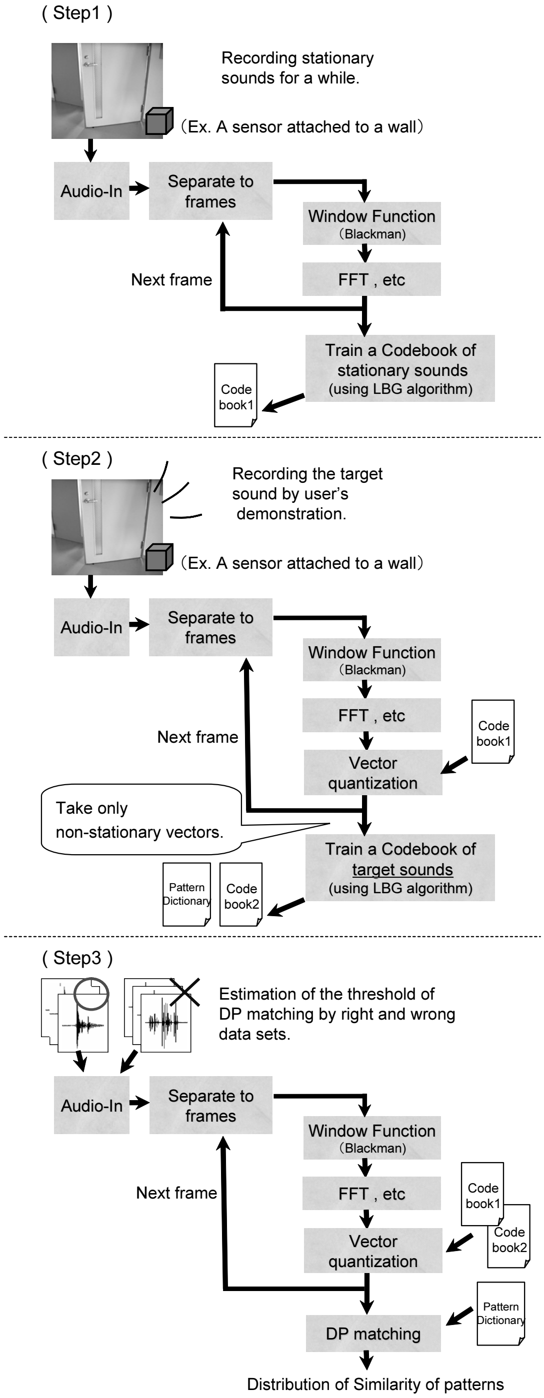

Figure 3. Process of Instant Learning

In third step, as test, a developer input correct or incorrect event sounds with annotations. By this operations, the system gets the set of minimum cost in DP matching, and can make threshold determination for an event detection judgment. The threshold is determinated by following two expressions. $P(S|n)$ is a rate of true detection when correct sounds inputted, and $P(N|s)$ is a rate of false detection when incorrect sounds inputted. The relation between two expression and threshold is shown in Fig. 4. By Fig. 4, the best threshold is around point where $P(S|n) = P(N|s)$.

$$P(S|n) = \frac{(Number\ of\ accepted\ incorrect\ sounds)}{(Total\ of\ incorrect\ sounds)} \tag{2}$$

$$P(N|s) = \frac{(Number\ of\ rejected\ correct\ sounds)}{(Total\ of\ correct\ sounds)} \tag{3}$$

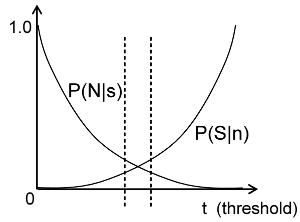

Figure 4. Relation between threshold and P(S|n), P(N|s)

2.3.3 Event detection

In event detection mode, using the code book and code pattern dictionary made in learning mode, the sensor recognizes the target sound and notifies when detected. The processing flow is extraction of feature quantities from each frame, vector quantization, and comparing the threshold and a matching cost. These are shown in Fig. 5.

