MUSICAL INSTRUMENT IDENTIFICATION BASED ON HARMONIC TEMPORAL TIMBRE FEATURES

Jun Wu, Yu Kitano, Stanislaw Andrzej Raczynski, Shigeki Miyabe, Takuya Nishimoto, Nobutaka Ono and Shigeki Sagayama
The Graduate School of Information Science and Technology, University of Tokyo
Tokyo 113–8656, Japan
{wu,kitano,nishi,onono,sagayama}@hil.t.u-tokyo.ac.jp

ABSTRACT

The Music Instrument Identification research is an important and difficult problem in Music Information Retrieval (MIR). In this paper an algorithm based on flexible harmonic model is proposed to represent the pitch in music by Gaussian mixture structure. The proposed algorithm models each spectral envelope of underlying harmonic structure to approximate the real music and uses EM algorithm to estimate the parameters. Not only is it able to estimate the multipitch (F0) but it also takes the attack problem (a kind of inharmonic structure at the beginning of some pitches) into account. The proposed algorithm makes it possible to envisage the use of timbre features derived from both harmonic part and attack part. Musical instrument recognition is then carried out by using SVM classifier. Experiment shows high performance of the proposed algorithm for instrument identification task.

1. INTRODUCTION

Musical instrument identification task includes both estimation of music pitches and identification of each pitch to specific instrument. Although it has been considered as difficult problem, some approaches such as using Cepstral coefficient [1], Temporal features [2], Spectral features [3] to deal with single instrument identification have been developed. For more difficult problem which is to identify the multi-instrumental polyphonic music, some previous research has also been done such as: frequency component adaptation [4], missing feature theory [5], and feature weighting to minimize influence of sound overlaps [6]. However, all of these researches need given correct F0 as the prior knowledge while in real application the correct F0 is not given actually.

In our previous work, a generative modeling of harmonic sound for multipitch analysis called Harmonic-Temporal Clustering (HTC) [7] is developed. HTC decomposes the spectral energy of the signal in the time-frequency domain into acoustic events, which are modeled by using acoustic object models with a harmonic and temporal 2-dimensional structure. Unlike conventional

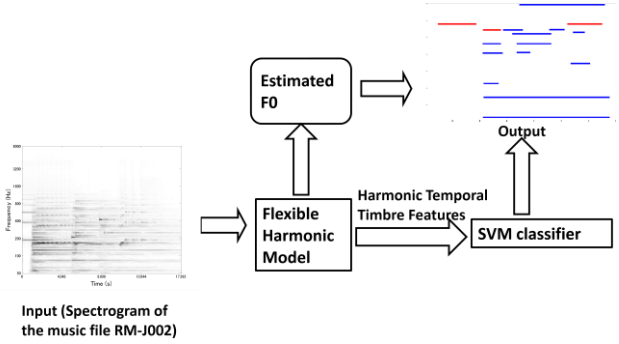

Figure 1. Flow chart of proposed system.

frame-wise approaches such as [8, 9], HTC deals with the harmonic and temporal structures in both time and frequency directions simultaneously. However, HTC was not able to deal with attack problem which widely exists in musical pitches.

In this paper, at first a flexible harmonic model capable of modeling both harmonic part and attack part of music is proposed to model the music pitches and estimate F0s. It uses Gaussian mixture structure to represent musical pitches and is able to estimate each mean parameter by EM algorithm from input musical signal. Then a new approach based on classifying each pitch into timbre categories according to their similarity with regard to the timbre features is proposed for musical instrument identification. The proposed algorithm can both estimates multiple pitches and identify the pitch to specific instrument. Therefore it will not need any given prior knowledge, which makes this new algorithm efficient for real application.

