24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

FEATURE EXTRACTION AND RECOGNITION OF WALL IMPACT NOISE

Zeng Xiangyang, Wang Qiang, Wang Lu, Lu Chenxiang
Key Laboratory of Ocean Acoustics and Sensing, Northwestern Polytechnical University
School of Marine Science and Technology, Northwestern Polytechnical University,
Xi'An, Shannxi, China, 710072
E-mail: zenggxy@nwpu.edu.cn

Recognition of the impact noises emitted by electrical-drill or hammer is significant in such applications as building sound event monitoring system. The methods based on spectral features have been proved to have poor performance, therefore, in this paper a novel temporal feature has been proposed considering the similarity and volatility in frequency domain and the difference in time domain between various types of noises. When using measured data to testify the performance, it can be found that the recognition accuracy of the temporal features is better than that of the popular spectral features such as MFCC and the feature is noise robust .

Keywords: Feature extraction, Wall impact noise, Acoustic signal recognition.

1. Introduction

Acoustic signal classification and recognition is very interesting and important issue in many fields such as speech recognition, underwater target detection and other environmental sounds classification [1][2]. At present, statistical learning based methods are usually applied for such tasks. Feature extraction is crucial in these methods. Traditional features extraction framework is always based on the difference of the spectrum characteristics of different classes, and the most popular one of them is Mel-Frequency Cepstrum Coefficient(MFCC) [3][4]. In many fields such as speech recognition, speaker identification, and other sound classification task, MFCC is very prevalent and can efficiently extract the most distinctive information of sound sources, i.e., the channel characteristics of vocal organs. However, the generation mechanisms of the acoustic signals are ever-changing. The spectrum based features cannot catch the discriminative features due to the similarity of spectrum between different classes. In another word, those features lack of the representation on the time domain characteristics.

In this paper a novel temporal feature extraction method has been proposed considering the similarity and volatility in frequency domain and the difference in time domain between various types of noises. In section II we analyze the characteristics of two types of acoustic signals and then introduce the novel feature extraction method. Finally in the experiments, the recognition performance of the traditional feature(MFCC) and the proposed features are compared.

2. Temporal Feature Extraction Method

2.1 Comparison Between Impact Noises Emitted by Electrical Drill or Hammer

In the application of building sound event monitoring, the impact noises emitted by electrical-drill or hammer is necessary to be classified. Fig.1 shows the spectrogram of electrical drill(right) and banging(left) sound. It can be seen that the spectral of banging sound varies from the beginning to the end and the high frequency component has an obvious reduction. The electrical drill sound is more stable than the banging sound and its distribution is similar to the beginning of the banging sound. If only the spectrum features are used, the recognition accuracy is not well. However, it can be found that these two types of sounds can be easily distinguished by our ears. The following two reasons probably can explain the phenomenon. Firstly, since the two excited forces are all broadband, the early stages of the two sounds become very similar in the frequency domain. Secondly, due to the highly time-variety characteristics of banging, the samples within this class are not similar to each other in different time interval. Therefore, alternative features should be applied for such cases.

