PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Detection of Boat Noise by a Convolutional Neural Network for a Boat Information System

Haruki YAMAGUCHI[1]; Kenji MUTO[2]

[1,2] Shibaura Institute of Technology, Japan

ABSTRACT

Some boat noises are perceived as noisy and annoying by people who live near canals. We have previously proposed an information system that uses audiovisual means to provide cellphone alerts of approaching noisy boats, but a problem with that system was camera-based detection of a boat approaching at night. In the present paper, we investigate using training data to detect boat noise in environmental sound by means of a convolutional neural network. To detect boat noise, training data are used involving spectrograms of the environmental sound. The spectrogram configuration is investigated to improve the detection of boat noise. From the results, when the spectrogram configuration has a time axis of 5 s and a frequency axis of 10–3,500 Hz, the detection performance has a highest accuracy of over 95%.

Keywords: Noise detection, Boat noise, Sound recognition, Convolutional neural network

1. INTRODUCTION

In daily life we are surrounded by numerous sounds, some of which are perceived as uncomfortable, such as the noise under a railway viaduct or from construction or engines. The canals of Tokyo in Japan have been used to carry food and supplies since the 1800s, and some are now also used for freight and sightseeing boats. Each day, more than 100 boats use one such canal, alongside which stand apartment and office buildings in close proximity (1). Of the various boats, tugboats generate high levels of noise when transporting sand or construction materials. The noise level from their engines can reach 70 dBA in a building, which is almost the same as that from a vacuum cleaner. Therefore, the people who live alongside this canal are exposed to unpleasant sounds.

The impression of sound depends on the environment. Miyagawa et al. showed that the evaluation of sound depends on whether it is accompanied by video (2). According to Abe et al., adding visual information to sound stimulation influences the quantitative evaluation of the sound (3). Kai et al. showed that providing visual information about a boat as a sound source decreases the impressions of noisiness, loudness, and annoyance (4).

In previous work, we proposed a cellphone-based noise information system that uses audiovisual means to decrease the annoyance of boats. This system for detecting boats comprises a camera and a microphone, but we developed the system to detect boats by means of only the camera. Consequently, a problem with the system was the difficulty of boat detection at night (5). In the present paper, we describe a method for detecting boats by using acoustic data recorded with a microphone. To detect boat noise, the sound detection system analyzes the audio signal by means of a convolutional neural network (CNN), which is used widely in image recognition. Although many researchers in the field of machine learning are studying how to recognize traffic noise, those sound recognition systems are not targeted at boat noise (6-9). We discuss the important parameters for the CNN trained on data that involve the frequency characteristics of the audio signal.

2. SELECTION OF TRAINING DATA FOR CNN

We describe the elements of the CNN training data involving boat noise. Using the CNN requires pretreating the spectrograms used as the input data for training the CNN. The process is widely used in the field of speech recognition. The spectrograms require the elements of time range, frequency range, frequency resolution, and overlap. In this section, we describe the choice of the spectrogram parameters. First, we recorded the environmental sound near the canal. We then had people manually classify the recorded sound.

[1] ma19085@shibaura-it.ac.jp; [2] k-muto@shibaura-it.ac.jp

Finally, we transformed the sound into spectrograms so that the CNN could learn the boat noise, which is why the system requires three-dimensional data.

2.1 Evaluation Method

We used the CNN to evaluate the spectrograms, and we calculated the accuracy P_{AR}. The results of the classified data in Section 2.4 are indicated by the labels "Boat" or "NoBoat" every second. The accuracy rate is given by

$$P_{AR} = \frac{R\,[s]}{T\,[s]} \times 100\ [\%],$$

where R is the number of labels per second classified by the CNN that matched those classified by people, and T is the total analysis time. From this formula, we calculated the accuracy rate of the spectrogram performance for each frequency range and each training time.

As the training data, we used sound recorded for 24 h on March 16, 2015 at an apartment veranda alongside the canal. Each second of this data set was classified by people, and the classification is described in detail in Section 2.3. For the data with which to verify the CNN training, we used sound recorded from 7:00 am to 7:30 am on March, 17 2015, each second of which was also categorized by people. The verification data were recorded in the same environment as were the training data.

