IOSES : An Indoor Observation System Based on Environmental Sounds Recognition Using a Neural Network

Non-member	Takehiko Ashiya	(Keio University)
Member	Masafumi Hagiwara	(Keio University)
Member	Masao Nakagawa	(Keio University)

The demand for security is more and more increasing these days because of high crime rate. This paper proposes a new Indoor Observation System based on Environmental Sounds recognition (IOSES) for a security system. The IOSES is useful for various kinds of needs such as office security, home security, and so on. It uses two microphones in order to detect the direction of the sound sources. The IOSES operates as follows. First, it analyzes a background noise in order to reduce it. Spectrum subtraction method is employed for the noise reduction. The IOSES detects a sound to be recognized using some parameters even if in a noisy environment. The IOSES decides the detected sound using a recognition result by a neural network, the characteristics of the detected sound (intermittence), and the direction of the sound source. The neural network plays an important role in the decision. Neural networks are effective since they do not require modeling of sound to be recognized. After these processes, the IOSES judges the situation of the indoor environment. Computer simulations and experiments in a real environment are carried out. The results indicate the effectiveness of the proposed IOSES for a practical use.

Key words : Indoor observation system, Neural network, Security system, Environmental sound recognition

1. Introduction

With the increase of crime rate, the concept of security is getting more and more important these days⁽¹⁾. There exist various kinds of securities: securities for life, securities for property, securities for data, etc.

Generally observation methods for security can be classified into three categories :

- Automatic observation by machine systems
- Observation by humans
- Observation by combination of humans and machines, for example, monitoring several monitor screens by humans in a central observation room.

There are various kinds of automatic observation systems. Observation is based on sensors such as a camera, an infrared sensor, a magnet sensor, a thermal sensor, etc. In a lot of security systems, a monitor camera is used as the input device, however, humans or guards should continue to watch the screen. This is because an automatic monitoring by a computer is very difficult technically. Even if possible, it requires huge amount of computations or an expensive computer

having high performance. On the other hand, humans are very costly. In addition, humans sometimes make mistakes: the reliability depends on the individual human. Therefore, it is desirable for us to make security systems machine operating.

How do we cope with the conflicting problem?

We take notice of environmental sounds. In our living environment, we are surrounded by various kinds of sounds; we call them "environmental sounds"⁽²⁾. We can guess or judge our environmental situation by hearing the environmental sounds. For example, we can guess someone is walking when we hear a sound of footsteps.

Since we recognize many kinds of environmental sounds, we can get initial information of a change in our environment and guess the situation. This is because an auditory function can be considered as the first perceptive function to notify us of a change in our situation.

Most of the recent researches on sound recognition are on speech recognition. However, development of the environmental sounds recognition system is very important not only for security systems but also for appliances

for deafmutes, measurements, etc. Therefore, we propose an Indoor Observation System based on Environmental Sounds recognition (IOSES) for the first step. The proposed IOSES has many features which are summarized in 5.

This paper is constructed as follows. Overview of the proposed IOSES is given in 2. The basic performance is described in 3. Experimental results are shown in 4.

2. Overview of the Proposed IOSES

In this section, we describe an overview of the proposed IOSES. We assume that the following processes are necessary to construct an observation system utilizing auditory information :

- Detection of a sound and the direction.
- Noise reduction of the sound.
- Estimation of intermittence of the sound.
- Recognition using spectrum of the sound.

The proposed IOSES operates based on the above procedure. The schematic description is shown in Fig. 1. We explain the details of each process.

2.1 Signal Detection Part

The proposed IOSES uses two microphones in order to detect the direction of the sound.

In the signal detection part, the IOSES detects occurrence of a sound in a noisy environment by monitoring a change of amplitude component of the input signal.

In order to detect a sound correctly, we should consider the followings :

- a) The threshold should be small to detect small

sounds.

- b) The threshold should be large to avoid fault detection by noise.

