American Journal of Networks and Communications
2016; 5(2): 11-16
http://www.sciencepublishinggroup.com/j/ajnc
doi: 10.11648/j.ajnc.20160502.11
ISSN: 2326-893X (Print); ISSN: 2326-8964 (Online)

Tone Quality Recognition of Instruments Based on Multi-feature Fusion of Music Signal

Zhe Lei[1], Mengying Ding[1], Xiaohong Guan[1], Youtian Du[1, *], Jicheng Feng[2], Qinping Gao[2], Zheng Liu[2]

[1]The School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China

[2]Xi'an Conservatory of Music, Xi'an, China

Email address:

lzleizhe@163.com (Zhe Lei), corianderherb@foxmail.com (Mengying Ding), xhguan@mail.xjtu.edu.cn (Xiaohong Guan), duyt@mail.xjtu.edu.cn (Youtian Du), fjc-fagott@hotmail.com (Jicheng Feng), gaoqinping@yahoo.com.cn (Qinping Gao), composerliu@hotmail.com (Zheng Liu)

[*]Corresponding author

To cite this article:

Zhe Lei, Mengying Ding, Xiaohong Guan, Youtian Du, Jicheng Feng, Qinping Gao, Zheng Liu. Tone Quality Recognition of Instruments Based on Multi-feature Fusion of Music Signal. *American Journal of Networks and Communications.* Vol. 5, No. 2, 2016, pp. 11-16.
doi: 10.11648/j.ajnc.20160502.11

Received: March 21, 2016; **Accepted**: March 31, 2016; **Published**: April 19, 2016

Abstract: The traditional expert-based instrumental music evaluation strategy can't meet the requirements of the rapidly accumulated audio data. The traditional strategy not only takes a high cost of human's energy and time but also may have some problems on consistency and fairness of judgment. This paper aims at designing a complete recognition and evaluation strategy to automatically identify the timber of wind instruments. We take the clarinet as example and propose a strategy based on multi-feature fusion and random forest. First, we use the identification of fundamental frequency algorithm to automatically distinguish the notes performed by the instruments. Second, we extract 3 types of features including MFCC, brightness and roughness to describe the instrumental signals. Then, considering two kinds of variants: note and tone quality, we design 5 strategies to remove the influence of different notes in the evaluation of tone quality. By analyzing these strategies, we explore the optimal strategy for the recognition. The final evaluation results over 840 music slices demonstrate the effectiveness of this method.

Keywords: Tone Quality, Timbre Analysis, Audio Signal Processing, Random Forest

1. Introduction

With the rapid development of digital music processing and storage technologies, the audio data accumulate continuously. Many art colleges have amassed vast amount of instrumental music data. At present, the evaluation of instrumental music is generally based on experts. However, the expert-based evaluation strategy not only takes a high cost of human's energy and time but also may have some problem on consistency and fairness of judgment. Hence, using information processing method to help evaluate the tone quality of instrumental music has great theoretical significance and practical values.

Now, the research on instrumental analysis and application mainly focuses on the fields including instrument category recognition, musical information retrieval, computer-assisted music understanding and instrumental tone quality analysis [1]. It is hard to retrieve the common digital music contents because they contain huge amounts of time-series data which have no well-defined semantic and structural style [2]. Some recent researches introduced the indexing and retrieval of instrumental tone quality based on distance matching technology [3]. Some researchers studied more common chords music in the real music environment to extend the previous research on the evaluation of single channel tone quality. They extracted the tone quality features and employed new data mining algorithms to identify and retrieve the interesting objects in massive audio files [4]. Computer-aided music understanding mainly refers to using computer technology and other relevant information technologies to

analyze music structures and contents [5]. The work can make computers automatically analyze the music structures, and thereby reduces the burden on related experts. Computer-aided music understanding is mainly applied in inferring musical form, music genres, music styles and music types [6]. Guo et al. presented a new approach to the instrumental tone quality analysis that can evaluate the level of the performer by instrumental tone quality recognition [7].