2.2 Spectrogram Data for CNN

To train the CNN to recognize sound, we used spectrograms. In this paper, we discuss the frequency range of the spectrograms. Each spectrogram comprises three-dimensional data on a time axis, a frequency axis, and a sound-level axis. We reason that the sound classification by the CNN depends on the spectrograms. Therefore, we discuss the spectrogram composition. We determined the spectrogram frequency range by focusing on the sound source of the boat. Boat noise contains engine sound, which is a low-frequency noise. We studied the 35 patterns listed in Table 1, in which we changed the range with regard to the maximum frequency for the spectrogram while the minimum frequency was set to 10 Hz. When the spectrogram shows the frequency, which generated by frequency bands. The number of frequency bands were 2 to 95. In this experiment, we searched for an influence on accuracy from those maximum frequency ranges of 50 Hz or higher.

For another spectrogram parameter, we considered the length of analysis time for a spectrogram to fit the boat noise in our previous study work, which result of time length at 5 s (10). Therefore, we used an analysis time of around 5 s to give the spectrogram train for CNN from previous study. Last one of the spectrogram parameter, Figure 1 shows two examples of the spectrograms used to train the CNN. Each spectrogram of the sound data was obtained using a fast fourier transform (FFT), for which the characteristic values were a sampling frequency of 24 kHz, a window size of 4,096 samples, and an overlap of 8.4%. This FFT window size can detect frequencies as low as 5.9 Hz. The example spectrograms shown in Figure 1 confirm that they differ according to the frequency range. The spectrogram to train CNN was calculated using the speech spectrogram functions from MATLAB (11). Hence the frequency bands of spectrograms were generated based on the bark scale.

Table 1 – Spectrogram frequency ranges used to train the convolutional neural network (CNN)

Frequency range [Hz]				
10 – 50	10 – 250	10 – 2500	10 – 6000	10 – 10000
10 – 75	10 – 500	10 – 3000	10 – 6500	10 – 10500
10 – 100	10 – 1000	10 – 3500	10 – 7000	10 – 11000
10 – 125	10 – 1250	10 – 4000	10 – 7500	10 – 11500
10 – 150	10 – 1500	10 – 4500	10 – 8500	10 – 12000
10 – 175	10 – 1750	10 – 5000	10 – 9000	
10 – 200	10 - 2000	10 – 5500	10 – 9500	

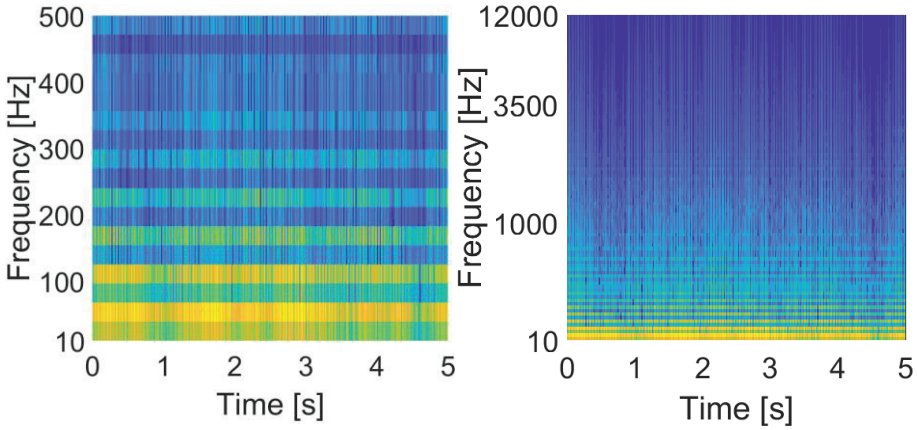

Figure 1 – Examples of spectrograms used to train CNN. Left: 10–500 Hz analyzed; right: 10–12,000 Hz analyzed.