In Section 2, at first the proposed flexible harmonic model is introduced. After the F0s are estimated by the proposed algorithm, the Harmonic Temporal Timbre Energy Ratio (HTTER) and Harmonic Temporal Timbre Envelop Similarity (HTTES) features are also generated from the proposed model. It is used to construct SVM-based classifier for identifying each pitch to specific musical instrument. In Section 3, the experimental results are demonstrated. At last, the conclusion is made in section 4. The overall flowchart of the proposed system is illustrated in Figure 1. The output of the proposed system is

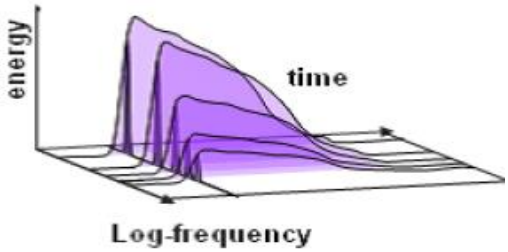

Figure 2. Profile of the *k*th pitch model

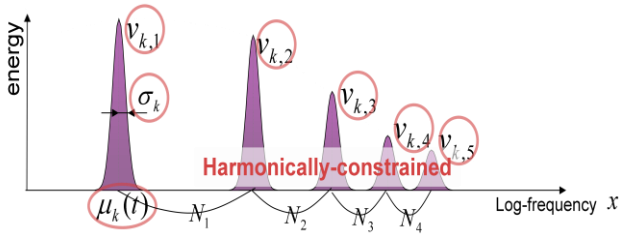

Figure 3. Cutting plane of $q_k(x,t;\theta)$ at time *t*.

the estimated multipitch of the musical signal and the different color represents different instrument.

2. HARMONIC TEMPORAL TIMBRE FEATURES FOR INSTUMENT IDENTIFICATION

2.1 Flexible Harmonic Model

In this section we discuss about how we build the flexible harmonic model from the observed power spectrogram series *W*(*x*;*t*) of input music signal, where *x* is log-frequency and *t* is time. The proposed model tries to approximate the power spectrogram by assuming it is the sum of *k* parametric models $q_k(x,t;\theta)$ (see Figure 2). $q_k(x,t;\theta)$ represents the *k*th pitch model in the music and θ represents the parameters in the model. One pitch model is composed of fundamental partial (F0) and N harmonic partials. The parameters of flexible harmonic model are represented in Table 1.

Given the pitch contour $\mu_k(t)$ in *k*th pitch model, the contour of the *n*th partial is $\mu_k(t) + log(n)$ (see Figure 3). The normalized energy density of the *n*th partial in the *k*th model can be assumed to be a multiplication of the power envelope of the *n*th partial $U_{k,n}(t)$ and the Gaussian distribution centered at $\mu_k(t) + \log(n)$, $U_{k,n}(t) \times \frac{v_{k,n}}{\sqrt{2\pi}\sigma_k} e^{-(x-\mu_k(t)-logn)^2/2\sigma_k^2} \quad n = 1,\dots,N, \quad (1)$

satisfying

$$\forall k, \sum_n v_{k,n} = 1.$$

Since we do not know in advance what the sources are, it is important to introduce a model as generic as possible for estimating the power envelope function. Therefore we should choose a function that is temporally continuous,

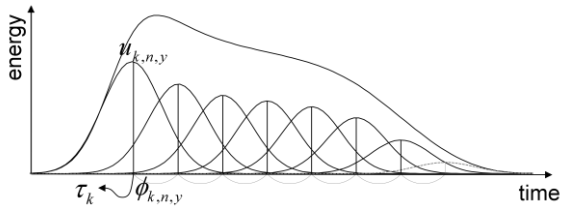

Figure 4. Power envelope function $U_{k,n}(t)$ at frequency *x*.