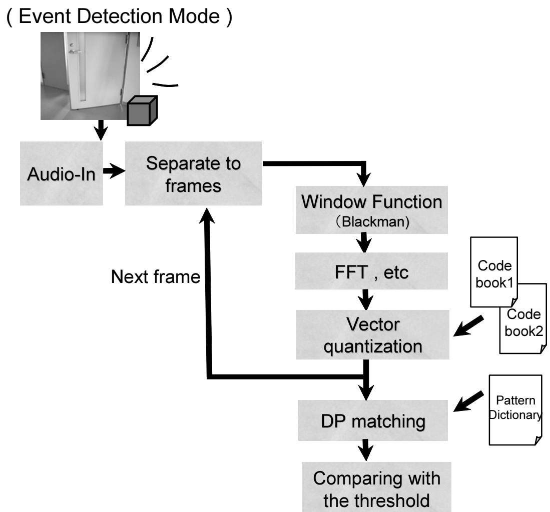

Figure 5. Process of Sound Recognition

3. Implementation

Based on the above design, we implemented a prototype of the instant learning sound sensor using C++. Fig. 6 is a screenshot of the sensor installation in instant learning mode, and a photo of an attached piezoelectric device on aspect of a drawer. The right side of Fig. 6 is a simple example of remote monitoring system using the proposed sensor, which sends a mail when a target event is detected. Like this example, by only using sound sensors, we can easily make sensors to monitor events.

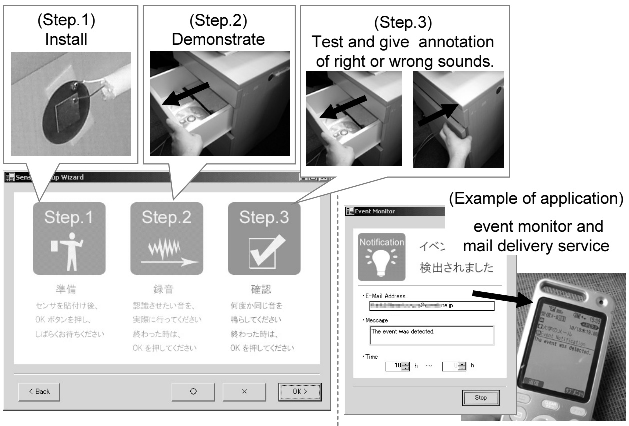

Figure 6. Screenshot of Sensor Installer and Example of use

Other examples we confirmed are sounds of pouring a tea into cup, typing a keyboard, etc.

4. Evaluation

We evaluated the recognition accuracy of our proposal sound sensor, using an example sound of sliding the drawer out in the section 2.

(1)First, we provided five correct data and five incorrect data to sensor. Incorrect data are sounds of closing actions.

(2)In the event detection mode, we inputted sounds of (a)correct event 21 times, (b)incorrect event that is closing actions 21times and (c) knock, etc, 20 times.

The results are shown in Table. 1 and 2. Table. 1 shows matching cost of input sounds with a target pattern. Matching cost below 9.49 results correct recognition. So, the best threshold is around 9.49. In this experiment, we leave a some margin for error, then sensor installer estimated the threshold is 9.69. Table. 2 shows the recognition accuracy of data (a), and the rate of false detection of data (b)(c). From these results, the proposal sensor is not always able to detect a target sound, but there are not false detections. We think that a low rate of false detection is more important than high recognition accuracy of correct events. Therefore these are enough practical results.

However, the feature quantity used to this experiment is not necessarily best for other sounds. Therefore, to support various type sounds, the sensor has to support more feature quantities and recognition algorithms. To reduce recognition errors, we can utilize outputs of multiple sound sensors.

Table 1. Matching Cost

	Matching cost
Correct data	$2.68 \sim 9.49$
Incorrect data	$11.3 \sim 13.6$

Table 2. Result of Evaluation

	Rate
(a)Recognition accuracy	76.2%
(b)Rate of false detection 1	0%
(c)Rate of false detection 2	0%

5. Conclusion

This paper describes a design and implementation of the instant learning sound sensor for ubiquitous computing. The recognition algorithm in this prototype system is consist of FFT and one stage DP matching for each frame. So, we think that the proposal sensor can be made as a small sensor device with a microcomputer. We will support more kinds of feature quantities and algorithms for various sounds recognition.

References

[1] Crossbow Technology, MOTE, http://www.xbow.com /Products/Wireless_Sensor_Networks.htm

[2] Smart-Its Project, http://www.smart-its.org/

[3] Jianfeng Chen, Alvin Harvey Kam, Jianmin Zhang, Ning Liu, Louis Shue, Bathroom Activity Monitoring Based on Sound, PERVASIVE 2005, 2005.

[4] Susan Cotterell, Ryan Mannion, Frank Vahid, Harry Hsieh, eBlocks - An Enabling Technology for Basic Sensor Based Systems, IPSN Track on Sensor Platform, SPOTS 2005, 2005.