Verification of Animal External Ear Model-based Microphone Sensor for Hazardous Sound Detection

Miho Yamada[1,*,†], Takeshi Kumaki[2,†] and Kyosuke Kageyama[3,†]

[1]*Graduate School of Science and Engineering, Kindai University, 3-4-1 Kowakae, Higashi-Osaka, Osaka, Japan*
[2]*Dept. of Electronic and Computer Engineering, Ritsumeikan University, 1-1-1 Noji-Higashi, Kusatsu, Shiga, Japan*
[3]*Dept. of Electrical, Electronic and Communication Engineering, Kindai University, 3-4-1 Kowakae, Higashi-Osaka, Osaka, Japan*

Abstract

Hazardous sounds represent a particular type of danger signal and should be heard clearly to ensure safety. Hearing-impaired people can be involved in accidents. This is exacerbated by the fact that they may have difficulty hearing hazardous sounds. Therefore, they spend time with a hearing-assistance dog in their daily lives. However, the number of hearing-assistance dogs has decreased recently, and they are less known in society and they may even be prevented from entering public facilities. Therefore, an animal external ear model-based mobile device that can detect hazardous sounds has been proposed. This device can alert the hearing-impaired person. In this paper, the animal external attachment with a human- and a cat-external ear model-verified for hazardous sound detection accuracy by microphone sensor is presented. Microphone sensors with and without a human- and a cat-external ear model are used. An ambulance siren and a bicycle bell are used as the hazardous sounds, and the spectral envelopes of each recorded sound are compared. The results show that the animal external ear model-based microphone sensor can more effectively detect the frequency features of the ambulance siren and the bicycle bell. A microphone sensor attached an external attachment with a human-external ear model can detect 2 $\sim$ 3 kHz frequency features with the highest sound pressure compared to the other microphones, regardless of the distance between the microphone sensor and the sound source. Also, a microphone sensor attached an external attachment with a cat-external ear model can detect wide range of frequencies with the highest sound pressure compared to the other microphone sensors, regardless of the distance between the microphone sensor and the sound source.

Keywords

Hazardous sound detection, External ear, Hearing-impaired people, Formant, Spectral envelope

1. Introduction

Various sounds are encountered in our daily lives. Among them, hazardous sounds pose possible danger signals. Our safety might depend on being able to hear hazardous sounds such as an emergency bell or the siren of an emergency vehicle. However, hearing-impaired people cannot hear these hazardous sounds. In 2016, there were about 300,000 hearing-impaired people in Japan [1]. The degree of hearing is different for each hearing-impaired person. But many hearing-impaired people can be impacted by accidents. Therefore, some hearing-impaired

The 6th International Symposium on Advanced Technologies and Applications in the Internet of Things (ATAIT 2024), August 19-22, 2024, Kusatsu, Japan

*Corresponding author.
†These authors contributed equally.
miho.yamada97@kindai.ac.jp (M. Yamada); kumaki@fc.ritsumei.ac.jp (T. Kumaki); kageyama@ele.kindai.ac.jp (K. Kageyama)

people spend time with a hearing-assistance dog. This dog supports hearing-impaired people and warns them of danger by touching or pulling them [2]. The hearing-assistance dogs are identified by wearing a cape with words like "Hearing-assistance dog". Thus, people can understand that there is hearing-impaired person in the area [3]. However, the number of hearing-assistance dogs has decreased recently, and only 52 hearing-assistance dogs work in Japan currently [4]. Additionally, hearing-assistance dogs aren't recognized well in society, and they are prevented from entering public facilities such as restaurants [5]. Thanks to having two ears, animals can judge the direction and source of sounds. The external ear consists of two structures: a pinna and an ear canal. The pinna has a sound collection effect and the canal has a resonance effect [6]. The evolution of animal's ears over time has supported them to enable the identification of hazardous sounds. Therefore, animals can detect enemies or danger in real time. Generally, a human, a dog, or a cat can hear sounds of approximately $0.020 \sim 20$ kHz, $0.065 \sim 50$ kHz, and $0.05 \sim 65$ kHz sounds, respectively [7, 8]. A cat can hear frequencies three times wider than humans.