parameter	Physical meaning
$\mu_k(t)$	Pitch contour of the *k*th pitch
w_k	Energy of the *k*th pitch
$v_{k,n}$	Relative energy of *n*th partial in *k*th pitch
$u_{k,n,y}$	Coefficient of the power envelop function of *k*th model, *n*th partial, *y*th kernel
τ_k	Onset time
$Y\phi_{k,n,y}$	duration (Y is constant)
σ_k	Diffusion in the frequency direction of the harmonics
$\tilde{\mu}_j$	Mean of *j*th Gaussian in attack model
$\tilde{\sigma}_j^2$	Diffusion in the frequency direction of *j*th Gaussian in the attack model
$\tilde{\alpha}_j$	Coefficient of *j*th Gaussian distribution in attack model

Table 1. Parameters of flexible harmonic model

nonnegative, having a time spread from minus to plus infinity (assuming the Gabor-wavelet basis as the mother wavelet) and adaptable to various curves. Assume the spectra are obtained by the wavelet transform (constant transform) using Gabor wavelet basis function, the frequency spread of the wavelet power spectra is close to a Gaussian distribution. The assumption was justified based on the generalized Parseval's theorem in [7]. To come up with a function satisfying all these requirements, we let the frequency spread of each harmonic component be approximated by a Gaussian distribution function when the spectra are obtained by the wavelet transform (constant Q transform) using Gabor wavelet basis function. Denote $U_{k,n}(t)$ as the power envelope of the *n*th partial.

$$U_{k,n}(t) = \sum_{\forall y} \frac{u_{k,n,y}}{\sqrt{2\pi\phi_{k,n}{}^2}} exp\left\{-\frac{(t-\tau_k-y\phi_{k,n})^2}{2\phi_{k,n}{}^2}\right\} \quad (2)$$

τ_k is the center of the Gaussian, which is considered as an onset time estimate, $u_{k,n,y}$ is the weight parameter for each kernel, which allows the function to have variable shapes for each harmonic partial (see Figure 4). $u_{k,n,y}$ is defined as the coefficient of the power envelop function of *k*th model, *n*th partial, *y*th kernel. It should be normalized to satisfy $\forall k, \forall y: \sum_y u_{k,n,y}(x,t) = 1$.

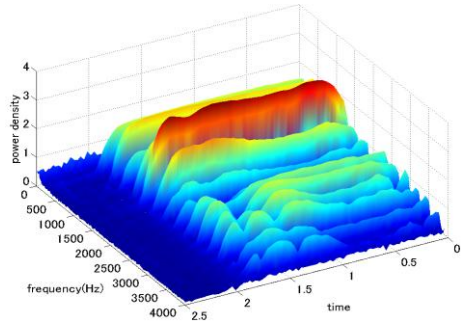

Figure 5. Power spectrogram of oboe sound.

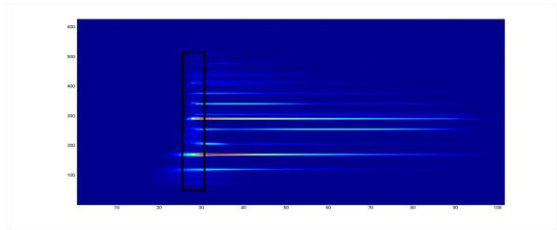

Figure 6. The spectrogram of attack in a piano pitch.

Figure 5 shows the spectrogram of oboe sound. Three axes are frequency, time and power density respectively. From the figure we can see that the envelope of each partial is different and has different information although there is also relationship between the partials. To approximate the envelop of each specific partial, the proposed model is actually estimating the parameters for each partial even in the same model $q_k(x,t;\theta)$.