As result, we use the following adaptive threshold:

$$\text{Threshold} = \max\{(P_{\max} - X[\text{dB}]), (N_{\text{ave}} + Y[\text{dB}])\} \quad \dots\dots\dots (1)$$

Where

- $P_{\max}$: peak power of the input signal
- N_{ave} : average of the noise power
- X, Y : variables corresponding to a) and b), respectively.

When the power of input signal exceeds the threshold, the IOSES considers that some sound occurs.

2.2 Noise Analysis Part

It is natural to assume that the monitored sound contains some background noises which should be reduced.

The noise analysis part plays the important role. It analyzes the background noise and extracts two indices to reduce the noise in the noise reduction part : average of the power level and the spectrum.

2.3 Noise Reduction Part and Sound Recognition Part

The IOSES employs the method of spectrum subtraction⁽⁴⁾ to reduce the background noise and get better recognition ability using two indices calculated at the noise analysis part. This method is effective and simple : subtraction in a frequency domain is carried out.

Since environmental sounds contain various kinds of sounds, it is impossible to construct a general model for all of the environmental sounds. This is one of the largest difference between the environmental sound recognition and a speech recognition.

Neural networks, which mimic an information processing process of brain, have been getting more and more attention especially for pattern recognition problems these days. Neural networks do not require detailed knowledge of the input and have an excellent learning ability and a generalization ability. Therefore we employed a neural network in the recognition part. The IOSES uses a multiple-layer neural network : 32 nodes in the input layer, 12 nodes in the hidden layer, and 8 nodes in the output layer. The input data for the neural network is the information of short-time (42 msec $\approx$ 1/6 k $\times$ 256, see Table 1.) power spectrum data calculated by FFT. The input data are normalized between [0, 1.0] at each frame. We determined the number of hidden nodes by trial and error. The number of output nodes corresponds to the number of sounds to be recognized. The

Fig.1. Schematic description of the proposed IOSES.

performance of multiple-layer neural network depends on the number of parameters such as the number of weights and hidden nodes. The methods are surveyed in ref⁽⁶⁾.

As for the learning, the IOSES employs DCP (Dynamically Control training Parameter) method⁽⁷⁾ which is the modified version of the back-propagation learning algorithm⁽⁸⁾.

2.4 Coarse Classification based on Intermittence Part

Generally speaking, each sound has its own characteristics. For example, a sound of footsteps is somewhat periodical, sound from a closed door is somewhat pulsive. The IOSES uses such information: detected signal can be roughly classified into three categories based on the intermittence of the signal: 1) pulsive sounds; 2) intermittence sounds; 3) continuous sounds.

The results of this process are used as additive information to decide the sound at the decision part.

2.5 Detection of Sound Source Direction Part

Let us consider indoor security. It is natural to assume that we know a lot of things about the area: a plan for the area, arrangement of desks, location of doors, etc.

In the IOSES, direction of sound source is estimated in order to make judgement of the situation in the area more reliable. The detection of sound source direction is estimated by calculating the cross-correlation coefficient of the signals from two microphones. It is expressed as:

$$f(\tau) = \frac{\frac{1}{n} \sum_{t=1}^n R(t)L(t+\tau)}{\sqrt{\frac{1}{n} \sum_{t=1}^n R^2(t)} \sqrt{\frac{1}{n} \sum_{t=1}^n L^2(t)}} \dots\dots\dots (2)$$

where $R(t)$, $L(t)$ are the input data at time t from the R-ch microphone, and that from the L-ch microphone, respectively, and τ shows the onset time difference between two microphones.

There are, of course, more sophisticated methods such as in ref⁽⁹⁾. However, since the target of this study is to propose an indoor observation system based on environmental sounds, there is large potential to improve the first version of IOSES.

2.6 Judgement of Environmental Situation

The IOSES judges environmental situation based on the following information:

1. result of coarse classification based on sound intermittence.
2. result of sound recognition by neural network.
3. result of sound source direction.

These results are used to judge the situation in the observation area. For example, the IOSES does not consider a sound as footsteps unless the sound is intermittent.

3. Basic Performance of the IOSES

In this section, we present basic performance of the proposed IOSES. We examined it under a computer fan noise environment.