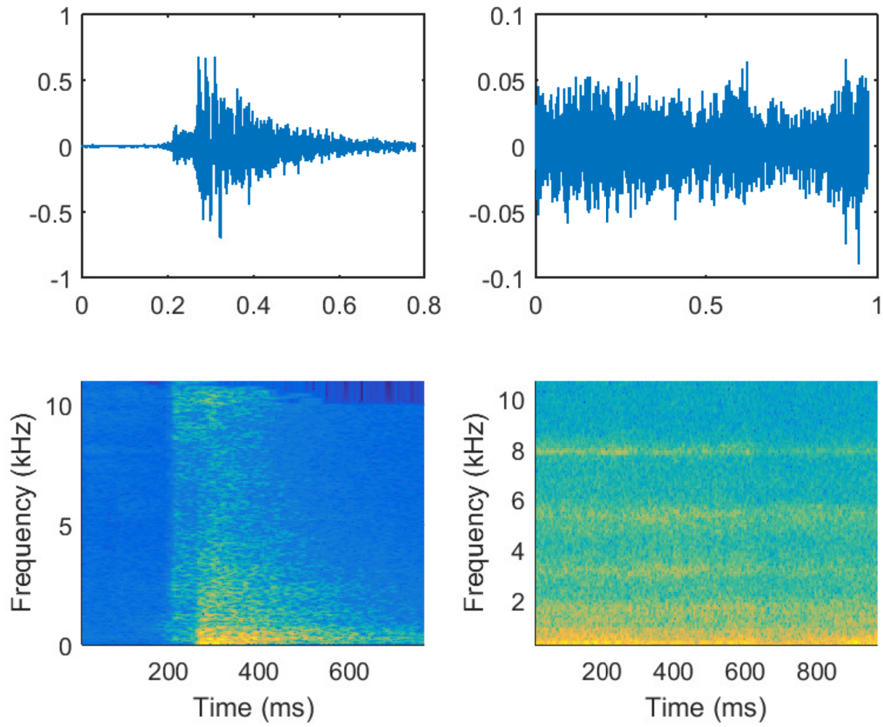

Figure 1: waveform and spectrogram of electrical drill(right) and banging(left) sound.

2.2 Temporal Feature Extraction Method

Here we propose a feature extraction method which can utilize the time-variant characteristics of aforementioned sounds. The schematic diagram of the feature extraction procedure is given by Fig.2.

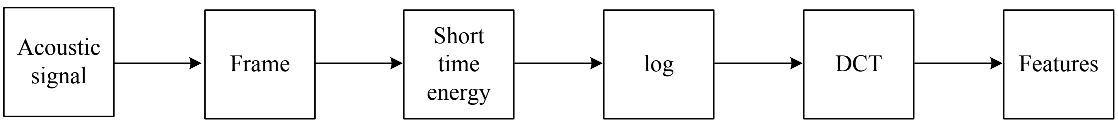

Figure 2: schematic diagram of feature extraction procedure.

We design the procedure to extract the features with different characteristics of signal in the time domain to overcome the low time resolution of the spectrum features. Generally the banging sound has an exponential decay(see Fig.3), and the short time energy is usually a part of the energy decay. The proposed method uses logarithm for better exploring the attenuation law of this kind of sound.

Then DCT(discrete cosine transform) is applied to further reflect the variation trend of the EDC and reduce the dimension of the features to wipe out the details of the curve. The first 5 DCT coefficients are reserved and they are corresponding to the low frequency components of the logarithmic energy decay curve(EDC). The feature extraction algorithm is shown in Alg.1.

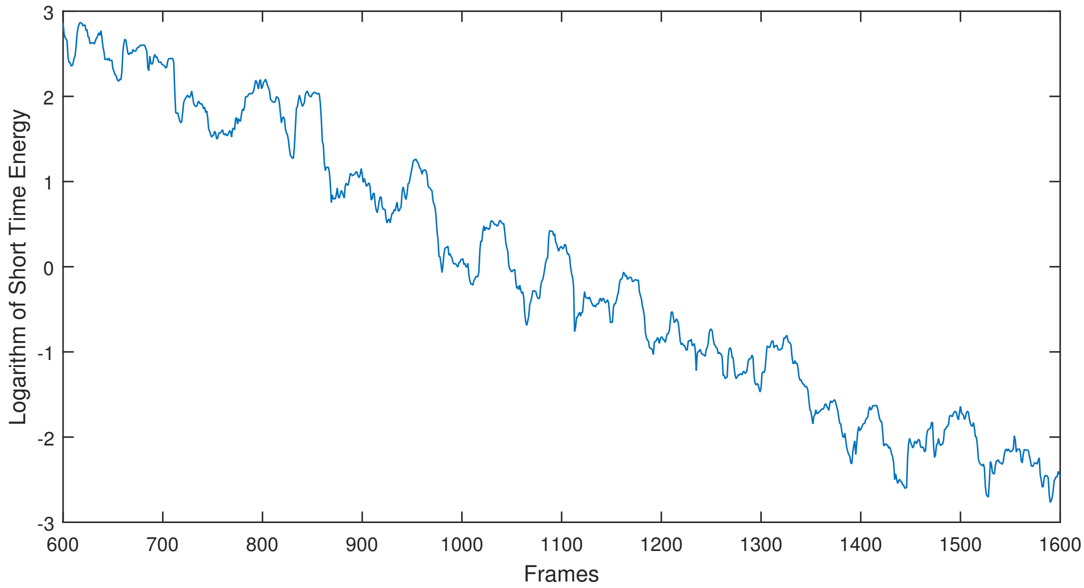

Figure 3: energy decay curve of the impulse sound.