This paper focuses on the analysis of tone quality of wind instruments, and evaluates the quality level of wind instruments through the recognition results of tone quality. We take clarinets as example and propose an identifying and evaluating method of tone quality based on multi-feature fusion. The proposed method includes 3 steps: 1) Note recognition. We first distinguish the notes performed by instruments based on the identification of fundamental frequency. 2) Multi-feature extraction. We extract 3 typical features including MFCC, roughness and brightness with good identifying and evaluating performance. 3) We use random forest algorithm to construct the basic classifier and then design 5 strategies, at last we analyze and evaluate these strategies. In summary, the main contribution of this paper is to propose an identifying and evaluating method of tone quality of wind instruments, analyze and compare 5 strategies, explore the influence on tone quality analysis raising by note difference from 5 types of strategies and give the optimal recognition strategy. At last, we do the experiment in the complete audio signals by using the optimal strategy.

2. Methodology

2.1. Scheme Description

In this paper we choose clarinet as example and propose a scheme of instrument tone quality recognition shown in Fig. 1.

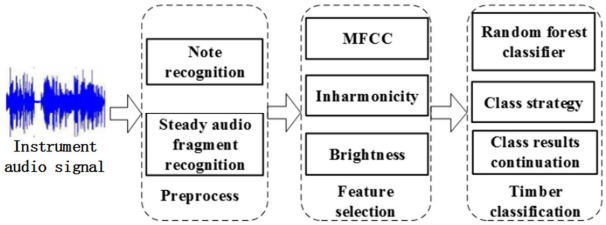


Figure 1. *Recognition strategy of instrumental tone quality.*

2.2. Note Recognition

Given an audio fragment $X=\{x_m \mid m=1,2,\cdots,M\}$, where x_m represents sampled point of the signal and M is the length. Different notes can show different characteristics which may influence the analysis results of tone quality, hence the note should be recognized before tone quality evaluation. In general, a note is denoted by note name and octave. In this paper, we take 28 different notes covering 4 octaves. Each octaves have 7 neighboring notes denoted by $\{c_i, c\#_i, d\#_i, f_i, g_i, a_i, a\#_i\}$ which correspond to $\{do, re, mi, fa, so, la, si\}$ respectively, where the letters and character '#' represent note name, and the subscript $i \in \{1,2,3,4\}$ represents the index of octaves.

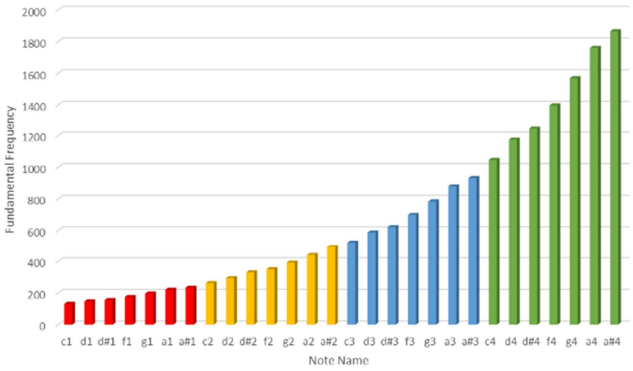


Figure 2. *Fundamental frequency of each note. The horizontal axis represents 28 different notes, the vertical axis represents the fundamental frequency, and the different colors indicate the different octaves.*

As shown in Fig. 2, different notes correspond to different fundamental frequency. We first use the method presented by Cheveigne et al. [8] to identify the fundamental frequency of notes, which is briefly described below:

1) Computing cumulative mean normalized difference function

$$d_t'(\tau)=\begin{cases}1 & ,if\ \tau=0\\ \dfrac{d_t(\tau)}{\sum_{j=1}^{\tau} d_t(j)/\tau} & ,otherwise\end{cases} \tag{1}$$

where

$$d_t(\tau)=\sum_{j=1}^{W}(x_j-x_{j+\tau})^2 \tag{2}$$

and searching for the values of τ to make the value of Eq. (2) zero.