3.1 Definition of SNR

In the IOSES, a sound to be recognized is considered to exist during the time when the input signal power is higher than some level (30 dB is used in this paper) below of its peak.

Since the background noise level is calculated at the noise analysis part, we define SNR (signal to noise ratio) as

$$\text{SNR} = 10 \log_{10} \frac{S}{N} \dots\dots\dots (3)$$

where S is the averaged signal power, and N is the averaged background noise power.

3.2 Data Sampling and Acoustic Analysis

We performed a lot of experiments to find that most of the frequency components in indoor sounds lies in less than 1 kHz. Therefore we set the sampling frequency as 6 kHz in order to consider frequency components up to 3 kHz.

In addition, the IOSES employs a non-linear frequency band division method to analyze the lower frequency components effectively.

3.3 Signal Input Part

In the IOSES, the interval of two microphones is set to be 34 cm: the difference of 1 sample delay at 6 kHz sampling rate corresponds to about 10 degrees in a direction (see Fig. 2). We consider that the accuracy of 10 degrees is reasonable for a practical use in an indoor environment.

3.4 Noise Analysis Part

The IOSES utilizes a background noise information to reduce the noise component and get more accurate recognition. In this part, an averaged amplitude is calculated and then the averaged noise power spectrum is calculated using FFT analysis.

3.5 Signal Detection Part

In this part, in order to detect a signal to be recognized, two variables, X and Y in eq. (1), are determined.

We carried out experiments to determine the variable X . Fig. 3 shows the relation between the variable X and the detection rate under a fan noise environment using

Fig. 2. Relation between time delay at microphones and sound source direction (theoretical).

Fig. 3. Relation between the threshold X and the detection rate under fan noise environment.

various kinds of sounds to be recognized. In this figure, when the IOSSES takes a fan noise for the other sound we denote the case as “over detection”.

We set X as 16 dB based on Fig. 3. We should mention that the determination of the value is easy and should be adapted to the environment where the IOSSES is used.

Concerning the variable Y , we performed another experiment under the condition $X=16.0$ dB. As a result of this experiment, we found that the IOSSES can detect a sound when $\text{SNR}=-5$ dB at $Y=14$ dB.

Fig. 4 shows an example of detection process of a door sound. Fig. 5 shows another example of a door

Fig. 4. An example of detection process of a door sound.

Fig. 5. An example of a door sound detection under a noisy environment ($\text{SNR}=-5$ dB).

sound detection under a noisy environment ($\text{SNR}=-5$ dB). We can see from this figure that the door sound components are successfully detected under the noisy environment.

3.6 Noise Reduction Part

The input signal is analyzed by Fast Fourier Transform (FFT). Short-time power spectrum on each frame is calculated after subtracting the averaged noise power spectrum to reduce the noise components and is inputted to the neural network in the sound recognition part. The conditions of this part is summarized in Table 1.

3.7 Coarse Classification based on Intermittence

The proposed IOSSES classifies the detected sound

Table 1. Conditions of noise reduction part.

Sampling rate	6kHz
Frame size	42msec
Frame shift	10msec
Number of FFT points	256
Acoustical analysis	Short-time power spectrum
Construction of NN	32-12-8
Learning method of NN	Error back-propagation
Property of Microphone	Single direction
Interval of Microphones	34cm

Table 2. Coarse classification of sounds based on intermittence.

Sound sources	Number of detected sections
Pulsive sounds	.1 ~ 3
Intermittence sounds	4 ~ 6
Continuous sounds	7 ~ 10

Fig. 6. An example of waveform of each sound based on intermittence.

coarsely based on the intermittence as mentioned in 2. 4. The IOSES detects the sections which contain the detected sound and counts the number of sections in the time interval of 2000 msec. Then coarse classification is carried out based on Table 2. Fig. 6 shows an example of waveform of each sound based on intermittence.

3.8 Detection of sound source direction part

Here we examine basic performance of the detection of sound source direction part. We used a door sound as an example. Fig. 7 shows an overview of the experiment. We rotated the microphone stand to detect the sound direction. The IOSES detects the direction based on the correlation of two microphones. Fig. 8 shows the results of the detection. Arrows indicate the peak of

Fig. 7. Overview of the experiment for the detection of sound source direction part.