Our new method can detect more hazardous sounds than only a microphone sensor, and alert a hearing-impaired person to hazardous sounds, like the hearing-assistance dog. An animal external ear model-based mobile device for hazardous sound detection has been proposed [9]. This proposed device implements an animal external attachment to a microphone sensor to pick up more detail of sounds. This device can detect hazardous sounds and warn the hearing-impaired persons who are wearing it. In this paper, the animal external attachment follows a human- and cat-external ear model, and has a verified hazardous sounds detection accuracy by the microphone sensor.

The rest of the paper is organized as follows. Section 2 outlines existing technologies. Section 3 explains the animal external ear model-based mobile device for hazardous sound detection. Section 4 describes how to process recorded sounds and its spectral envelope. Section 5 explains the experimental method. Section 6 shows the spectral envelopes of recorded sounds as experimental results, and Section 7 gives our conclusions.

2. Existing technologies

There are four existing technologies to detect hazardous sounds. The first system detects hazardous sounds and situations using clustering by the complete linkage method and probability modeling of daily sounds [10]. These systems need to perform a lot of calculations to process a huge amount of data. And, the second system that monitor the elderly using sound can manage health and detect danger for the elderly to monitor indoor living environment mainly from acoustic signal observed using a microphone sensor. In this system, acoustic events are detected from audio signals observed by microphone sensor using machine learning and obtains information about daily life [11]. However, incorporating machine learning may increase processing requirements and power consumption. Therefore, These systems making it difficult to implement on a mobile device. The third system can detect and determine daily sounds and non-daily sounds which are possibly hazardous using a microphone array [12]. This system is stationary and has a height of about 3 m. So, it is difficult to carry and use. The fourth system is hearing aid. A hearing aid helps the hearing-impaired people hear better. However, they have some disadvantages, such as high performance at a very high price, and the muffled sound and echoes of one's own voice can be bothered.

3. The animal external ear model-based mobile device for hazardous sound detection

The animal external ear model-based mobile device for hazardous sound detection is explained in this section. Overview of this device is shown in Fig. 1 (i). The hazardous sound is an alarm sound, siren, and emergency bell of an emergency vehicle which is designed to make people notice some danger. This device continuously pics up the sounds from the microphone sensor attached to an animal external attachment. The animal external attachment is shaped somewhat like the external ear of an animal (Fig. 1 (i) (a)). The picked-up sound is analyzed with a microcomputer and is determined to be hazardous sound (Fig. 1 (i) (b)). If the sound is judged to be a hazardous sound, the device notifies the hearing-impaired person using LED or other methods (Fig. 1 (i) (c)). This device is implemented as a mobile device.

4. Flow of recorded sound processing

Figure 1 (ii) shows the flow of recorded sounds processing in this experiment. The purpose of this process is to derive the spectral envelope. A sound is recorded by the microphone sensor. And, the recorded data are processed by a short-time Fourier transform (STFT) to find a spectrogram [13]. The power spectrum defined as $\log|\mathrm{X}(\mathrm{k})|^2$, is extracted from one row of the spectrogram. And a discrete Fourier transform (DFT) is applied to each row to obtain the cepstrum. The cepstrum defined as $\mathrm{CEP}(\mathrm{k})$, is considered to be in the form of

$$\mathrm{CEP}(\mathrm{k}) = \mathrm{DFT}(\log|\mathrm{X}(\mathrm{k})|^2)$$

From the cepstrum, the quefrency above a certain threshold is set to 0 to extract the low quefrency components of the cepstrum. Quefrency is a twisted phrase of "frequency" and