2.3 Sound Classification for Boat Notification System

The CNN requires classification when trained using sound data. For training, we categorized each second of sound data using one of three labels, namely. the Boat label if the sound data contained boat noises, the NoBoat label if the sound data contained no boat noises, and the NoTraining label, which is categorized as intermediate between the "Boat" and "NoBoat" labels. To assign these labels to the sound data, we listened to the latter through headphones to detect any boat noise in each second of the sound data. For the training data, we used a 24 h recording made on March 16, 2015.

In this study, the CNN was not trained on the NoTraining label to avoid any fuzziness regarding the boat sound. On the labeling, we judged perceptible by the ear the period of boat noise. Figure 2 shows the labeling process schematically. During periods classified with the Boat label, perception of boat noise was evaluated, except for the initial 10 s and the final 10 s.

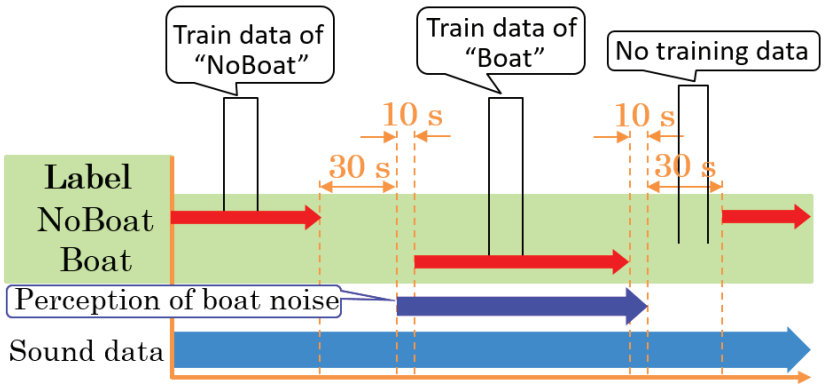

Figure 2 – Environmental sound including the passage of a boat and time diagram of sound classification of boat noise to train the CNN. Sound data given the "Boat" label contain boat noise, and those given the "NoBoat" label exclude boat noise. The labels were assigned by people based on perceived boat sounds.

2.4 Training Model of CNN

In this research, we used a CNN comprising 24 layers, which included five convolutional layers (11). The first layer of the CNN was for inputting the three-dimensional spectrogram data, which compress by a time value, frequency value and sound volume value. The size of input data of spectrogram for CNN was 350 × (2 to 95) × 1. Each convolutional layer was filtered by a 3 × 3 filter. The 20% data from previous layer was eliminated on dropout layer. The CNN trained spectrogram with the sound classification label. Figure 3 shows the training of the CNN model.

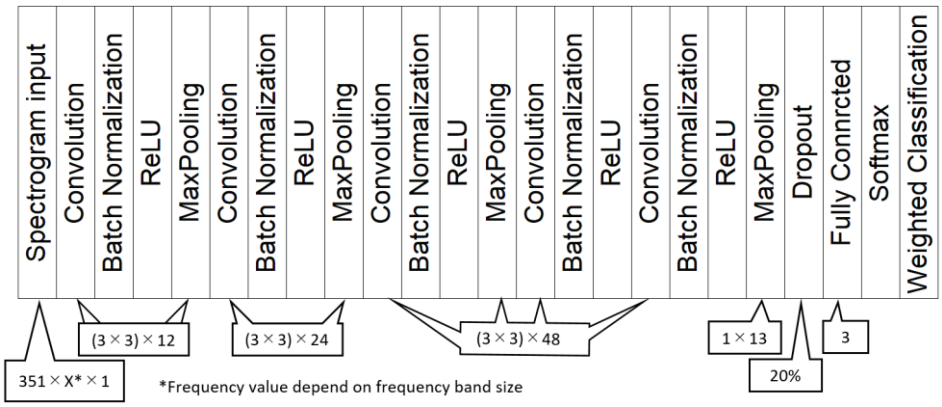

Figure 3 – Layers used to train the CNN when using spectrograms as input data.

