Music Instrument Classification using Nontonal MFCC

Yi Wu, Qi Wang and Ruolun Liu*

Digital Audio Lab, Shandong University at Weihai, Shandong, China

*Corresponding Author: ruolun.liu@sdu.edu.cn

Keywords: Timbre recognition, nontonal MFCC, instrument classification.

Abstract. Combined with the sounding mechanism and cepstrum, a new model is proposed to describe the timbre more precisely, together with the nontonal Mel-frequency cepstral coefficients (NMFCC) derived from the nontonal spectral content which relates closely to the resonator. A better performance is observed from the experiment results of five classifiers over the isolated instrument samples of 13 instruments of different instrument families. The NMFCC method outperforms the one using MFCC with an accuracy rate of 97.7% for individual instruments.

Introduction

Music instrument classification is a challenging problem in many timbre related tasks. Timbre is the attribute of auditory sensation in terms of which a listener can judge that two sounds having the same loudness and pitch are dissimilar. Kostek [1] has studied the classification of 12 instruments with MPEG-7 descriptors and features based on the wavelet transform. Based on the source-filter model, Dubnov [2] studied the importance of excitation by means of Higher Order Statistics (HOS). Constant Q transform is used by Brown et al. [3] to identify four instruments of the woodwind family. Another study on taxonomic feature analysis for recognition of classical instruments was described by Deng et al. [4]. Though the MFCC of music sound have been proved helpful more than once in the instrument timbre classification [6-8], little has been progressed in improving the accuracy rate [8,9]. In general, most classifiers are established on a highly redundant feature set [2,4] instead of an individual feature.

In this paper, instrument timbre analysis is investigated for the western instruments. The significant factors in the timbre identification process are re-examined by the nontonal spectral content of the samples. Based on this nontonal spectrum, a new analytical model is proposed, together with a new timbre feature, nontonal MFCC (NMFCC). Compared with MFCC, the NMFCC can distinguish the timbre difference more clearly, and shows a better performance in the music instrument classification.

Sounding Models of the Musical Instrument

The framework of source-filter model originated from speech production, has been used for decades in speech coding and synthesis [10]. Similar to the speech signal, the music signal can also be modeled as a convolution between excitation and resonator. So in the frequency domain, the spectrum of music sound, $X(\omega)$, is the product between the spectra of excitation and resonator, $P(\omega)$ and $H(\omega)$,

$$X(\omega) = P(\omega)H(\omega). \tag{1}$$

In general, each instrument has its unique timbre because of its special shape and material, and these can significantly affect the frequency response of the resonator. In the frequency domain, the spectral envelope is an approximated description of $H(\omega)$, so the timbre feature is often obtained through extracting the shape of the spectral envelope.

According to the "partial+noise" model, music sound

$$x[n] = \sum_{r=1}^{R} A_r[n]\cos(\theta_r[n]) + e[n] \tag{2}$$

where $A_r[n]$ and $\theta_r[n]$ are the instantaneous amplitude and the phase of the r-th sinusoid respectively, and $e[n]$ is the noise component. The model assumes that the sinusoids are the stable partials of the sound, which is the tonal part. The remaining "residual" is modeled as a filtered white noise of a time varying filter [11], which is less reasonable in term of the sounding mechanism of the most instruments.

The proposed synthetical sounding model combines the source-filter model with the "partial+noise" model where the excitation is divided into sinusoids and residual components,

$$x[n] = \{\sum_{r=1}^{R} B_r[n]\cos(\varphi_r[n]) + u[n]\} * h[n] \tag{3}$$

where $B_r[n]$ and $\varphi_r[n]$ are the amplitude and phase of the r-th sinusoid in the excitation, $u[n]$ is the white noise, and $h[n]$ is the impulse response of the resonator. In other words, the music sound is divided into two parts, tonal part and nontonal part. Each part is produced by its own excitation, but the resonator is the same one.

According to the source-filter model, the envelope of the instrument sound spectrum is not exactly the representation of resonator frequency response. Some vibration experiments of instrument show that the excitation also has its own characteristic spectral envelope [12], which affects the shape of the estimated resonator spectral envelope to some extent. As Fig. 1 shows, the spectral envolpoe shape of the tonal part has big difference with the resonator specrtral envelope, while the outline of nontonal spectral is close to the resonator's one. It is more appropriate to capture the timbre information of resonator through nontonal spectral envolope.