Algorithm 1 feature extraction method

1: **Input:** Signal $x(n), n = 1, 2, \cdots, N$, frame size N_1, time shift N_2
2: **Initialize:** Frame the signal with $x_f(h, l) = x(hN_1 + l), h = 0, 2, \cdots, H, H = floor[(N - N_1)/N_2], l = 1, 2, \cdots, N_1$
3: **Output:** C
4: **for** $h = 0$ to H **do**
5: Calculate the short time energy of each frame:
$$\mathbf{E}^{(h)} = \sum_{l=1}^{N_1} \left(\mathbf{x}_f(h, l)^2\right)$$
6: Calculate the logarithmic EDC
$$\mathbf{X}^{(h)} = \log(\mathbf{E}^{(h)})$$
7: Perform DCT on the logarithmic EDC
$$\mathbf{X}_{\log}^{(h)}(i) = \sqrt{\tfrac{2}{N}} \sum_{k=1}^{K} \mathbf{X}^{(h)}(k) \cos\left(k - \tfrac{1}{2}\right) \tfrac{i\pi}{K}$$
8: Reserve first 5 DCT coefficients as the final features
9: **end for**

3. Experimental Results and Comparison

To test the algorithm, measured sound signals are collected at a building site. The database contains 2 different classes: electrical drill and banging sound. The sample rate is set as 22,050 Hz. The frame length is 256 data point. We employ the short-time energy for beginning point detection and intercept the following 0.2 s as a sample. For the electrical drill sound, we detect the beginning of the signal and then segment it with 0.2 s with non-overlapping. Finally 757 and 115 training samples, 106 and 16 testing samples, are collected for electrical drill sound and banging sound respectively. Since the last time of the two kinds of sounds is different, the data length of the two classes is imbalanced.

The ratio between the number of electrical drill sound and banging sound samples is almost 6.6. This also brings difficulty for the classification because generally the classifier intends to classify the new samples into electrical drill sound if the features lack of distinguishing ability.

We use MFCC as the contrast feature. Support vector machine(SVM)[5] is applied as classifier. The classification accuracy is defined as the ratio of the number of correctly classified samples to the number of total testing samples. In all of the training and testing conditions of both MFCC and the proposed feature, the parameters of SVM are set to obtain the best performance in the experiments. The results are shown in Tab.1. It can be found that the proposed feature outperform the MFCC.

Table 1: Comparison of Classification Accuracy(%) with MFCC and Proposed Feature

Method	MFCC	Proposed feature
electrical drill	97.31	**98.11**
banging	79.51	**93.75**

We show the ROC(Receiver Operating Characteristics) curve in Fig.4 to evaluate the classification performance of the proposed feature and MFCC. Fig.4 shows that the proposed feature can achieve higher AUC(Area under ROC Curve) value than MFCC.

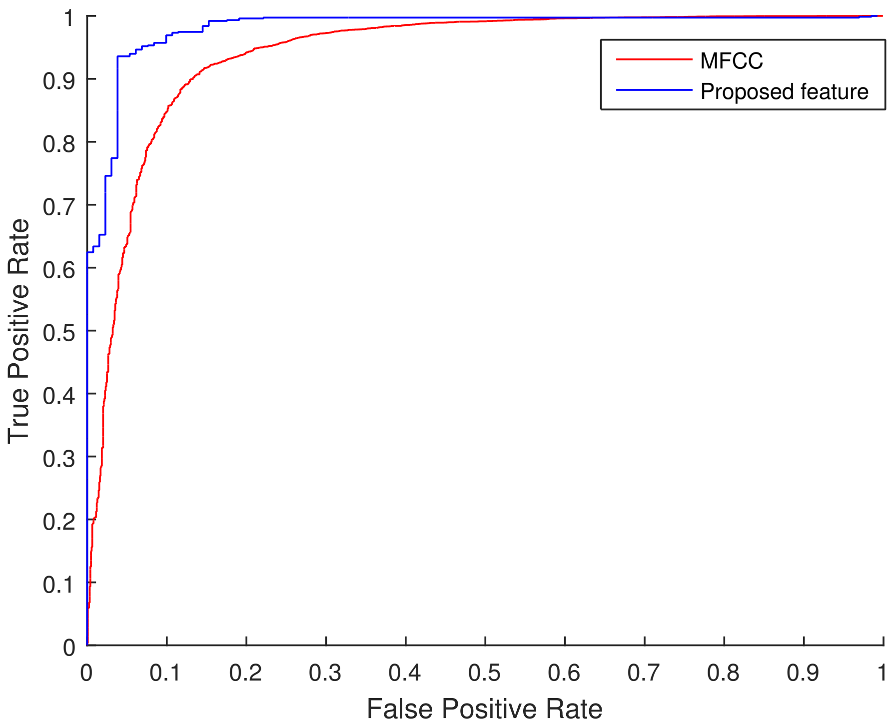

Figure 4: ROC curve of MFCC and Proposed feature.