2.5 Recording Conditions

As training data, we used sound recorded for 24 h by a microphone (MI-1233; Onosokki Co. Ltd) on March 16, 2015 on an apartment balcony alongside the canal. The microphone was sited 28 m above the level of the canal and 16 m horizontally from the canal. The recording conditions were the same as those used in our previous paper. On the day of the recording, the weather was variously sunny or cloudy and the wind was light. The sound was recorded at a sampling frequency of 24 kHz and was quantized into 16 bits.

3. RESULTS FOR FREQUENCY RANGE

3.1 Results

We performed an experiment on boat noise detection to improve the frequency range of the spectrogram of data for training the CNN using 30 min verification of the environmental sound data near the canal. Figure 4 shows the relationship between the accuracy rate of the CNN and the frequency range of each spectrogram. In this experiment, the highest accuracy rate was 95.6% for the frequency range of 10–3,500 Hz, and the lowest accuracy rate was 94.4% for the frequency range of 10–3,000 Hz. Those frequency ranges were the ones that were most effective for detecting boat noise in this experiment. An accuracy rate of 75–80% was obtained in the low-frequency range of 10–50 Hz, and an accuracy rate of over 90% was obtained in the 10–125 Hz range. Those results show that the frequency characteristics between 50–125 Hz were influential for recognizing boat noise using the spectrograms. In the frequency range of 10–1,000 Hz, an average accuracy rate of 93.1% was obtained, and this accuracy rate agrees with results of previous studies (10). However, the accuracy rate dropped to below 90% in the frequency ranges of 10-4,500 Hz and 10–10,000 Hz.

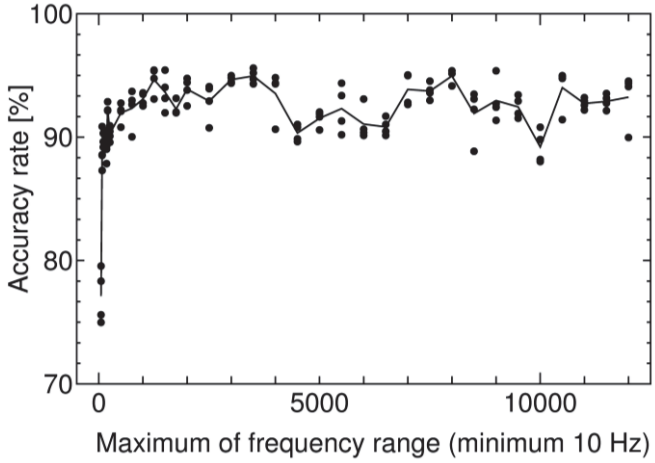

Figure 4 – Relationship between CNN accuracy and frequency range. The horizontal axis shows the upper limit (from 10 Hz) of the spectrogram frequency range, and the vertical axis shows the accuracy rate of the CNN using the training data of the spectrogram for each frequency. The line shows how the average accuracy varies with the maximum frequency.

3.2 Discussion

From the results of using spectrograms analyzed by a CNN to recognize boat noise, the present system could recognize all the boats in the reported experiment. Regarding the recognition accuracy, when the upper limit of the frequency range exceeded 125 Hz, the accuracy rate mostly exceeded 90%.

From the experimental results regarding the frequency range, much of the information required for recognizing boat noise lies in the 50–1,250 Hz frequency characteristics, which improve the average accuracy by 17.6%. Figure 5 shows that boats equipped with a diesel engine have peak frequencies below 50 Hz. It is effective to use not only the range around the fundamental frequency of boat noise but also the range of higher frequency for boat noise detection.