The model $q_k(x,t;\theta)$ is expressed as a mixture of Gaussian mixture model (GMM) with constraints on the kernel distributions: supposing that there is harmonicity with *N* partials modeled in the frequency direction, and the power envelope is described using *Y* kernel distribution in the time direction. The model can be written in the form

$$q_k(x,t;\theta) = \sum_n \sum_y S_{k,n,y}(x,t;\theta) \quad (3)$$

And the Kernel distribution can be written in the form

$$S_{k,n,y}(x,t;\theta) = \frac{w_k v_{k,n} u_{k,n,y}}{2\pi\delta_k \emptyset_k} e^{-\frac{(x-\mu_k(t)-log n)^2}{2\sigma_k^2} - \frac{(t-\tau_k-y\emptyset_{k,n})^2}{2\phi_{k,n,y}^2}} \quad (4)$$

Therefore the model $q_k(x,t;\theta)$ is the mixture of Gaussian distribution $S_{k,n,y}(x,t;\theta)$. And the whole model is the mixture of the pitch model $q_k(x,t;\theta)$.

2.2 A New Model for Attack Problem

In this section we discuss about how we model attack of the harmonic instruments. The term *attack* is defined to represent the inharmonic phenomenon at the very beginning of some pitches played by harmonic instruments. In the attack part the harmonic structure appears slightly

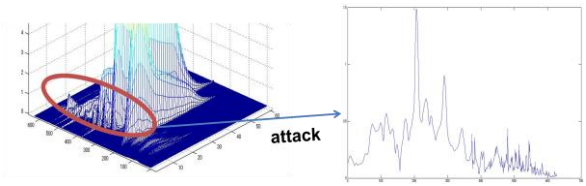

Figure 7. Power spectrum of attack in a piano pitch.

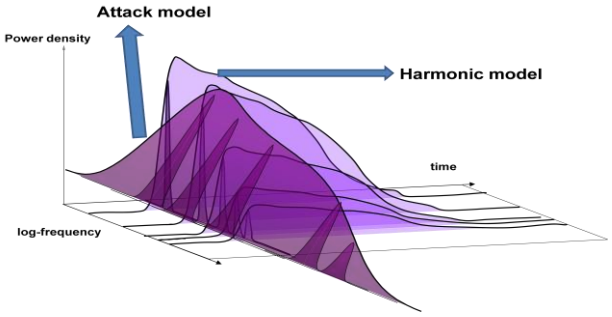

Figure 8. The representation of the proposed model.

unclearly. For example, Figure 6 is the spectrogram of a piano pitch, at the beginning of the pitch we can see the attack part which is indicated by rectangle. In attack part, the harmonic partial cannot be distinguished clearly like the pitch without attack, which makes the harmonic temporal modeling difficult.

To show it more clearly we draw the three dimensional power spectrum in the left part of Figure 7, which is not very harmonic model such as Figure 3. The power envelop of attack shown in the right part of Figure 7 is modeled by another Gaussian mixture model in frequency domain with the correlation with the harmonic part.

The attack model in time axes was represented by the following equation.

$$U'_{k,n}(t) = \frac{U_{k,n,0}}{\sqrt{2\pi}\emptyset_{k,n}} e^{-(t-y\emptyset_{k,n})^2/2\emptyset_{k,n}^2} \quad (5)$$

Therefore in time direction, it is modeled as a Gaussian distribution which is correlated with the harmonic part.

The attack model in frequency axes was represented by a Gaussian mixture model.

$$F(x) = \sum_{j=1}^{m} \alpha_j g(x,\mu_j,\sigma_j^2) \quad (6)$$

$g(x,\mu_j,\sigma_j^2)$ is a component Gaussian distribution characterized by means μ_j, covariance σ_j^2 and weight of its component distributions α_j. The parameters are updated by using EM algorithm in next section.

Therefore the whole proposed model was composed by the harmonic model part and attack model part, which is shown in Figure 8. The harmonic model part is same as Figure 2 while the attack model part is Gaussian mixture model in log-frequency direction.