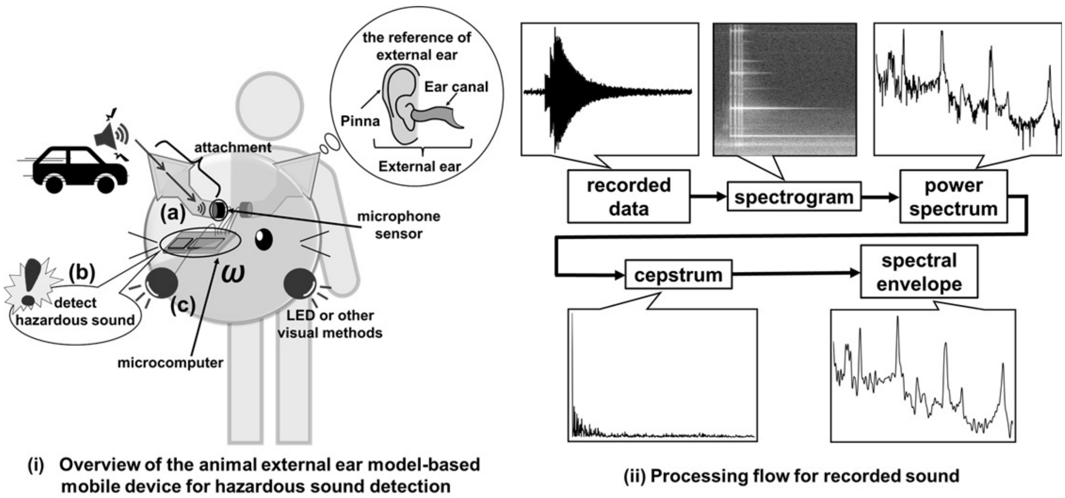

Figure 1: Overview of the animal external ear model-based mobile device for hazardous sound detection and processing flow for recorded sound.

horizontal axis of cepstrum. Such quefrency includes components of the spectral envelope, which is defined as $\log(|\mathrm{highpass(k)}|^2)$. Then, an inverse discrete Fourier transform (IDFT) is applied to the resulting data to find the spectral envelope, which configured as follows $\mathrm{SPE(k)}$:

$$\mathrm{SPE(k)} = \mathrm{IDFT}(\log|\mathrm{highpass(k)}|^2)$$

The spectral envelope includes formants, which are frequency characteristics of a specific frequency region that characterize the sound. In particular, the spectral envelope is useful when indicating vocal-tract characteristics of a human voice. By observing formants can get the frequency characteristics of a vocal tract [14]. In speech recognition, Japanese vowels have long been discriminated by their formant differences. When formants are sorted from lowest to highest frequency, the first and second formants are important clues to discriminate Japanese vowels. Non-human sound sources do not have a human vocal tract. However, it is considered that by observing the spectral envelope, it is possible to confirm the frequency characteristics of any sound.

5. Experimental method

Figure 2 shows the experimental situation. The temperature is 16 degrees Celsius. Three microphone sensors are placed 1 m and 3 m from the sound source. They have the same microphone sensor (model: KY-037). MIC_only has only a microphone sensor. MIC_human has a microphone sensor and attached an external attachment with a human-external ear model. An external attachment with a human-external ear model has the pinna and the ear canal, the length of the ear canal is approximately 0.03 m. MIC_cat has a microphone sensor and external attachment with a cat-external ear model. An external attachment with a cat-external ear model has the pinna and the ear canal, the ear canal is L-shaped and the length of the ear canal is approximately 0.06 m.

The external attachment of all microphone sensors is printed and created with a 3D printer. The used material is thermoplastic polyurethane resin. All microphone sensors are facing sideways. The sound sources are the ambulance siren produced by a speaker (model: ASP-W50N-K) and the bicycle bell. The ambulance siren has frequencies 0.77 kHz and 0.96 kHz at

