Towards A Sound Recognition System for Traditional Malay Musical Instruments

Norhalina Senan
Faculty of Computer Science and Information System
Universiti Teknologi Malaysia
Malaysia
012-701821454
halina@uthm.edu.my

Ali Selamat
Faculty of Computer Science and Information System
Universiti Teknologi Malaysia
Malaysia
019-7701757
aselamat@utm.my

ABSTRACT
In recent year, studies on music content analysis especially on musical instrument recognition system are developed extensively. However, almost all the studies are developed based on the Western musical instruments. Meanwhile, study on non-Western musical instrument especially on traditional Malay musical instrument is still lacking. With the enormous amount of instruments data and features schemes, adapting the existing techniques for recognizing the traditional Malay musical instrument sound might not be as easy due to the differences in the feature schemes, amount of sound samples and recording environment. Thus, the existing techniques that have been proposed for automatic musical instrument recognition system will be reviewed and evaluated especially on their performance in achieving the highest accuracy rate. In addition, the motivation, strategy and the factors that can affect the development of the traditional Malay musical instrument also will be discussed.

Keywords
Malay musical instruments, sound recognition system.

1. INTRODUCTION
Traditionally, almost all local musicologists recognize the musical instruments by their own knowledge gathered from the seminar, books or other references source. Some of them are capable to recognize the instruments by the sound produced. This is done through their own experience and practice. However, human perception can incorporate errors, due to partial misinterpretation, incorrect or inconsistent judgment of similar sound from different type of instruments, outside interference such as noise or perceived bias [8].

With the advances of digital signal processing and computational techniques, automatic musical instrument recognition has becomes an important aspect of music information retrieval (MIR). Mackay and Fujinaga [14] also claimed that automatic musical instrument recognition using machine learning is better in producing a good result compared to human capability. This automatic recognition system also can have numerous potential applications. For instance, recognizing and analyzing the content of the instrument can lead to more knowledge about the different musical styles and can be further utilized for computer-assisted musical instrument tutoring [8][20]. Furthermore, it also can be enhanced as a validation or quality control tool in musical instrument manufacturing.

Several studies have been conducted regarding these issues [4][5][7][12][23]. However, almost all the studies are developed based on the Western musical instruments. Meanwhile, study on non-Western musical instrument especially on local (Malaysian) environment is still lacking.

Thus, this paper discusses the overall issues and a review of the techniques to the development of the musical instrument recognition system especially for traditional Malay musical instrument. Section 2 presents the motivation of the research. A review of the technologies and the related issues are presented in the next session.

2. MOTIVATION
With the various potential applications that can be developed from this automatic musical instrument recognition system, there were significance need to explore further in this area especially when it involved a new domain which is traditional Malay musical instrument. Although there are various algorithms that have been discovered, adapting the existing algorithms or techniques in the new domain not be as simple due to the differences in the feature schemes and the activities involve in the recognition process.

In addition, the implementation of musical instrument recognition systems still has restricted practical usability due to the certain problem especially to find the right feature extraction scheme for the musical instrumentals sound. It is also fascinating to see that the feature schemes adopted in current research are all highly redundant. Hence, by simply selecting all the features available, it might give poor performance. This is because some feature give poor separability among different classes and some are highly correlated [12]. Furthermore, Deng et al. [3] also found that some of the features that they used in their research were not reliable for giving robust classification result. All these show that one of the most crucial issues in musical instrument recognition is to select the right feature extraction scheme from sound database. Hence, a new method will be investigated and derived correspondingly in order to produce a better result.

Therefore, the objectives of this research are (a) to study the different feature schemes of traditional Malay musical instrument sound; (b) to design and formalize a new feature selection method for identifying a good feature combination schemes; and (c) to validate the generated feature combination schemes using classifier.
3. LITERATURE REVIEW
This section discusses briefly the overview of traditional Malay musical instruments and several topics related to musical instrument recognition and feature analysis such as feature extraction schemes, feature selection techniques and classification algorithm used to validate the performance of selected features.

3.1 Traditional Malay Musical Instruments
The traditional Malay musical instruments are believed to have originated from different countries and cultures. For instance, the kompang was brought to the Malay Region by the traders from the Middle East in the thirteenth century [1]. Besides Arab countries, some of the instruments were also invented from other countries such as angklung which believed was brought by migration from Indonesia [16].

The traditional musical instruments play an important role in traditional Malay culture. The instruments were mainly used to accompany traditional dance such as kuda kepang and mak yong, wedding ceremony, traditional theater such as wayang kulit (shadow puppet) and religious function such as Maulud Nabi and berzanji [16].