Fig. 8. Experimental result of detection of sound source direction. (sound source locates on left side of the microphone stand)

cross-correlation for each direction and they correspond to the detected direction. From these figures we can see that the IOSES detected the direction of sounds correctly.

We performed other experiments under a computer fan noise environment and confirmed that the IOSES can detect the direction of sound sources even when SNR is about 0 dB.

3.9 Recognition Target

We selected the following sounds to be recognized.

1. Door sound of opening and closing
2. Window sound of opening and closing
3. Chair squeak sound
4. Drawer sound of opening and closing

Fig. 9. Recognition rate for learning data under noise.

5. Beep sound of a computer
6. Reset sound of a computer
7. Arrival sound of a fax
8. Telephone call sound
9. Fan noise of a computer

These kinds of sounds can be occurred in an ordinary office. If someone tries to invade an office, he or she should open a door or a window. If the invader wants to steal something, he or she would try to search in drawers. If the invader wants data in a computer, he or she should use the computer.

The proposed IOSSES utilizes the information of sound duration time (intermittence). For example, although both the beep sound of a computer and the arrival sound of a fax have almost the same spectrum, the IOSSES can discriminate them by considering sound duration time.

As concerns the data for the neural network, we prepare 15 data samples: 5 are for learning and 10 are for evaluation.

3.10 Recognition of Indoor Environmental Sounds

In order to examine the effectiveness of the noise reduction part, we carried out an experiment under a computer fan noise environment.

Fig. 9 shows recognition rate for learning data under a noisy environment. From this figure it can be observed that the noise reduction part worked very well especially under the SNR of 10 dB. In addition, from our extensive experiments we should mention that SNR is above 10 dB at most cases.

4. Experimental Results of IOSSES

We carried out experiments to examine the total performance of the proposed IOSSES. Some results are shown here.

Fig. 10. Overview of the experimental area.

4.1 Conditions of the experiments

First, we describe conditions of the experiments. We asked a subject to act naturally.

Fig. 10 shows the overview of the area where experiments were performed. As a background noise there was a computer fan noise in this room.

We show the results of the following two experiments. (Experiment-1)

A subject started from the room B, moved across the hallway, and came into the room A. In this process, door sounds of the room A, room B and footsteps sound occurred by the subject. Then the subject walked through the room A and opened the door to the room C. (Experiment-2)

A subject came into the room A, starting from the hallway and walked to the personal computer (pc), turned on the pc and touched the keyboard. In this process, reset and beep sounds of the pc occurred by the subject.

Next, the subject turned to the work station (ws) position, and opened and closed the drawer of the desk. In this process, a chair squeak sound and a drawer sound occurred. Then, the subject moved to the window, and opened and closed it.

4.2 Coarse Classification Based on Sound Intermittence

As explained in 2.4 and 3.7, the proposed IOSSES classifies sounds based on the intermittence. Table 3 shows the result of the experiments under a computer fan noise environment. Although it is difficult to classify

Table 3. Result of sound source classification based on intermittence.

Pulsive sound	Door (room A, room B, room C), Reset, Beep
Intermittent sound	Footsteps (ex1, ex2) , Keytouch, Chair, Drawer
Continuous sound	Window

Fig. 11. Detection of direction of door sound for an example of stationary sound source (experiment-1).

Fig. 12. Detection of direction of footsteps for an example of moving sound source (experiment-1).

sounds having similar spectrum, the sound intermittence information is very useful. We confirmed that the coarse classification process of IOSES can be practical.

4.3 Detection of Sound Source Direction

Figs. 11 and 12 show results of the detection of sound source direction. Fig. 11 shows the detected direction of a door sound as an example of stationary sound source.

Fig. 13. Detection of the direction of indoor environmental sounds (experiment-2).

Fig. 12 shows that of footsteps as an example of moving sound source. We can see that the subject moved from the front of the microphones to the left. Also it can be observed that the IOSES could detect the sound source correctly under a noisy environment even if the sound source moved.