2) Setting an absolute threshold and choosing the smallest value of τ that gives a minimum of d' deeper than that threshold.

3) Making parabolic interpolation. In the previous steps, if the period is not a multiple of the sampling period, the estimation may be incorrect by up to half the sampling period. Hence, this step is to reduce the estimation error by parabolic interpolation.

4) Best local estimation. For each time index *t*, we search for a minimum of $d_\theta'(T_\theta)$ for θ within a small interval $[t-T_{\max}/2, t+T_{\max}/2]$, where T_θ is the estimation at time θ and is the largest expected period. Generally $T_{\max}$ =25ms. The fundamental frequency of the audio signal is the inverse of its estimated period T.

We split the audio signal into multiple frames with the length of 882 points and the overlap of 50%. Every segment's fundamental frequency is estimated. When the ratio of the estimated pitches of two adjacent segments is between $2^{-1/24}$ and $2^{1/24}$, the two segments can be regarded as belonging to one note.

Generally, each note has three states: transient state, quasi-steady state and decay state. In our work, we only focus on the quasi-steady part because it is the most important component for tone quality evaluation and it excludes disturbances of transient state and decay state on feature extracting. To avoid the inaccuracy caused by transitional state between adjacent notes, we extract 80% in the middle of each note as the quasi-steady segment. As shown in Fig. 3, the segment between the two green dashed lines shows a complete note, and the segment between two red dashed lines denotes the quasi-steady state of this note.

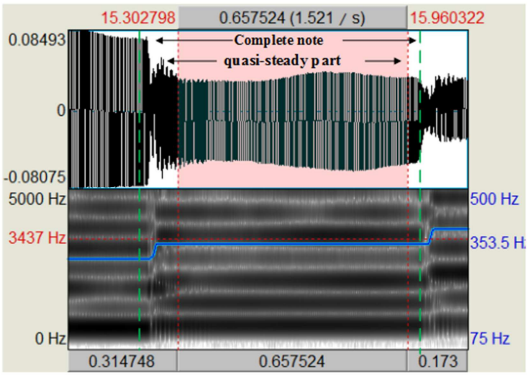


Figure 3. *The quasi-steady state of a note.*

2.3. Feature Extraction

The information in frequency domain of instrumental signal plays an important role in instrument tone quality analysis. This paper adopts 3 kinds of features to describe the instrumental signal, they are Mel-Frequency Cepstral Coefficients (MFCC), roughness and brightness.

2.3.1. MFCC Feature

In previous research of instrument type recognition, MFCC is a very typical feature, and it integrates the instrumental pronouncing mechanism and human auditory perception effectively [9]. Some research suggests that MFCC is one of the most outstanding property in the single-feature classification scheme [10]. Based on this intuition, we choose MCFF as features for the instrumental tone quality recognition. In this paper, we take 12 dimensional MFCC feature vector, which is denoted by F_{MFCC}.

2.3.2. Roughness Feature

Roughness is mainly used to describe the harmonious degree of a sound segment [11]. When two acoustic sources are with unequal frequency, the harmonic sound will make people feel turbid and harsh, and this feeling calls inharmony. Roughness is one of the features that can evaluate the quality of musical instruments. It is calculated through the dissonance of every two harmonic components of sound.

We apply Fourier transform to each note segment and extract the frequency and amplitude of each harmonic component and represent them by f_i and A_i respectively. Then, we calculate the dissonance of every two harmonic components and sum them over all the component pairs follows:

$$D_F = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} d\left(f_i, f_j, A_i, A_j\right) \tag{3}$$

where $d(f_1, f_2, A_1, A_2) = \min(A_1, A_2)[e^{-b_1 s(f_2 - f_1)} - e^{-b_2 s(f_2 - f_1)}]$, $s = \frac{x^*}{s_1 f_1 + s_2}$, and b_1, b_2, s_1, s_2 are coefficients that can be determined by experience, x^* is the point in which the roughness achieve the maximum value.