2.3 Updating Equations Using EM Algorithm

The proposed method uses EM algorithm for the parameter estimation. We assume that the energy density

W(x;t) has an unknown fuzzy membership to the *k*th model, introduced as a spectral masking function $m_k(x,t)$. To minimize the difference between the observed power spectrogram time series *W*(*x*;*t*) and the pitch model $\sum_k q_k(x,t;\theta)$, we use the Kullback–Leibler (KL) divergence as the global cost function;

$$J=\sum_k \iint_D m_k(x,t)W(x;t)log\frac{m_k(x,t)W(x;t)}{q_k(x,t;\theta)}, \quad (7)$$

under the constraint;

$$\sum_k m_k\ (x,t)=1, 0<m_k\ (x,t)<1, \forall x, \forall t, \quad (8)$$

The problem is regarded as the minimization of (7).

The membership degree $m_k(x,t)$ (spectral masking function) of *k*th pitch model can be considered as the weight of the *k*th model in the whole spectrogram model. It is unknown at the beginning and need to be estimated. On the other hand, the spectrogram of the *k*th model is modeled by a function $q_k(x,t;\theta)$, where θ is also unknown. The proposed model is optimized by using EM algorithm, where the E-step updates $m_k(x,t)$ with θ fixed and the M-step updates θ with $m_k(x,t)$ fixed.

The *k*th model is composed of fundamental partial and harmonic partials. We use another masking function $m_{k,n,y}(x,t)$ that decomposes the *k*th partitioned cluster $m_k(x,t)W(x;t)$ into the {*n*,*y*}th subcluster. Therefore $m_{k,n,y}(x,t)$ can be considered to be the weight of each Gaussian distribution of the *k*th model. We apply the Jensen's inequality for the cost function and derive the following function:

$$J_k \triangleq \iint_D m_k(x,t)W(x,t)log\frac{m_k(x,t)W(x,t)}{\sum_{n,y}S_{k,n,y}(x,t;\theta)}dxdt \leqq J_k^+ \triangleq \sum_{n,y} \iint_D m_k(x,t)m_{k,n,y}(x,t)W(x,t)\,log\frac{m_k(x,t)m_{k,n,y}(x,t)W(x,t)}{S_{k,n,y}(x,t;\theta)}dxdt. \quad (9)$$

The equality holds when

$$m_{k,n,y}(x,t)=\frac{S_{k,n,y}(x,t;\theta)}{\sum_n\sum_y S_{k,n,y}(x,t;\theta)}, \quad (10)$$

satisfying the following conditions:

$$\sum_n\sum_y m_{k,n,y}\ (x,t)=1, \forall k$$

$$0<m_{k,n,y}(x,t)<1, \forall n, \forall y.$$

The E-step is realized by the following equation.

$$m_k(x,t)m_{k,n,y}(x,t)=\frac{S_{k,n,y}(x,t;\theta)}{\sum_k\sum_n\sum_y S_{k,n,y}(x,t;\theta)} \quad (11)$$

The M-step can be realized by the iteration of the update the parameters depending on each acoustic object

$$\begin{cases} a=\sum_n\sum_y \iint_D\ y(t-\tau_k)l_{k,n,y}(x,t)dxdt \\ \mathrm{b}=\sum_n\sum_y \iint_D\ (t-\tau_k)^2 l_{k,n,y}(x,t)dxdt \end{cases} \quad (12)$$

$$\phi_k^{(i)}=\frac{-a+\left(a^2+4b\omega_k^{(i)}\right)^{1/2}}{2\omega_k^{(i)}} \quad (13)$$

$$u_{k,n,y}^{(i)}=\frac{1}{d_u+w_k^{(i)}}\left(d_u\bar{u}_{k,y}+\iint_D\ l_{k,n,y}(x,t)dxdt\right) \quad (14)$$

$$\sigma_k^{(i)}= \frac{1}{w_k^{(i)}}\sum_{n,y}\iint_D\ \left(x-\mu_{k0}^{(i)}-\log n\right)^2 m_{k,n,y}^{(i)}(x,t)W(x,t) \quad (15)$$

$$\tau_k^{(i)}=\frac{1}{w_k^{(i)}}\sum_{n,y}\iint_D\ \left(t-y\phi_k^{(i-1)}\right)l_{k,n,y}^{(i)}(x,t)dxdt \quad (16)$$