In general, the traditional Malay musical instruments can be classified into four (4) categories which are membranophones, idiophones, chordophones and aerophones [21]. Membranophones and idiophones are also known as percussion instruments. These instruments are the largest and most important instruments in Malay traditional music. Figure 1 illustrates the category of the Malay musical instruments.

However, there are still several remains problem that need to be tackled in producing a good recognition system.

3.2 Automatic Musical Instrument Recognition
Automatic musical instrument recognition is a systematic approach that able to identify the complex features of the musical signals from the musical instruments database automatically. This is concern as the first step in developing a wide variety of potential applications such as musical tutoring system, automatic music transcription, multimedia databases annotation and automatic pirated detection [4][14][20].

In general, automatic musical instrument recognition process involved three (3) main steps which are feature extraction, feature selection and classification. Figure 2 shows the architecture of the musical instrument recognition system. There are various algorithms that have been explored in solving problem for each step in automatic musical instrument recognition system.

One of the most crucial issues of automated musical instrument recognition is to find the best feature schemes or properties [4][5][12]. This is important because different types of musical instruments have their own different types of feature schemes. In addition, features are fed to pattern recognition architecture as the input and are the basis in the lead of the classification process.

3.2.1 Feature Extraction
In automatic musical instrument recognition, one of the challenges is the ability to distinguish between instrument sounds. The challenges become more difficult when the instruments are played in a group and involved a complex mixture of instruments. Thus, feature extraction plays an important role for this purpose.

Feature extraction is the process of obtaining digital representation of large amounts of information contains in music instrument, music genre and many other fields. Deng, Simmermacher and Cranefield [3], explained that the extracted audio feature schemes can be used to interpret music with a less of human supervision. Furthermore, computational and learning cost become as major constraints in pattern recognition problem. Hence, by implementing feature extraction, these problems can be solved by reducing the amount of data required.

Various feature schemes have been identified and adopted by past research either by individual sets or combination of them. Typically, the features consist time-domain and frequency-domain features. Thus, this research will try to use three different extraction feature categories proposed by [4], which are mel-frequency cepstral coefficients (MFCC) features, perception-based features and MPEG-7 based features which consist both of features domain. The mean and standard deviation are calculated for each of the features for the classification purposes.

3.2.1.1 Mel-Frequency Cepstral Coefficients
Mel-frequency cepstral coefficients feature have been used not only in musical instrument classification but also in other audio processing area such as music genre and speech processing [4]. The MFCC value is computed directly from the power spectrum [17]. Typically, the first thirteen coefficients have been found to be most useful in musical instrument sounds features. The effectiveness of MFCC in identifying different type of audio features have been discovered in [5][18].
3.2.1.2 MPEG-7

The MPEG-7 provides a standard description of the audio in a pattern recognition framework. The descriptions consist of solid mathematical information about the audio signal [2]. The MPEG-7 low-level descriptors have seventeen (17) temporal and spectral descriptors divided into the following groups six (6) groups which are basic, basic spectral, signal parameters, timbral temporal, timbral spectral and spectral basis [15]. The details of each groups was stated in Figure 3.

![Figure 3. Overview of Audio Framework including Descriptors](image)

The MPEG-7 features were used in [22] to optimize sound representation in musical instrument recognition. The result shows that combination of temporal and spectral features produce the best result with 68.4% accuracy for k-NN classifier compared using single feature itself. Deng, Simmermacher and Cranefield [16] assessed seven (7) features of MPEG-7 descriptors in their work. The features are Harmonic Centroid (HC), Harmonic Deviation (HD), Harmonic Spread (HS), Harmonic Variation (HV), Spectral Centroid (SC), Long-Attack k-Time (LAT), and Temporal Centroid (TC). However, they found that some MPEG-7 features were not reliable for giving robust classification result. This is because they only use several features and intend using more MPEG-7 in future.

3.2.1.3 Perception-based

Perception-based features were extracted from multiple segments either in time-domain or frequency-domain of a sample signal. There are five (5) features in this category which are zero-crossing rate, root-mean-square, spectral centroid, bandwidth and flux.

3.3 Feature Selection

Feature selection, also known as feature reduction, can be defined as a technique of choosing the most relevant features for building robust classifier. By removing the irrelevant features, the performance of the classifier can be improved by reducing the “curse of dimensionality”, enhancing generalization capability, and reducing learning and computational cost. In addition, many researcher also claimed that even an optimal classifier difficult to classify accurately if the poor features are presented as the input [12][3][14]. The feature selection algorithms can be classified into two (2) main categories which are filter and wrapper algorithms [10]. The filter algorithm use the initial set of the features, and then applies the selected feature subset to the clustering algorithm, whereas, the wrapper algorithms incorporates the clustering algorithm in the feature search and selection. Essid, Richard and David [6] claimed that the wrapper algorithm more efficient than the filter algorithm, but more complex.