2.3.3. Brightness Feature

Brightness is another important audio feature that denotes the percentage of energy of spectrum segment in which the frequency is higher than the cut-off frequency f_k [12].

$$B = \frac{\sum_{n=k}^{N} x_n e^{-\frac{2\pi i}{N}kn}}{\sum_{n=0}^{N} x_n e^{-\frac{2\pi i}{N}kn}} \tag{4}$$

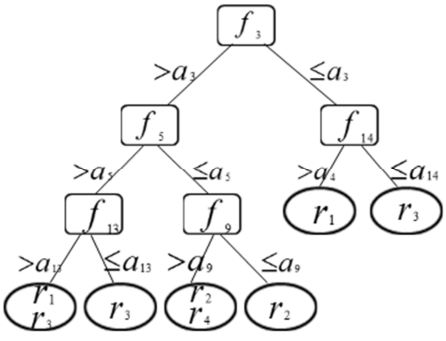
where k is the index of cut-off frequency.

2.4. Recognition Scheme Based on Random Forest

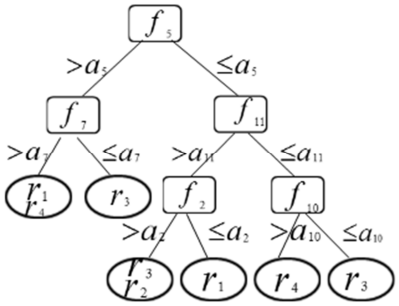
2.4.1. Random Forest

Each audio sample can be represented by a feature vector X based on the aforementioned feature extraction. Our work uses random forest as the basic classifier. Random forest is a kind of combined classifier, which combines Bootstrap method and Classification And Regression Tree (CART) algorithm together and constructs an ensemble of decision tree classifier model [13]. The classifier can be denoted by $\{h_1(X), h_2(X), \ldots h_N(X)\}$, and each component denotes a tree. The final classification result is determined by all of the decision trees in the ensemble.

In the stage of constructing decision trees, we first create a new data set by extracting t samples from training set using BootStrap method and selecting features f from the feature vector randomly, and then use CART algorithm to build a decision tree. Finally, by repeating these processes N times, we can obtain an ensemble of N decision trees $\{h_1(X), h_2(X), \ldots h_N(X)\}$. Fig. 4 shows an example of the achieved trees.



(a) h_1



(b) h_2

Figure 4. *An example of random forest.*

In this figure, f_i is the selected feature for the corresponding node and a_i is the threshold for f_i in the classification.

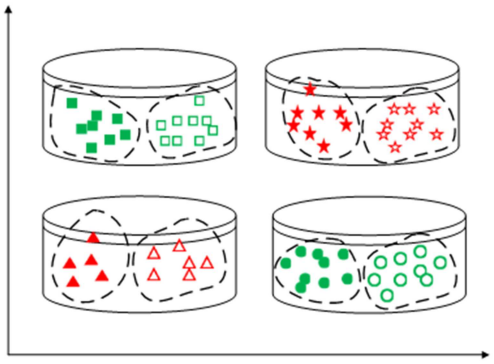
In the stage of classification, the category of each example is determined by voting of all the decision trees based on the following formula:

$$c_p = argmax_{r_i}\left(\frac{1}{N}\sum_{n=1}^{N} I\left(\frac{m_{h_n,r_i}}{m_{h_n}}\right)\right) \tag{5}$$

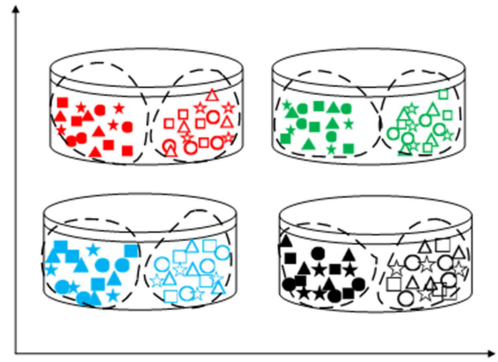
where N is the number of the decision trees, $I(\cdot)$ is the indicator function, m_{h_n,r_i} is the result of category r_i predicted by the tree h_n, m_{h_n} is the number of leaf nodes in the tree h_n.