$$v_{k,n}^{(i)}=\frac{1}{d_v+w_k^{(i)}}\left(d_v\bar{v}_n+\sum_y\iint_D\ l_{k,n,y}^{(i)}\ (x,t)dxdt\right) \quad (17)$$

$$w_k^{(i)}=\sum_{n,y}\iint_D\ l_{k,n,y}^{(i)}\ (x,t)dxdt \quad (18)$$

$$l_{k,n,y}^{(i)}(x,t)=m_k^{(i)}(x,t)m_{k,n,y}^{(i)}(x,t)W(x,t) \quad (19)$$

$$\beta_j(x)=\frac{\alpha_j g(x,\mu_j,\sigma_j^2)}{\sum_{j=1}^m \alpha_j g(x,\mu_j,\sigma_j^2)} \quad (20)$$

$$\tilde{\mu}_j=\frac{\int \beta_j(W(x;t))W(x;t)dx}{\int \beta_j(W(x;t))dx} \quad (21)$$

$$\tilde{\sigma}_j^2=\frac{\int \beta_j(W(x;t))(x-\tilde{\mu}_j)^2dx}{\int \beta_j(W(x;t))dx} \quad (22)$$

$$\tilde{\alpha}_j=\frac{\int \beta_j(W(x;t))dx}{\int xdx} \quad (23)$$

Since each step of this update rule can reduce the objective function (9) successfully, the iteration of these update steps can yield to locally optimal parameters.

2.4 Harmonic Temporal Timbre Features

In polyphonic music, different signals are very often overlapped so that the analysis and identification of each signal or each pitch are difficult. For solving this problem, we need to retrieve as much information from each signal or pitch as possible to find the specific instruments' patterns and identify them. The characteristic of instruments' spectral energy of each harmonic partial can be used for identifying specific instrument. There are many differences between the shapes in the spectrum of the harmonic partials, the temporal structure and the envelop similarity of the harmonics. Therefore we consider that the characteristic in timbre of specific instrument is derived from the difference of harmonic temporal timbre energy and harmonic temporal timbre envelope shape. The shapes of acoustic events classified into the same timbre category or same instrument should look alike regardless of the pitch, power, onset timing and duration. Besides the spectral envelope features such as σ_k and temporal features such as $Y\phi_{k,n,y}$ and $u_{k,n,y}$, we define the Harmonic Temporal Timbre Energy Ratio (HTTER) and Harmonic Temporal Timbre Envelop Similarity

(HTTES). HTTER defines the features of the energy ratio of the harmonic temporal timbres. HTTES defines the difference between the envelop shapes of the harmonic temporal timbres.

$$HTTER_{\mathrm{k,n,n'}} = \frac{\sum_y S_{k,n,y}(x,t;\theta)}{\sum_y S_{k,\mathrm{n'},y}(x,t;\theta)} \quad (24)$$

$$HTTES_{k,n,n'} = \int \left(U_{k,n}(t) log \frac{U_{k,n}(t)}{U_{k,n'}(t)} \right) dt + \int \left(U_{k,n'}(t) log \frac{U_{k,n'}(t)}{U_{k,n}(t)} \right) dt \quad (25)$$

3. EXPERIMENTS

To evaluate the proposed algorithm, we did the experiments with the music notes chosen from the RWC music database [11]. Since the RWC database also includes the MIDI files associated with each real-performed music signal data, we will evaluate the accuracy by comparing the estimated fundamental frequency and the MIDI files. The accuracy for instrument identification experiment is the multiplication of the accuracy for F0 estimation and the accuracy for identifying each pitch to corresponding instrument.

Using the corresponding MIDI data as references, the accuracy for instrument identification is computed by

$$\text{Accuracy} = \frac{\mathrm{X} - \mathrm{D} - \mathrm{I} - \mathrm{S}}{\mathrm{X}} * \text{Accuracy}_{instrument}$$

where

X is number of the total frames of the voiced part;

D is number of deletion errors;

I is number of insertion errors;

S is number of substitution errors.