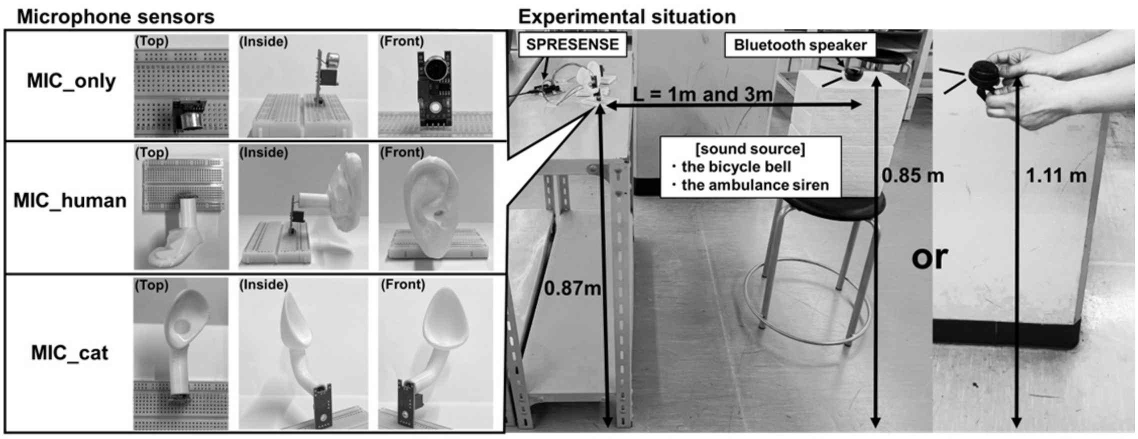

Figure 2: Experimental situation.

intervals of 0.65 s. The bicycle bell is rung by hand. The sound sources are recorded for 15 s, and the recorded sounds are processed in section 3. Three microphone sensors are connected to a SPRESENSE [15], which has a main board (CXD5602PWBMAIN1) and an extension board (CXD5602PWBEXT1). SPRESENSE can record sounds at a sampling frequency of 48 kHz.

6. Experimental result

6.1. The ambulance siren

Figure 3 shows the spectral envelope of the ambulance siren for distances of 1 m and 3 m, respectively. The horizontal axis is frequency and the vertical axis is power. The power is decreased by about 3 dB due to change the distance between the microphone sensors and the sound source from 1 m to 3 m in Fig. 3. It is shown that each formant of the ambulance siren can be detected at around 0.77 kHz and 0.96 kHz at intervals of 0.65 s regardless of the distance between the microphone sensors and the sound source from Fig. 3. Also, Fig. 3 shows that MIC_human detects formant of 2 $\sim$ 3 kHz frequency features with the highest sound pressure compared to the other microphone sensors, regardless of the distance between the microphone sensor and the sound source. On the other hand, MIC_cat detects formants of wide range of frequencies with the highest sound pressure in the range of 0 $\sim$ 7.5 kHz compared to the other microphone sensors, regardless of the distance between the microphone sensor and the sound