2.4.2. Recognition Strategy

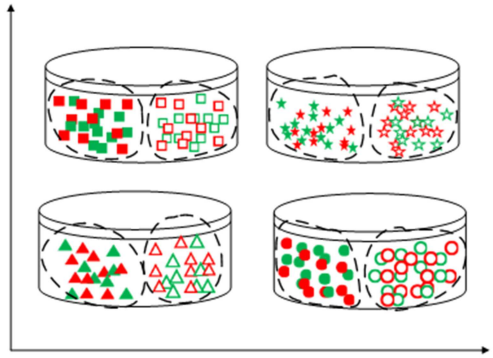
To recognize the level of tone quality, we focus on solving a classification problem and predicting that which category the audio signal is classified into. However, the variance of tones is raised not only by the tone quality, but also by the variance of notes. Moreover, the latter may lead to larger difference between two tone examples than the former in many cases. The difference among notes mainly reflects in the octave and note name. To analyze how the notes effect the results and achieve the evaluation results independent of the notes, we present and compare the following 5 different strategies. 1) All-notes classification strategy (AN): To build a single classification model over the whole examples of different tone quality and different notes. 2) Single-note strategy (SN): To build multiple classification models, each corresponding to the examples of the same note. 3) Equal-octave strategy (EO): To build multiple classification models, each corresponding to the examples of 7 notes in the same octave. 4) Homonymic-note strategy (HN): To build multiple classification models, each corresponding to the examples of the notes with the same note name. For example, $\{c_1, c_2, c_3, c_4\}$ are the homonymic notes and there is an octave between two neighboring notes. 5) Mixture-note strategy (MN): To build multiple classification models, each corresponding to the examples of the combination of some notes, for example $\{c_i, c\#_i,\}$, $\{d\#_i, f_i\}$, $\{g_i, a_i, a\#_i\}$ are three sample groups to build the models. Fig. 5 shows the latter four strategies.



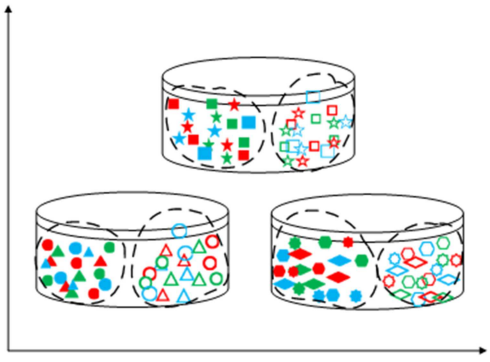
(a) SN



(b) EO



(c) HN



(d) MN

Figure 5. *The illustration of four classification strategies: (a) SN, (b) EO, (c) HN and (d) MN. Different shapes of points represent different note name and different colors represent different octave. Each cylinder represents a classifier, and each closed, dashed curve corresponds to a class, i.e., a level of tone quality.*

In the practical environment, we first split a continuous music signal into multiple fragments by note recognition and each fragment corresponds to a note. Then we use above recognition algorithm to handle each fragment. Finally, we fuse the recognition result of each fragment to achieve the final result. Suppose the continuous audio signal s_l can be split into k fragments denoted by $\{x_{l_1}, x_{l_2}, \cdots, x_{l_K}\}$. Based on the recognition algorithm, we can get the recognition matrix $P = [p_{ki}]_{K \times C}$, where p_{ki} denotes the probability of sample x_{l_k} belonging to class r_i:

$$\Pr(x_{l_k} \mid r_i) = \frac{1}{N} \sum_{n=1}^{N} I\left(\frac{m_{h_n, r_i}}{m_{h_n}} \right) \tag{6}$$