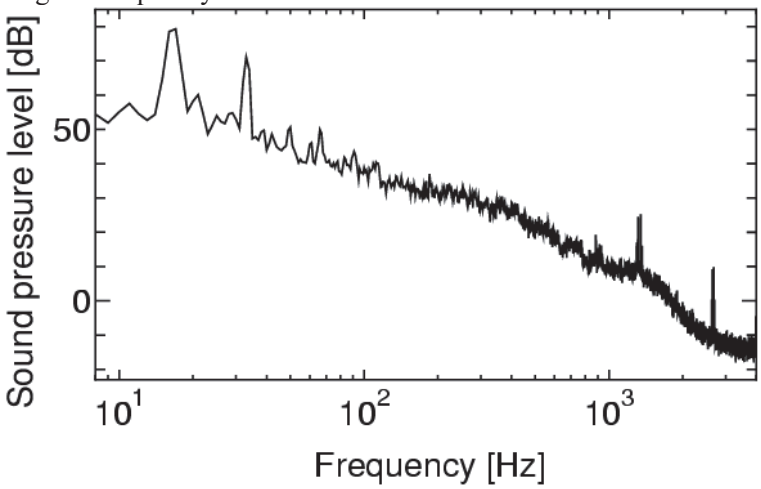

Figure 5 – Example of frequency characteristics of tugboat noise. In this case, the peak frequencies of the boat noise were 17 Hz and 33 Hz.

4. CONCLUSIONS

We proposed a boat notification system for residents along a canal to reduce the impression of boat noise. This system required a method for detecting boats from environmental sound, and we proposed doing so by using a CNN to analyze spectrograms. In the reported experiment, we focused on how the spectrogram parameters depends on the characteristics of the boat noise. We discussed the parameter of frequency characteristics on spectrogram for improving the detection of boat noise. The results showed the frequency range of 50–1,250 Hz to be a valid one for detecting boats using a CNN trained by spectrograms. In this paper, we focused on the boat detection using noise, which was influenced by low-frequency sounds originating from boat engines. In future work, we will classify boat sounds according to the category of boat.

REFERENCES

1. Muto K., Akahira T. A result of 24-hours measurement of small boat noise in residential area along canal. Proc. 12th Western Pacific Acoustics Conference 2015; 7 December 2015; Singapore 2015. pp.221-224.
2. Miyagawa M., Suzuki S., Aono S., Takagi K. The effect of visual information on the impressions of environmental sound (in Japanese). The journal of the acoustical society of Japan; vol. 56. No.6. pp.427-436, 2000.
3. Abe K., Sato G., Takane S., Sone T. Influence of visual information on loudness evaluation of environmental sound (in Japanese). Technical report of noise and vibration in ASJ. N-2005-34. 2005. pp.1-8.
4. Kai M., Muto K. Evaluation level of boat noise on presentation time length of video for noise alert system (in Japanese). Autumn Meeting Acoustical Society of Japan; 2018. pp.419-420.
5. Akiyama T., Kobayashi Y., Kishigami J., Muto K. CNN-Based Boat Detection Model for Alert System Using Surveillance Video Camera. Proc. 7th Global Conference on consumer Electronics 2018; Nara, Japan 2018. pp.634-635.
6. Zhang X., Zou Y., Shi W. Dilated convolution neural network with LeakyReLU for environmental sound classification. IEEE 2017 22nd International Conference on Digital Signal Processing, 2017.
7. Salamon J., Bello P Juan. Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification. IEEE Signal Processing Letters, Vol. 24, Issue. 3, pp.279-283,

2017.
8. Chu S., Narayanan S., Kuo C.-C. J. Environmental Sound Recognition with Time–Frequency Audio Features. IEEE Transactions on Audio, Speech, and Language Processing, Vol. 17, Issue 6, pp.1142-1158, 2009.
9. Medhat F., Chesmore D., Robinson J. Environmental Sound Recognition Using Masked Conditional Neural Networks. Advanced Data Mining and Applications: 13th International Conference, pp.373-385, 2017.
10. Muto K., Yamaguchi H. Study of train data with environmental sound for boat noise detection by convolutional neural network (in Japanese). Spring Meeting Acoustical Society of Japan; 2019. pp.363-364.
11. MathWorks. Speech Command Recognition Using Deep Learning. https://mathworks.com/help/deeplearning/examples/deep-learning-speech-recognition.html; April 16 2019.