Fig. 13 shows the overall results of the experiment-2. The IOSES could detect all of the sound sources correctly.

4.4 Sound Recognition based on Frequency Analysis

The results of “sound recognition by neural network” part (see Fig. 1) are shown in Tables. 4 and 5.

In the experiments, first, signal detection is carried out for each input and then the sound is recognized by the neural network at each frame. Each number in these tables shows the number of forms recognized as the corresponding sound.

For example, in Table 4, 17 frames of the “Door (room B)” sound is recognized as a door sound. We used the following notations :

- dor : Door sound of opening and closing
- wnd : Window sound of opening and closing
- dsk : Drawer sound of opening and closing
- chr : Chair squeak sound
- fan : Fan noise of computer.
- tel : Telephone call sound
- bep : Beep sound of computer
- rst : Reset sound of computer

These tables adopt confusion matrices. The IOSES should have the ability to infer unknown or unlearned sounds. Therefore in these experiments, some sounds

Table 4. Recognition result of detected frame (experiment-1)

	dor	rst	bep	chr	dsk	wnd	tel	fan
Door (room B)	17	0	3	10	0	2	0	6
Footsteps (Hallway)	6	1	2	9	0	1	1	3
Door (room A)	20	0	1	15	0	1	0	0
Footsteps (room A)	19	0	0	17	2	2	1	3
Door (room C)	12	0	0	8	4	6	0	0

Table 5. Recognition result of detected frames (experiment-2)

	dor	rst	bep	chr	dsk	wnd	tel	fan
Door (room A)	35	0	2	23	0	4	1	1
Footsteps	17	0	1	12	4	2	1	4
Reset	0	29	16	0	0	1	0	0
Beep	0	0	15	0	0	2	0	0
Keytouch	3	0	1	8	4	1	0	0
Chair	9	0	4	34	1	6	1	0
Drawer	24	0	1	1	26	8	3	1
Window	6	0	0	21	14	51	0	0

(footsteps sound and keytouch sound) were not learned on purpose.

As for the unlearned sounds, footsteps sound and keytouch sound, they are categorized to the similar sounds. For example, a sound of footsteps is likely to be categorized as a door sound or a chair squeak sound. However the IOSSES can infer that the detected sound is not a door sound nor a chair squeak sound because of the coarse classification of sound source based on the intermittence.

As for the learned sounds, it can be observed from these tables that all of the learned sounds were correctly recognized.

4.5 Judgement of Situation based on Recognition Result

As an indoor observation system, the IOSSES utilizes 3 kinds of processings to judge the detected sounds more correctly. They are based on the sound intermittence, the sound source direction, and the sound spectrum. In addition, area arrangement information such as shown in Fig. 10 is also available, the IOSSES can judge the detected sound by using these data.

For example, a door sound recognized in the sound recognition part can be confirmed in the detection of sound source direction part: the door sounds should come from about 45 degrees right side of the microphones or the front of the microphones (see Fig. 11). From Figs. 10, 11, Tables 4, and 5, the IOSSES could judge the door sounds correctly. In addition, as a result of Fig. 11, since the peak value of cross-correlation coefficient of the door of room A is larger than that of room B, we can confirm that the door position of room B is farther

than that of room A.

For example of an unlearned sound, a keytouch sound was hard to recognize by Table 5. However, by using the results in the coarse classification based on intermittence part, the sounds considered as some pulsive sound.

Concerning a sound of footsteps, it was not learned on purpose in this case. However, the IOSSES could infer it by utilizing all of the data obtained in the IOSSES. As a result of Table 3, the IOSSES inferred an existence of intermittent sound. Then the IOSSES inferred that the intermittence sound is moving from front to left according to Fig. 13. As a result, the IOSSES judged the intermittence sound as a sound of footsteps.

As explained so far, we confirmed that the proposed IOSSES can recognize the environmental sounds pretty well and judge the situation in the environment.