$\text{Accuracy}_{\text{instrument}}$ is the accuracy for identifying each pitch to corresponding instrument by comparing with the corresponding MIDI data. The right part of Figure 9 showed the result of applying proposed algorithm for RM-J012 in RWC database [11]. In the estimated F0 the piano pitch is represented by blue lines while the flute pitch is represented by red lines. It was compared with the MIDI data in the left part of Figure 9 for calculating the accuracy. In MIDI figure, the piano part is represented by blue lines while the flute part is represented by yellow lines.

271 music signal pieces (including 6 instruments: 32 altosax pieces, 36 guitar pieces, 88 piano pieces, 45 violin pieces, 36 flute pieces and 34 oboe pieces) chosen from the RWC music database [11]. 70% of the signal pieces were selected randomly as the training data. Then the proposed model was applied to generate the training features. SVM classifier was generated from the training features. The testing data was selected randomly from the rest 30% music pieces and mixed randomly to generate new polyphonic signals.

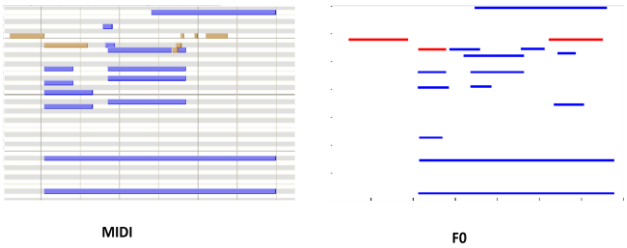

Figure 9. Corresponding MIDI file and the Estimated F0 for RM-J002 from RWC database

	2 instruments (%)	3 instruments (%)	4 instruments (%)
NMF	58.4	52.7	41.5
Proposed	74.8	60	50.7

Table 2. Recognition accuracy of NMF algorithm and the proposed algorithm

In Table 2, the proposed algorithm was compared with the NMF algorithm which is widely used by researchers for multipitch estimation and instrument identification. [12] [13] First, the F0 is estimated by using NMF pitch transcription algorithm. Therefore, each pitch was identified to specific instrument by using SVM classifier to classify the pattern of each estimated pitch. At last, the accuracy was calculated by comparing the estimated pitch and instrument category and the corresponding MIDI data. The proposed algorithm preponderate over the NMF approach for 16.4% for 2 instruments task, 7.3 % for 3 instruments task and 9.2% for 4 instruments task.

Recognition accuracy of instrument identification by using 12 dimension MFCC features and proposed features is shown in Table 3. It shows the accuracy of identifying the correct instrument for each corresponding pitch from the polyphonic test signals which contain 2 instruments (for example guitar and piano), 3 instruments and 4 instruments respectively. The proposed algorithm preponderate over the MFCC features for 6.8% for 2 instruments task, 7.4% for 3 instruments task and 6.4% for 4 instruments task.

4. CONCLUSION

The motivation of this research is to develop an algorithm for musical instrument identification without given preconditions such as correct F0s. The proposed algorithm models each spectral envelope of underlying harmonic structure to approximate the real music as close as

	2 instruments signals (%)		3 instruments signals (%)		4 instruments signals (%)	
	MFCC	Proposed	MFCC	Proposed	MFCC	Proposed
altosax	73.6	77.2	46.8	52.9	40.5	47.1
guitar	68.5	73.8	51.4	58.7	38.7	46.8
piano	79.1	86.7	66.5	73.3	54.3	63.6
violin	66.7	76.5	60.2	67	48.5	53
flute	56.8	69.5	47.1	56.8	45	51.4
oboe	57	65.2	43.7	51.3	38.9	42.2
Total accuracy	67	74.8	52.6	60	44.3	50.7

Table 3. Recognition accuracy of instrument identification by using MFCC and proposed features

possible and uses the EM algorithm to estimate the parameters. New features such as Harmonic Temporal Timbre Energy Ratio (HTTER) and Harmonic Temporal Timbre Envelop Similarity (HTTES) are proposed to generate classifier for instrument identification. The proposed algorithm was intuitive and efficient for solving the musical instrument identification problem, which was proved by the experiments.