Feature Extraction and Evaluation

Kaminskyj and Nielsen [6,7] have used MFCC in a broad series of instrument classification studies. It was demonstrated that, the MFCC based feature scheme gives the best classification performance [4]. The MFCC provides a rather compact representation of the spectral envelope and is more musically meaningful than other representations. In order to extract the resonator part from sound spectrum completely, the NMFCC is proposed based on the above synthetical model. The NMFCC extraction can be simply done by replacing amplitude spectrum in MFCC calculation by the nontonal amplitude spectrum (NAS) which is a new decisive factor in the perceptual instruments recognition.

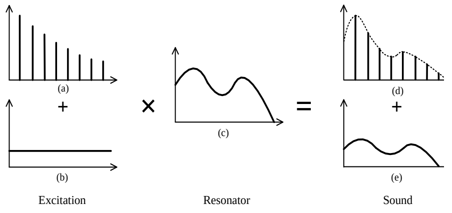

Fig.1 Synthetical model of instrument sounding.

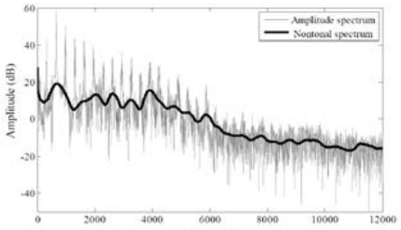

Fig.2 Orignal and nontonal spectra of violin E4.

Fragoulis [13] also verified that NAS contains more timbre information than tonal part does. The nontonal content extraction first spots each partial within a region that covers the whole harmonic spectral peak. The region width of partial's lobe is defined as 10% of the pitch frequency. By zeroing all these spectral regions, the tonal spectral lobes are eliminated. Next, interpolation and smoothing will be conducted in the residual part. In this way, the NAS curve is obtained. Fig. 2 shows the original spectrum and its nontonal part of a violin sound of E4. Compared with the original spectrum outline, several new peaks appear in the NAS curve, especially the ones located between 2000 - 4000 Hz. It is obvious that the peak located around 4000 Hz are more significant in the NAS, which is often seen in the formant of voice.

In order to show the advantages of NMFCC, the traditional MFCC is also used in the same classifiers. Fig.3 shows the MFCC and NMFCC of 4 instruments from 3 families of string, woodwind, and brass. Every instrument contains 37 samples of different pitches. Both MFCCs and NMFCCs are calculated frame by frame and an average is obtained within one sample, using

40-channel Mel filters. Only the first 12 coefficients are considered. A good feature should show significant difference between different instruments while keep stable in one instrument. As it can be seen clearly, the NMFCC tends to better satisfy the both requirements than MFCC does. It's also worth pointing out that the timbre of violin and viola are quite similar, even an experienced auditor cannot distinguish them from each other at certain pitches. However, the NMFCCs of the two instruments differ from each other obviously. The MFCC variances are 40.2, 41.7, 75.0, and 52.4 for violin, viola, flute, and trumpet in the top row of Fig. 3, whereas in the 2nd row, the NMFCC variances drop to 13.4, 10.7, 15.2, and 6.2 for the same instruments respectively.

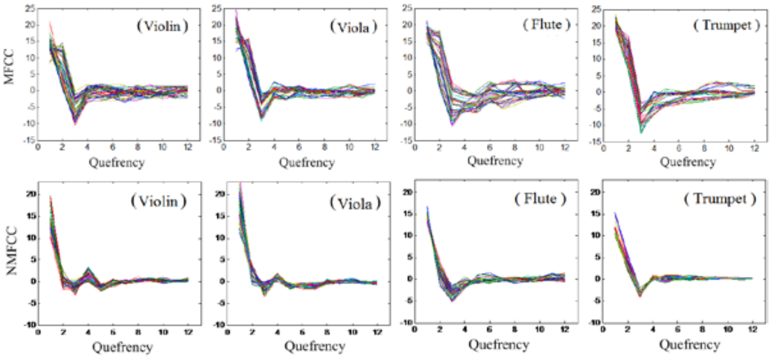

Fig.3 MFCCs and NMFCCs for violin, viola, flute and trumpet

Classification Experiments

The dataset is taken from the MUMS (McGill University Master Samples). All the notes are recorded using WAV format with the sampling frequency of 44.1 kHz. The experiments are conducted on 513 monophonic isolated notes of 13 instruments of string, brass, and woodwind families as indicated by Table I. The classifiers are established with leave-one-out cross validation**.**

Table 1. Music instruments for the classification experiments

Family	Instrument
String	Violin, Viola, Cello, Double bass, Piano, Guitar
Woodwind	Flute, Cor anglais, Oboe, Basson
Brass	Trumpet, Horn, Trombone