Like formula (5), we adopt the voting strategy to get the final evaluation result for the continuous music signal based on the following form:

$$c^* = argmax_{r_i} \left(\sum_{k=1}^{K} \Pr(x_{l_k} \mid r_i) \right) \tag{7}$$

Furthermore, the tone quality of an instrument generally does not just locate on the predefined discrete rank of quality. Therefore, we can evaluate the tone quality of an instrument with the continuous recognition result instead of discrete one shown in Eq. 7 by using the average value of a probability distribute overall the ranks. Given C ranks of tone quality named $r_1, r_2, \cdots, r_C$ in order of increasing tone quality, the continuous recognition result can be achieved by:

$$c_{co}^* = \frac{1}{K} \sum_{k=1}^{K} \sum_{i=1}^{C} i \cdot \Pr(x_{l_k} \mid r_i) \tag{8}$$

3. Experimental Result

3.1. Dataset and Experiment Parameter Selection

In this paper, we select 4 ranks of clarinets with different tone quality, and number them 1,2,3,4 in the order of increasing tone quality. 28 different notes in 4 neighboring octaves are performed, each note is repeated 30 times, and then 840 examples are obtained. According to the feature extraction method in 2.4, we extract features for each note sample. Then the dataset is split randomly into two sets, 70% is for training and the rest for test. In the experiment, we use 100 decision trees to build the random forest classifier, in each tree we choose 5 features and the depth of 4.

3.2. Experimental Result

To evaluate the classification results, we use the correct classification rate defined as follows:

$$R = \frac{CN}{TN} \tag{9}$$

where CN is the number of examples that are correctly classified and TN is the total number of the examples of one rank.

Based on the method proposed above, we take 5 classification strategy to recognize the tone quality of instruments. Table 1 shows the correct classification rate in each strategy. Rank1, Rank2, Rank3, Rank4 represent the correct classification rate in each corresponding level of tone quantity and AVGR represents the average rate of all level of tone quantity.

Table 1. *The correct classification rate of five strategy.*

	AN	SN	EO	HN	MN
Rank1	76.91%	88.65%	76.06%	84.11%	79.80%
Rank2	69.63%	80.56%	66.44%	78.62%	78.23%
Rank3	69.21%	85.11%	70.99%	76.00%	73.33%
Rank4	70.26%	82.67%	59.60%	72.11%	64.32%
AVGR	71.64%	84.24%	71.30%	77.71%	73.61%

From table 1 we can see that single-note strategy (SN) gets the best result. In this strategy, the difference among notes have been removed before the samples are used to build the model, and the classifier only needs to consider the difference of tone quality for one note. That is, each group of examples only include the examples with the same note, and then it has lower within-class divergence and higher between-class divergence. Therefore, the classifier can achieve better performance in the recognition of tone quality. Comparing equal-octave strategy (EO) to homonymic-note strategy (HN), we find that the latter strategy has higher accuracy rate. This result illustrates that the samples with same note name have lower within-class divergence. Mixture-note strategy (MN) combine the note name and the octave, so we can infer that the divergence of the samples is between EO and HN strategy. The table also shows that the result of mixture-note strategy is between that of equal-octave strategy and homonymic-note strategy.

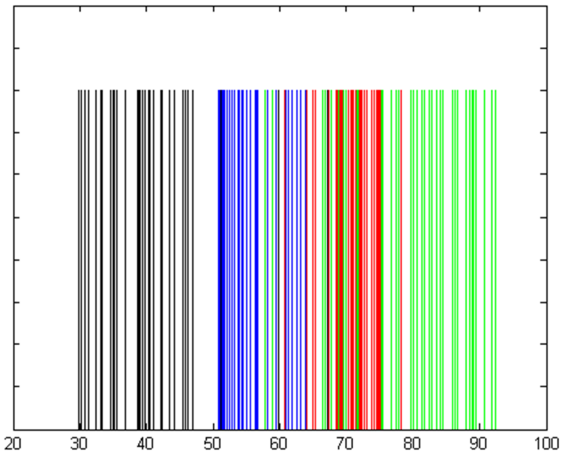


Figure 6. *The evaluation result based on the continuous music. The horizontal axis represents the continuous result of the music signal. The black, blue, red and green part respectively denote four different ranks of clarinets in order of increasing tone quality. The center position of each rank is artificially located at 40,55,70,85 for an appropriate show.*