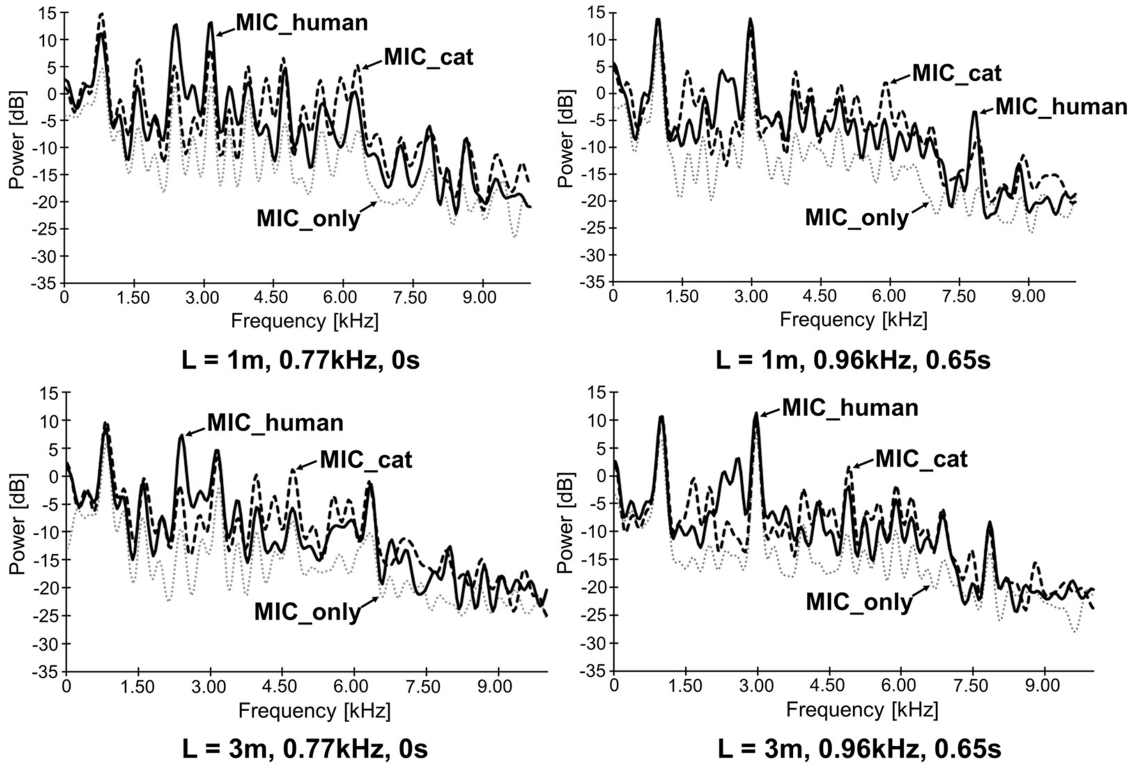

Figure 3: Spectral envelope of the ambulance siren.

source. And the sound pressure detected by MIC_human and MIC_cat are almost greater than the sound pressure detected by MIC_only.

6.2. The bicycle bell

The spectral envelope of the bicycle bell is shown in Fig. 4 for distances of 1 m and 3 m, respectively. The horizontal axis is frequency and the vertical axis is power. The power is decreased by about 1 dB due to change the distance between the microphone sensors and the sound source from 1 m to 3 m in Fig. 4. Figure 4 shows that each formant of the bicycle bell can be detected five places at around 2.0 kHz, 3.6 kHz, 5.0 kHz, 8.5 kHz, and 15 kHz, regardless of the distance between the microphone sensors and the sound source. Also, Fig. 4 shows that MIC_human detects formant of 3.6 kHz with the highest sound pressure compared to the other microphones regardless of the distance between the microphone sensor and the sound source. On the other hand, MIC_cat detects formants of 5.0 kHz, 8.5 kHz, and 15 kHz with the highest sound pressure compared to the other microphones, regardless of the distance between the microphone sensor and the sound source. In other words, MIC_human can detect approximately $2 \sim 3$ kHz frequency features and MIC_cat can detect a wide range of frequency features in the range of $0 \sim 18$ kHz from Fig. 4. And, the sound pressure detected by MIC_human and MIC_cat are almost greater than the sound pressure detected by MIC_only.

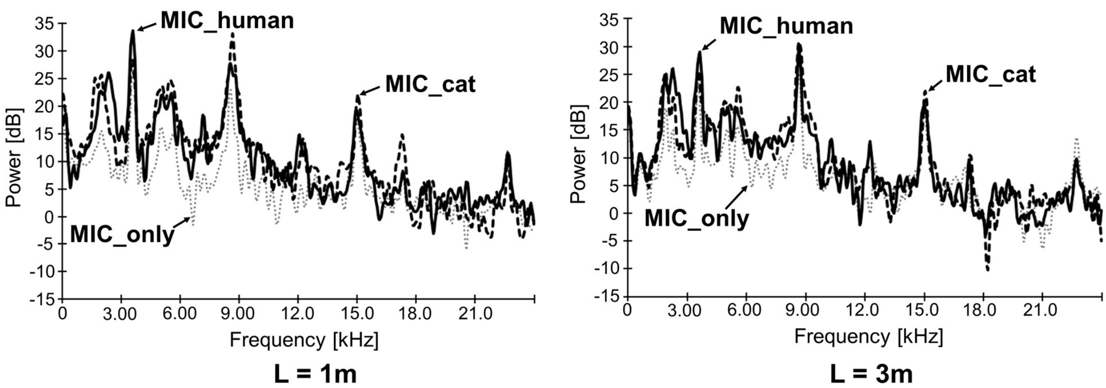

Figure 4: Spectral envelope of the bicycle bell.