Liu and Wan [12] studied the feature selection for automatic classification of musical instrument sounds using filter algorithm which is sequential forward selection technique. This technique is convenient to provide a sub-optimized set of features. The results shows that the modified k-NN classifier using 19 selected features (6 temporal, 8 spectral, and 5 coefficients) achieves highest accuracy of 93.5%.

Other techniques on filter algorithms are Information Gain (IG), Gain Ratio (GR), and Symmetrical Uncertainty (SU). These techniques have been applied by [16] to find the best features schemes for instruments classification of classical music. IG filter produces ranked feature sets of the best classification accuracy. The average accuracy is decreased with the less features number; nevertheless the classification rate still can achieve almost 80% for only 3 features.

Meanwhile, study on wrapper algorithms also have been presented by several researchers. For instance, Essid, Richard and David [6] and Mackay and Fujinaga [14] applied Genetic Algorithm (GA) in their work. The GA perform better in [14] but less efficient when compared to others wrapper algorithm which is Inertia Ratio Maximization using Feature Space Projection (IRMFSP) in [6]. Essid, Richard and David [14] claim that the selection of fitness function for GA structure also can affect the overall performance.

3.4 Feature Validation via Classification

Classifier is used to verify the performance of the selected features. The accuracy rate achieved by the classifier will be analysed to identify the effectiveness of the selected features. Achieving a high accuracy rate is important to ensure that the selected features are the best relevance features that perfectly serve to the classification architecture which able to produce a good result.

However, the performance of the overall classification system is not only depends on the features used. There is also significance to ensure that the classifier is able to analyze and extract the implicit information of these features into an intelligible form [19]. There are various classification algorithms that have been used in musical instrument recognition system such as Support Vector Machine (SVM) [4][6][11], k-Nearest Neighbours (k-NN) [4][5][9][12] and Artificial Neural Network [4][13][23].

Herrera, Yeterian and Gouyun [9] studied the classification of drum kit sounds into three (3) main categories which are super-category, basic-category and sub-category. k-NN, K* (modified k-NN), Canonical Discriminant Analysis, C4.5 and Partial
Decision Trees (PART) were adopted as classifier. The highest accuracy rate based on the feature sets and categories varied from 90.7% to 99.3%.

The classification of the instrument into individual and four (4) groups of instrument’s family which are brass, woodwind, piano and string has been discussed by [3]. They used five (5) classifiers which are SVM, k-NN, Naïve Bayes, multi-layer perceptron (MLP) and Radial Basic Function (RBF). In individual classification, 3-NN achieved average accuracy of 98.4% over four instruments. However, even full feature set would not help much in classified woodwind instrument. Meanwhile, 1-NN produced highest accuracy of 96.5% for “Selected 17” features in family classification.

Wicaksana, Hartono and Foo [23] evaluated two types of Neural Network structure which are single network and composite network in giving highest accuracy rate in recognizing musical instruments. The study made use of 25 sounds samples from six (6) different instruments. The single network is slightly better when only four (4) instruments are to be classified. However, composite network significantly achieved better accuracy for identifying six (6) instruments. It can be concluded that composite network will produced better accuracy when more instruments are to be identified.

Liu and Wan [12] analyzed 351 instruments sounds from five (5) different families. The main objective of their research is to identify the effectiveness of selected features on classification performance. Three (3) classifiers are used which are Nearest Neighbour (NN), k-NN and Gaussian Mixture Model (GMM). The result shows that the performance increases when more features are used. The best feature sets for different classifiers are different. The k-NN classifier using the best 19 features achieves highest accuracy of 93%.

The new feature selection and classification strategy were introduced by [6] using pairwise classification technique with Hastie-Tibshirani approach. Ten (10) individual instruments were used in this study. SVM with RBF kernel is the most successful classifier with average accuracy rate of 87%.

4. METHODOLOGY

Figure 4 shows the methodology of this research which consist five (5) main activities which are data acquisition, data pre-processing, feature extraction, feature selection and classification. The brief description for each phases of this methodology are as follows:

4.1 Data Acquisition
The 71 sounds of traditional Malay musical instruments were downloaded freely from www.rickshriver.net/hires.htm. Distribution of the sounds into categories is shown in Table 1. However, the collection of the sounds is very limited. As solution, the sounds will be self-recorded soon.