We also evaluate the tone quality of four different rank of clarinets based on continuous music signal with Eq.8. Each music signal includes multiple different notes, and we fuse the evaluation for each notes together and achieve the final evaluation result of the instrument. The continuous evaluation result of each note is shown in Fig. 5. We notice that the all the classification results of different level of instruments distribute around the center of the corresponding category, and only a very low proportion of examples are misclassified into the other categories.

4. Conclusions

This paper presents an identifying and evaluating method of tone quality of wind instruments. In this method, we fuse 3 kinds of effective features and use the random forest as the basic classification. To remove the influence of different notes in the evaluation of tone quality, we introduce 5 recognition strategies and give the optimal classification strategy. The experimental result shows that our method can recognize the rank of wind instruments effectively. For future work, we will test the method on more wind instruments and try to explore more features to improve the accuracy.

Acknowledgements

This work is supported in part by the National Natural Science Foundation (6137540,61221063), Fundamental Research Funds for the Central Universities (xjj2013090) and 111 International Collaboration Program, of China.

References

[1] Yu-Hsiang H, Chao-Ton S. Multiclass MTS for saxophone timbre quality inspection using waveform-shape-based features [J]. IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics A Publication of the IEEE Systems Man & Cybernetics Society, 2009, 39(3): 690-704.

[2] Paulus J, Klapuri A. Music Structure Analysis Using a Probabilistic Fitness Measure and an Integrated Musicological Model. [C] ISMIR 2008, 9th International Conference on Music Information Retrieval, Drexel University, Philadelphia, PA, USA, September 14-18, 2008. 2008: 369-374.

[3] Typke R, Veltkamp R C, Wiering F. Searching notated polyphonic music using transportation distances [C]. Acm Multimedia Conference. 2004: 128-135.

[4] Downie S, Nelson M. Evaluation of a Simple and Effective Music Information Retrieval Method [C] Research & Development in Information Retrieval. SIGIR '00 Proceedings of the 23rd annual international ACM SIGIR conference on Research and developm, 2000: 73-80.

[5] K. Roger B. Dannenberg, Ning Hu. Pattern Discovery Techniques for Music Audio [J]. Journal of New Music Research, 2003, 32(2): 63-70.

[6] Yu Y, Zimmermann R, Wang Y, et al. Recognition and Summarization of Chord Progressions and Their Application to Music Information Retrieval[C]. Multimedia (ISM), 2012 IEEE International Symposium on. 2012: 9-16.

[7] Guo J, Ding M, Guan X, et al. Timbre identification of instrumental music via energy distribution modeling[C]. Proceedings of the 7th International Conference on Internet Multimedia Computing and Service. ACM, 2015: 1-5.

[8] Cheveigné A D, Kawahara H. YIN, a fundamental frequency estimator for speech and music [J]. Journal of the Acoustical Society of America, 2002, 111(4): 1917-30.

[9] Valero X, Alias F. Gammatone Cepstral Coefficients: Biologically Inspired Features for Non-Speech Audio Classification [J]. IEEE Transactions on Multimedia, 2012, 14(6): 1684-1689.

[10] Logan B. Mel Frequency Cepstral Coefficients for Music Modeling [C]. In International Symposium on Music Information Retrieval. 2000.

[11] Sethares W A. Tuning, Timbre, Spectrum, Scale [M]. Springer London, 2005.

[12] Vassilakis P N. Perceptual and Physical Properties of Amplitude Fluctuation and their Musical Significance [J]. Acta ibérica radiológica-cancerológica, 2001, 28(4): 119-128.

[13] Biau G. Analysis of a Random Forests Model [J]. Journal of Machine Learning Research, 2010, 13(2): 1063-1095.