7. Conclusion

Our proposed method can detect more hazardous sounds more than only a microphone sensor, and alert a hearing-impaired person to hazardous sounds, like the hearing-assistance dog. In this paper, the animal external attachment with a human- and a cat-external ear model-verified hazardous sounds detection accuracy by the microphone sensor. Three different microphone sensors are compared, and the spectral envelopes of two recorded sounds are detected. The results show a microphone sensor attached an external attachment with a human-external ear model can detect $2 \sim 3$ kHz frequency features with the highest sound pressure compared to the other microphone sensors, regardless of the distance between the microphone sensor and the sound source. Also, a microphone sensor attached an external attachment with a cat-external ear model can detect wide range of frequencies. This enables the highest sound pressure possible compared to the other microphone sensors, regardless of the distance between the microphone sensor and the sound source. Thus, the animal external ear model-based microphone sensor can more effectively detect the frequency features of the ambulance siren and the bicycle bell. Therefore, it makes easier to detect hazardous sounds by adapting the external attachment of the microphone sensor to match each hazardous sound in our proposed device. In the future, experiments will be conducted in various environments, including noisy environments and changing temperatures. In addition, various external ear models will be verified to detect hazardous sounds more effectively by changing the shape and material of external attachment. Furthermore, effectiveness of our proposed method which detects other hazardous sounds that are common in daily life is verified.

References

[1] https://www.mhlw.go.jp/toukei/list/dl/seikatsu_chousa_c_h28.pdf (Accessed, Apr. 2024)

[2] https://www.hojyoken.or.jp/what/hearingdog (Accessed, Apr. 2024)

[3] https://www.hearingdogjp.org/about/workandrole (Accessed, Apr. 2024)

[4] https://www.jsdrc.jp/hojoken/chodoken_suu/ (Accessed, Apr. 2024)

[5] https://grapee.jp/99344 (Accessed, May. 2024)

[6] J. Melloui, O. Bouattane, and J. Bakkoury, "Study of the effect of a cause of tinnitus on the resonant frequency of the outer ear," 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), pp. 1–4, 2020.

[7] S. L. Mak, S. L. Au, W. F. Tang, C. H. Li, C. C. Lee, and W. H. Chiu, "A Study on Hearing Hazards and sound measurement for Dogs," 2022 IEEE International Symposium on Product Compliance Engineering – Asia (ISPCE-ASIA), pp. 1–4, 2022.

[8] https://www.signia.net/ja-jp/blog/local/ja-jp/cat-dog-human-hearing/ (Accessed, Apr. 2024)

[9] M. Yamada, T. Kumaki, and K. Kageyama, "Proposal of hazard sound detection system based on the external ear model," IPSJ SIG Technical Report, Vol. 2023-EMB-64, No. 9, pp. 1–4, Nov. 2023.

[10] K. Suzuki, M. Tanaka, and H. Inoue, "Clustering and modeling of hazardous or non-hazardous sound in daily life," 2012 Proceedings of SICE Annual Conference (SICE), Japan, pp. 1991–1995, 2012.

[11] https://www.fujitsu.com/jp/about/resources/case-studies/trends/pdf/cs-201710-connectedlife.pdf. (Accessed, Jun. 2024)

[12] M. Kawamoto, F. Asano, and K. Kurumatani, "A Security Monitoring System of Detecting Unusual Sounds and Hazardous Situations by Sound Environment Measurement Using Microphone Arrays," IPSJ SIG Technical Report, Vol. 2008-UBI-19, No. 66, pp. 19–26, Jul. 2008.

[13] J. Allen, "Short term spectral analysis, synthesis, and modification by discrete Fourier transform," IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 25, No. 3, pp. 235–238, Jun. 1977.

[14] M. Yokoyama, "Distinction of timbre of violin by spectral envelope," IPSJ SIG Technical Report, Vol. 2020-SLP-132, No. 26, pp. 1–6, May. 2020.

[15] https://www.sony-semicon.com/ja/products/spresense/index.html (Accessed, Apr. 2024)