4.2 Data Pre-Processing
The length of the sound files range from 0.1 second to around 60 seconds. As proposed by [12], every audio file is divided into frames of 256 samples. However, different size sample will be tested to see whether the size sample will affect the performance of the overall process. Each frame will be overlapped about 50%. This is to ensure there are no missing signals during the segmentation. The hamming-windowed filter is applied to each frame in order to improve the quality of the signal.

4.3 Feature Extraction
The segmented signal will be used to extract the sounds features. Three (3) categories of features extraction schemes will be extracted which are perception-based, MFCC and MPEG-7 as applied by [5]. Matlab 7.0 will be used to extract the features. The mean and standard deviation of the frame-based features will be computed for each audio file.

4.4 Feature Selection
Initially, the existing techniques which have been exploited by other researcher which is Principal Component Analysis (PCA)
will be used to select the feature. Then, the effectiveness of the algorithm will be analyzed. After that, the algorithm will be enhanced in order to improve the performance of the algorithm and the classification accuracy respectively.

4.5 Classification
The selected features will be further assessed using the classifier. The existing classifier that will be used such as Independent Component Analysis, Neural Network and Support Vector Machine will be used. Then, the result will be compared. This is done in order to identify the effectiveness of the proposed technique. The effectiveness is determined from the success rate of sound classification produced by the classifier. Finally, a report of data analysis from the testing will be tabulated.

5. DISCUSSION
In general, from the previous work, we found that there are various techniques and features that have been used and able to produce the best accuracy rate when classifying the sounds. However, as claimed by [4] there are still no standard benchmarking to assess the best techniques and features used. One of the reasons is every different researcher use a different set of data that contains a different type of musical instrument. With this constraint, therefore, direct comparison of these various approaches is quite difficult. In addition, with the different types of features schemes, the implementation of existing techniques to this research also must be considered and cannot be applied directly.

From the literature, we found that the research in this domain stills open problems and there were several issues that need to be tackled in future. One of the most crucial issues in musical instrument recognition system is to find the best feature schemes or sounds properties [5][4][12]. This means that more attention should be given in the data pre-processing, feature extraction and also features selection process.

Hence, several factors that might affect the overall performance of the automatic traditional Malay musical instrument recognition system were identified. Table 2 shows the factors that must be considered for each stage of the recognition process.

<table>
<thead>
<tr>
<th>Process</th>
<th>Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sound Recording</td>
<td>a) Acoustic environment</td>
</tr>
<tr>
<td></td>
<td>b) Player (gender, age)</td>
</tr>
<tr>
<td></td>
<td>c) Solo/Mixture</td>
</tr>
<tr>
<td></td>
<td>d) Time</td>
</tr>
<tr>
<td></td>
<td>e) Number of the data set</td>
</tr>
<tr>
<td>Feature Extraction</td>
<td>a) Length of the sounds</td>
</tr>
<tr>
<td></td>
<td>b) Frames size</td>
</tr>
<tr>
<td></td>
<td>c) Number of overlap</td>
</tr>
<tr>
<td></td>
<td>d) Filter Technique</td>
</tr>
<tr>
<td></td>
<td>e) Types of Features</td>
</tr>
<tr>
<td>Feature Selection</td>
<td>a) Method/algorithm (wrapper/filter)</td>
</tr>
<tr>
<td></td>
<td>b) Number of selected features</td>
</tr>
<tr>
<td>Classification</td>
<td>a) Method/algorithm</td>
</tr>
<tr>
<td></td>
<td>b) Number of cross-validation folds</td>
</tr>
</tbody>
</table>

6. CONCLUSION AND SUGGESTED WORK
In conclusion, automatic musical instrument recognition is still an open problem. A number of techniques have been applied in the past that differ in the features used to describe the important of recognition strategy. However, there are potential ways to improve the algorithms especially on features selection issues that have major influence to the recognition accuracy. Thus, this study has a significant importance to find better mechanisms for sound recognition problem. This is become more exciting when it involved different types of musical instruments which are the Malay traditional musical instruments instead of Western musical instruments.

Future work includes investigating the specific features that can improve the recognition performance. Additionally, different feature selection and classification techniques will be assessed to identify the best techniques which can produced the highest accuracy rate. Then, the techniques will be improved in order to achieve highest accuracy.

7. ACKNOWLEDGMENTS
Special thanks to Dr. Musa Mokji, lecturer from Faculty of Electrical Engineering, Universiti Teknologi Malaysia for his kindness in teaching the Matlab programming for Digital Signal Processing.

8. REFERENCES