5. REFERENCES

[1] J. C. Brown, "Computer identification of musical instruments using pattern recognition with cepstral coefficients as features," Journal of the Acoustical Society of America, vol. 105, no. 3, pp. 1933–1941, 1999.

[2] A. Eronen and A. Klapuri, "Musical instrument recognition using cepstral coefficients and temporal features," in Proc. ICASSP, vol. 2, pp. 753–756, Istanbul, June, 2000.

[3] G. Agostini, M. Longari, and E. Pollastri, "Musical instrument timbres classification with spectral features," EURASIP Journal on Applied Signal Processing, vol. 2003, no. 1, pp. 5–14, 2003.

[4] T. Kinoshita, S. Sakai, and H. Tanaka, "Musical sound source identification based on frequency component adaptation," in Proc. IJCAI-CASA, pp. 18–24, Stockholm, Sweden, July-August, 1999.

[5] J. Eggink and G. J. Brown, "Application of missing feature theory to the recognition of musical instruments in polyphonic audio," in Proc. ISMIR, Baltimore, USA, Oct, 2003.

[6] T. Kitahara, M. Goto, K. Komatani, T. Ogata, and H. G. Okuno "Instrument Identification in Polyphonic Music: Feature Weighting to Minimize Influence of Sound Overlaps" EURASIP Journal on Advances in Signal Processing, Vol.2007, Article ID 51979, 15 pages, 2007.

[7] H. Kameoka, T. Nishimoto, Shigeki Sagayama, "A Multipitch Analyzer Based on Harmonic Temporal Structured Clustering," IEEE Trans. Audio, Speech and Language Processing, vol.15, no.3, pp. 982–994, Mar, 2007.

[8] A. Klapuri, "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness," IEEE Trans. Speech and Audio Proc., vol.11, no.6, pp. 804–816, 2003.

[9] M. Goto, "A real-time music-scene-description system: Predominant-F0 estimation for detecting melody and bass lines in real-world audio signals," in Proc. ISCAJ, vol. 43, no. 4, pp. 311–329, 2004.

[10] K. Miyamoto, H. Kameoka, T. Nishimoto, N. Ono, S. Sagayama, "Harmonic-Temporal-Timbral Clustering (HTTC) For the Analysis of Multi-instrument Polyphonic Music Signals," in Proc. ICASSP, pp. 113-116, Apr, 2008.

[11] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "RWC music database: Popular, classical, and jazz music database," in Proc. ISMIR, pp. 287–288, Paris, Oct, 2002.

[12] E. Vincent, N. Bertin, and R. Badeau, "Harmonic and inharmonic nonnegative matrix factorization for polyphonic pitch transcription," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP), pp. 109-112, Las Vegas, March, 2008.

[13] T. Heittola, A. Klapuri and T. Virtanen, "Musical Instrument Recognition in Polyphonic Audio Using Source-Filter Model for Sound Separation," in Proc. ISMIR, pp. 327–332, Kobe, Oct, 2009.

[14] J. Wu, Y. Kitano, T. Nishimoto, N. Ono, S. Sagayama, "Flexible Harmonic Temporal Structure for modeling musical instrument," International Conference on Entertainment Computing (ICEC 2010), Seoul, South Korea, Sep, 2010.

[15] K. Itoyama et al. "Integration and Adaptation of Harmonic and Inharmonic Models for Separating Polyphonic Musical Signals," Proc. ICASSP 2007, Vol. I, pp. 57-60, April 2007.