RECOGNITION OF ENVIRONMENTAL SOUND RECORDED BY MOBILE PHONE

Reona Mogi / Satoshi Innami / Hiroyuki Kasai
The University of Electro-Communications
mogi@appnet.is.uec.ac.jp / innami@appnet.is.uec.ac.jp / kasai@is.uec.ac.jp

ABSTRACT

This paper presents a feature extraction and recognition method of recorded environmental sound by mobile device. We propose a new method to improve environmental sound recognition rate by combining Independent Component Analysis (ICA) and Matching Pursuit (MP) method. Simulation results showed that there was improvement in the recognition rate.

Keywords

Environmental Sound, auditory scene recognition, feature extraction, independent component analysis, matching pursuit

1. INTRODUCTION

Recently, sophisticated mobile devices like smart phones have been developing tremendously. This type of device allows us to take photos and record surrounding sounds easily. Moreover, we can utilize many services which use various sensors such as acceleration sensor. If we can recognize context from sound captured in many places, we can understand situation of the place which is not able to recognize from photo and map. Because mike of mobile phone is for calls, performance to capture sounds is inferior to recorder's mike. In addition, captured sounds may include unintentional noises. We propose acoustic feature extraction method to consider this effect and applied to sound recorded by mobile phone.

2. FEATURE EXTRACTION

We extract acoustic feature because of environmental sound recognition. Environmental sounds have various features, such as traffic, alarm and birdsong. Mel Frequency Cepstral Coefficient (MFCC) is a traditional frequency-domain feature value. In MFCC, it is difficult to extract correct feature value from sound sources which contain noise. Besides that, MFCC also has difficulty to extract feature value from sources which changes frequently within a short duration. We believe that ICA [1] is able to extract sound characteristics even the source was corrupted by noise because components within the source were assumed to be independent and thus prediction is possible. Moreover, MP [2] is a method to extract sound characteristics in time-domain. MP feature value is more robust to changes in sound. We propose acoustic feature extraction method by combination ICA and MP, and investigated the effectiveness of the proposal method.

3. CLASSIFICATION METHOD

Environmental Sounds are recognized by feature extraction value. We confirmed that the recorded sounds could classify by supervised data. As environmental sound classification method, we utilized K-Nearest Neighbor (kNN) and Gaussian Mixture Model (GNN), and compared recognition rate of two classification methods.

4. EXPERIMENT RESULT

We performed simulation experiments using real environmental sounds recorded by iPhone. We prepared 6 classes sound types; birdsong, station wicket, park, railway crossing and downtown. We derived supervised data of each class from 732 sound data. Subsequently, we classified 312 test data into 6 classes based on kNN and GMM.

Figure 1 shows the result of recognition rate by means of GMM of 4 mixtures and kNN clustering as k = 3. As shown in Fig. 1, it was confirmed that ICA feature extraction is higher recognition rate than MFCC. Moreover, it was confirmed that GMM is higher recognition rate than kNN. Therefore, GMM is able to represent correctly environmental sounds. Figure 2 shows the result of GMM recognition when number of mixture was changed. It confirms that number of mixture is high recognition rate between 2 and 8. Regarding MFCC, it seems that recognition rate is constant, because this experiment didn't be considered time change such as delta MFCC.

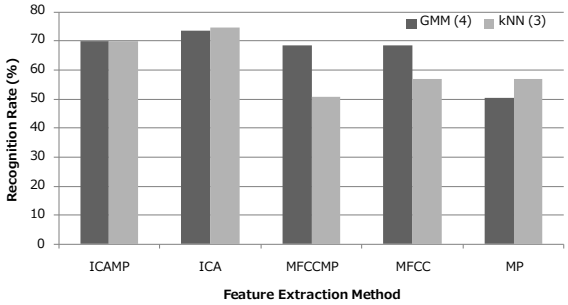

Figure 1. Recognition Rate of Each Feature Extraction

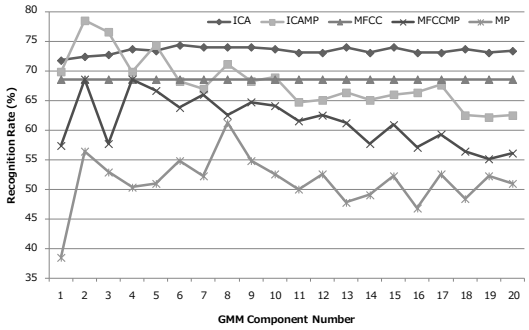

Figure 2. Recognition Rate by GMM Component Number

5. CONCLUSION AND FEATURE WORKS

This paper reported an actual recorded environmental sound recognition method using ICA and MP. From the simulation results, higher recognition rate was achieved by combining these two features. We are investigating method to recognize background sound and characteristic sound at once.

6. REFERENCES

[1] A. Hyvärinen, and E. oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13 (4.-5), pp. 411-430, 2000.

[2] Antti J. Eronen, Vesa T. Peltonen, Juha T. Tuomi, Anssi P. Kalapuri, Seppo Fagerlund, Timo Sorsa, Gaëtan Lorho, and Jyri Huopaniemi, "Audio-based context recognition," *IEEE Transaction on Audio, Speech, and Language Processing*, vol. 14, no. 1, pp. 321-329, 2006.