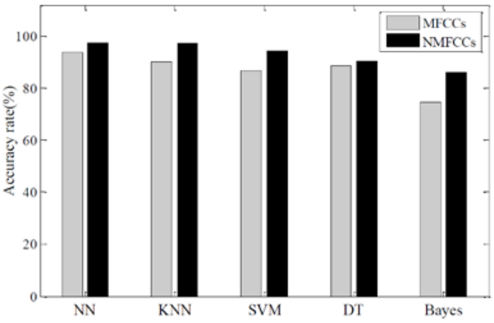

Fig. 4 Accuracy rate for different classifiers

The Naive Bayes classifier gives the worst result since tones from different instruments are not completely independent. Whereas the neural network (NN) classifier has the highest accuracy rate of 97.7%. The experiments are implemented in two hierarchies. At the top level of the taxonomy, instruments were divided into families; at the bottom level, the individual instruments are recognized. Fig. 4 illustrates the overall results in terms of accuracy rate of both MFCC and NMFCC, where the accuracy rates are improved by 4.0%, 6.9%, 7.4%, 1.9%, 11.3% with NN, KNN, SVM, DT and Naive Bayes classifiers respectively. The averaged accuracy rate over the five classifiers reaches to 93.2% of NMFCC from the 86.9% of MFCC.

Conclusion

This paper presents a new feature named by NMFCC that reflects better the timbre contribution from resonator. The feature comes from a synthetical model which combines the instrument sounding mechanism with the human hearing process. Through the massive experiments, the NMFCC shows significant disparity for different instruments while keeping stable for one instrument. Higher accuracy rates are also observed in the experiments with five different classifiers.

Though the experiment evaluations show evidences that the NMFCC is a promising feature in the music instrument timbre analysis, there are still some interesting issues to be further studied. For instance, there are different articulation styles in playing on any one instrument. Whether the NMFCC can be applied in more broad fields like instrument sound synthesizing or morphing, is also a valuable research subject.

Acknowledgements

This work is financially supported by the Shanghai Key Lab of Information Security Management Technology Fund (AGK201608) and Shandong Nature Science Foundation (ZR2016FM44).

References

[1] B. Kostek, Musical instrument classification and duet analysis employing music information retrieval techniques, Proc. IEEE, vol. 92, no. 4, pp. 712–729, 2004.

[2] S. Dubnov and A. Ben-Shalom, Review of ICA and HOS methods for retrieval of natural sounds and sound effects, in Proc. ICA, 2003.

[3] J. C. Brown, O. Houix, and S. McAdams, Feature dependence in the automatic identification of musical woodwind instruments, J. Acoust. Soc. Amer., vol. 109, no. 3, pp. 1064–1072, 2001.

[4] J. Deng, C. Simmermacher, and S. Cranefield, A study on feature analysis for musical instrument classification, IEEE Trans. Syst., Man, Cybern. B, vol. 38, no. 2, pp. 429–438, 2008.

[5] D. Bhalke, C. Rao, and D. Bormane, Stringed musical instrument recognition using fractional fourier transform and linear discriminant analysis, in Proc. ICICT, 2014, pp. 647–651.

[6] I. Kaminskyj and T. Czaszejko, Automatic recognition of isolated monophonic musical instrument sounds using knnc, J. Intell. Inf. Syst., vol. 24, no. 2/3, pp. 199–221, 2005.

[7] A. Nielsen, S. Sigurdsson, L. Hansen, and J. Arenas-Garcia, On the relevance of spectral features for instrument classification, in Proc. ICASSP, vol. 2, 2007, pp. 485–488.

[8] Barbedo J G A, Tzanetakis G. Musical instrument classification using individual partials [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2011, 19(1): 111-122.

[9] Newton M J, Smith L S. A neurally inspired musical instrument classification system based upon the sound onset [J]. The Journal of the Acoustical Society of America, 2012, 131(6): 4785-4798.

[10] Q. Fu and P. Murphy, Robust glottal source estimation based on joint source-filter model optimization, IEEE Trans. Speech Audio Process., vol. 14, no. 2, pp. 492–501, 2006.

[11] X. Serra, *Musical signal processing*, 1997, Chapter: Musical sound modeling with sinusoids plus noise, pp. 91–122.

[12] E. Bavu, J. Smith, and J. Wolfe, Torsional waves in a bowed string, Acta Acustica united with Acustica, vol. 91, no. 2, pp. 241–246, 2005.

[13] D. Fragoulis, C. Papaodysseus, M. Exarhos, G. Roussopoulos, T. Panagopoulos, and D. Kamarotos, Automated classification of piano-guitar notes, IEEE Trans. Speech Audio Process., vol. 14, no. 3, pp.1040–1050, May 2006.