5. Conclusions

In this paper, we have proposed the IOSSES (Indoor Observation System based on Environmental Sound recognition). As can be imagined from the name, the IOSSES judges the situation of the indoor environment based on a new concept, the environmental sound recognition.

As mentioned in 1, since the IOSSES uses sound data for security, it has the following features:

- It can detect a change of situation in our environment by monitoring environmental sounds under a noisy environment.
- It can guess a change in our environment even if a sound comes from blind spots.
- The amount of sound data is much smaller than that of image data: Too many computations are not required.
- It is possible to construct more sophisticated system by combining the proposed system with an existing security system.
- Since it employs a neural network, various kinds of environmental sounds can be considered.
- In addition, several room acoustical effects such as reverberation, echo pattern, natural frequency, etc. can be reduced by learning ability of neural network.

We carried out computer simulations and experiments in a real environment. The results have indicated the effectiveness of the IOSSES for a practical use.

Finally we have to emphasize that the concept of environmental sound recognition is important not only for security systems but also for appliances for deaf

peoples, various kinds of detectors, and so on.

Acknowledgment

We wish to express our deep gratitude to KUMAHIRA Security Foundation for supporting the research.

REFERENCES

- (1) S. Takenaka: "Home Security", Denki-shoin (1990) (in Japanese)
- (2) T. Ashiya, N. Nakagawa: "A Proposal of a Recognition System for the Species of Birds Receiving Birdcalls —An Application of Recognition Systems for Environmental Sound—", IEICE, Vol. E76-A, No. 10 1858 (1993)
- (3) D. E. Rumelhart, D. E. Hinton & R. J. Williams: "Learning Internal Representations by Error Propagation", Parallel Distributed Processing, Explorations in the Microstructure of Cognition, Vol. 1, Foundations, The MIT Press (1986)
- (4) S. F. Boll: "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", *IEEE Trans., ASSP.*, Vol. ASSP-27, 2, 113 (1979)
- (5) J. Huang, N. Ohnishi & N. Sugie: "Research on a sound source localization system suggested by living organisms—Localization of a single sound source in an anechoic environment—", IEICE, Vol. J71-A, No. 10, 1780 (1988)
- (6) R. Reed: "Pruning algorithms—A survey", *IEEE Trans. Neural Network*, Vol. 4, No. 5, 740, Sept. (1993).
- (7) M. Nakamura & K. Shikano: "A study of English word category prediction based on neural networks", *IEEE ICAS-SP89*, p. 731, (1989)

Takehiko Ashiya (Non-Member) He was born in Hyogo, Japan, on February 14, 1970. He received the B. E. and M. E. degrees in electrical engineering from Keio University in 1992 and 1994, respectively. His research interest includes environmental sound recognition. Since 1994 he has been employed by The Kansai Electric Power Company.

Masafumi Hagiwara (Member) He was born in Yokohama, Japan, on October 29, 1959. He received the B. E., M. E. and Ph. D degrees in electrical engineering from Keio University, Yokohama, Japan, in 1982, 1984, and 1987 respectively. Since 1987 he has been with the department of electrical engineering, Keio University, where he is now an Associate Professor. From 1991 to 1993, he was a visiting scholar at Stanford University. He received Niwa Memorial Award, Shinohara Memorial Young Engineer Award and IEEE Consumer Electronics Society Chester Sall Award, Ando Memorial Award in 1986, 1987, 1990 and 1994, respectively. His research interests include neural networks, fuzzy systems and genetic algorithms. Dr Hagiwara is a member of the IEEE, JNNS and INNS.

Masao Nakagawa (Member) He was born in Tokyo, Japan, on November 19, 1946. He received the B. E., M. E. and Ph. D degrees in electrical engineering from Keio University, Yokohama, Japan, in 1969, 1971 and 1974, respectively. Since 1973 he has been with the department of electrical engineering, Keio University, where he is now a Professor. His research interests include spread spectrum communications, consumer communications and sound processing. He received IEEE Consumer Electronics Society Chester Sall Award, in 1990. He was the executive committee chairman of International Symposium on Spread Spectrum Techniques and Applications 1992.