We also test the performance of the proposed feature and MFCC under different SNR conditions. SNR is set from -5 dB to 10 dB with 5 dB as the step. Fig.5 shows that the performance of the proposed feature is stable and outperform MFCC in all testing conditions.

4. Conclusion

In this paper we proposed a short time energy based feature extraction method for the wall impact noise. Different from the traditional spectrum based features, this feature can catch the time variant

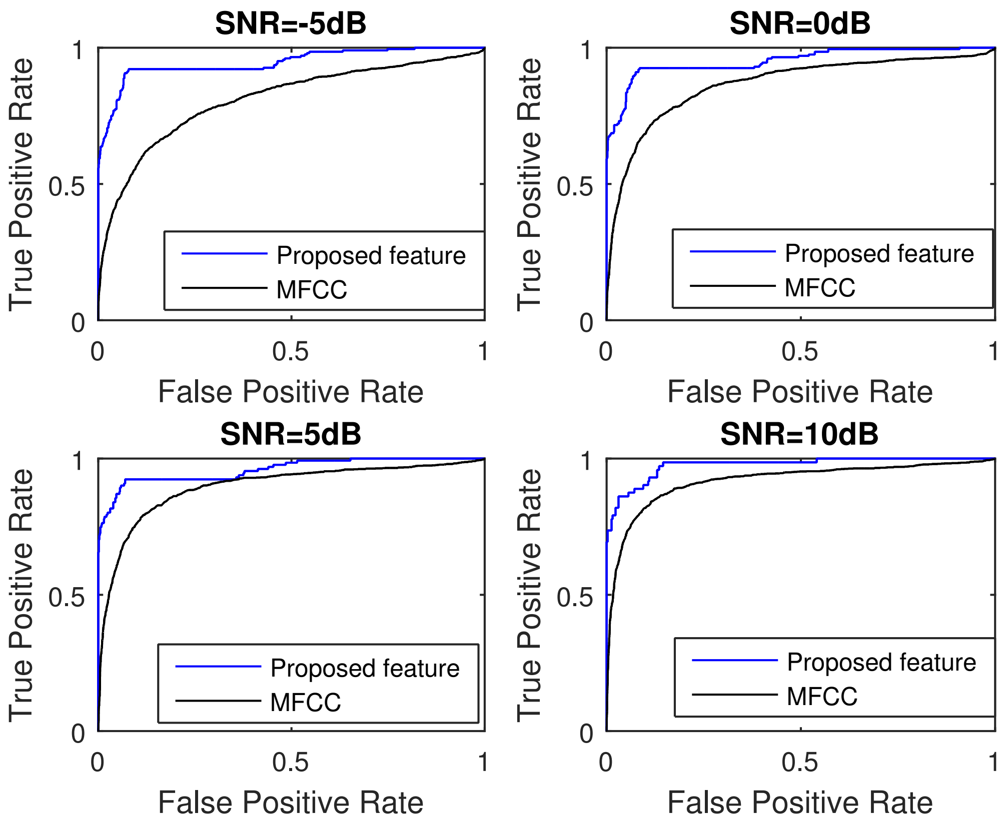

Figure 5: ROC curve under different SNRs.

characteristics of the noise signals and can be applied to recognize two different kinds of noises: electrical drill and banging noises. Experimental results show that the proposed feature is more robust than the traditional MFCC, which indicates that the temporal features has potential in recognize time-variant noises.

REFERENCES

1. Ntalampiras, S. A novel holistic modeling approach for generalized sound recognition, *Signal Processing Letters, IEEE*, **20** (2), 185–188, (2013).

2. Nakagawa, S., Wang, L. and Ohtsuka, S. Speaker identification and verification by combining MFCC and phase information, *Audio, Speech, and Language Processing, IEEE Transactions on*, **20** (4), 1085–1095, (2012).

3. Hung, W. W. and Wang, H. C. On the use of weighted filter bank analysis for the derivation of robust mfccs, *Signal Processing Letters, IEEE*, **8** (3), 70–73, (2001).

4. Mak, B. A mathematical relationship between full-band and multiband mel-frequency cepstral coefficients, *Signal Processing Letters, IEEE*, **9** (8), 241–244, (2002).

5. Chang, C. C. and Lin, C. J. LIBSVM: A library for support vector machines, *ACM Trans. Intell. Syst. Technol.*, **2** (3), 1–27, (2011).