Acoustic event detection of ultrasonic signals for Inaudible sound communications using CNN

Kosei OZEKI†, ***Student Member***, **Naofumi AOKI**†, ***and*** **Yoshinori DOBASHI**†, ***Members***

1. Introduction

This study has developed a communication system using inaudible bands of sound signals. Unlike radio communication systems using advanced wireless devices, it only requires the legacy devices such as microphones and speakers employed in ordinary telephony communication systems [1]. In this study, we have investigated the possibility of acoustic event detection using a convolutional neural network (CNN)[2] for 40 channels ultrasonic signals.

2. Proposed technique

We attempt to design a communication system based on frequency shift keying (FSK) modulation that embeds 40 ch sound signals in inaudible bands between 18 kHz and 20 kHz. As shown in Fig.1, sound signals including guard intervals to adjust their density are employed. The guard intervals are (a) 100 %, (b) 50 %, and (c) 25 % of the signal length. In this paper, we prepared four datasets. The datasets labeled (1)-(3) are composed by (a)-(c) guard interval signals, respectively. And the dataset labeled (4) was randomly composed by (a)-(c) guard interval signals.

Fig.2 shows the proposed CNN structure for acoustic event detection. Its kernel size is 3x3. The activation functions is ReLU. Max pooling, batch normalization, and 25 % dropout on all layers except the final layer are employed. The max pooling reduces the dimension only in the frequency direction in order to maintain the time resolution. The final layer uses a kernel size of 1x1 convolution. Its activation function is sigmoid. For the learning process, ADAM is used for the optimizer. The loss function is binary cross entropy. The learning rate is 0.001 for 200 epochs.

we designed classifiers labeled (1)'-(4)' from the trained datasets (1)-(4), respectively. In total, 16,000 sound signals and correct event sets was prepared as the learning data. In addition, 1,600 verification data which were not used in the learning process was prepared. In the experiment, the four classifiers recognized unused signals from the datasets (1)-(4). The results are shown in Table 1. The recognition accuracy of each classifier was over 95 %. The proposed technique shows good results even if the SNR of the sound signals was degraded.

†The authors are with Graduate School of Information Science and Technology, Hokkaido University, Sapporo, 060-0814 Japan.

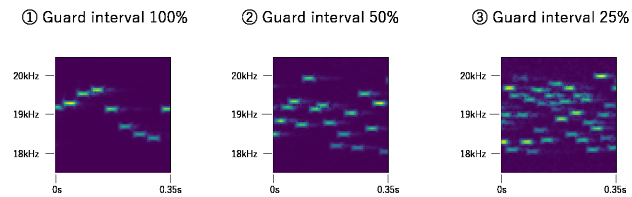

Fig. 1 Spectrogram patterns used in learning

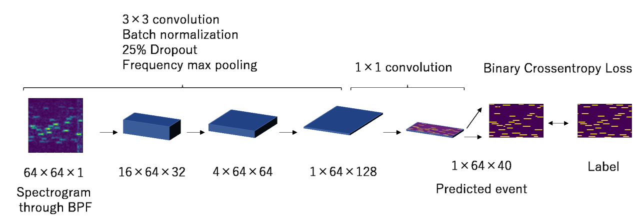

Fig. 2 CNN structure for acoustic event detection

Table 1 The recognition accuracy

Guard intervals of test datasets	Classifiers (1)' 100 %	(2)' 50 %	(3)' 25 %	(4)' random
(1) 100 %	99.94 %	97.59 %	91.92 %	96.45 %
(2) 50 %	99.85 %	99.85 %	96.80 %	98.84 %
(3) 25 %	79.63 %	95.90 %	99.10 %	91.72 %
(4) random	99.72 %	98.70 %	96.35 %	98.23 %

3. Conclusions

In this study, the recognition accuracy of ultrasonic signals is evaluated by acoustic event detection using CNN and compared by adjusting the signal density with a guard interval. As a result, it is indicated that the proposed technique may be applicable to inaudible sound communications. We plan to conduct some more evaluation considering real situation as a future work.

4. Acknowledgments

This was supported by Smart Solution Technology, Inc, and the Hokkaido University Ambitious Doctoral Fellowship.

References

[1] K.Ozeki, et al, "Improving the recognition accuracy of a sound communication system designed with a neural network," IEICE, vol.E104-A, no.11, 2021

[2] Zhang, et al, "ROBUST SOUND EVENT RECOGNITION USING CONVOLUTIONAL NEURAL NETWORKS", in Proc. of ICASSP(2015)