Towards a MIR System for Malaysian Music

Shyamala Doraisamy
Dept. of Multimedia
Faculty of Comp. Sc. & IT
University Putra Malaysia
shyamala@fsktm.upm.edu.my

Hamdan Adnan
Dept. of Music
National Arts Academy
Ministry of Culture, Arts and Heritage
Malaysia
hamdan_adnan@hotmail.com

Noris Mohd. Norowi
Dept. of Multimedia
Faculty of Comp. Sc. & IT
University Putra Malaysia
noris@fsktm.upm.edu.my

Abstract

Systems for the archival of musical documents digitally and development of digital music libraries are currently being researched and developed extensively. However, adapting these systems for the archival and retrieval of Malaysian music materials might not be as straightforward due to the distinct differences in musical structure and modes of non-Western music. This paper covers the motivations for the creation of a MIR system for Malaysian Music and outlines the plans for its development.

Keywords: Malaysian Music, Digital Libraries, Music IR, Genre Classification, *N*-grams.

1. Introduction

This rapid development in MIR systems research and development is quite evident. However, almost all of the systems developed are largely based on Western music. Interest in the incorporation of non-Western music data such as Korean, Indonesian and Malaysian is growing [1,2,5].

The study by Norowi et.al [1] investigates automated genre classification of Traditional Malaysian Music. A consensus on the definition of Malaysian music can be difficult to achieve. Based on a very broad and general view, it encompasses music from various cultures -- Malays, Indian, Chinese and native music of those from Sabah and Sarawak (East Malaysian states on the Island of Borneo), as the Malaysian community is multi-racial. The modern music (contemporary, pop, etc.) of these various cultures further expands the genre classes of Malaysian music. However, we define the scope of traditional Malaysian music based on the study by Nasuruddin [4], and this is discussed in the following section.

In general, traditional Malaysian music is no longer listened to very widely, and the Malaysian society in general simply associate it as music played for traditional ceremonies. Most are ignorant of its forms, genres and

also the possibility of this rich heritage disappearing. Apart from it not usually being the favourite choice for listening when compared to pop music (either Western, Malay, Chinese, Indian, etc.), there are also very few musical documents available for several of the genres. For the genres with relatively large collections, either in private collections or with government agencies, these may not always be easily accessible. There are also genres where there are no notation systems, and the music is passed on orally from generation to generation.

This paper discusses the need to use state-of-the art MIR technologies for the archival and retrieval of traditional Malaysian music digitally and a review of technologies towards its development. Section 2 presents an overview of Malaysian Music. A review of MIR technologies and the related problems of its use with Malaysian music are presented in the next section.

2. Traditional Malaysian Music

This section discusses briefly the genres and musical structures of traditional Malaysian music and the available collections.

2.1 Genres and musical structures

Using the term Malaysian Traditional Music would encompass all of the traditional music in Malaysia including Sabah and Sarawak. Nasuruddin [4] categorises these into six broad categories: shadow puppet music, dance theatre, music with Indonesian influence, percussion music and *nobat*, syncreatic Malaysian music, and music from Sabah and Sarawak. He also attempts to notate the microtonal instruments using standard Western notation. However it is just an approximation and does not reflect the actual pitches.

For the purpose of developing the MIR, these genres would have to be reclassified as tonal, microtonal and un-pitched, based on the instruments used and the vocal styles involved during the actual performance.

A number of genres from the syncreatic forms, and several based on the Indonesian influence such as *Gamelan* and *Caklempung* have been notated using Western notation or tablature. These can be loosely classified as tonal, and other genres such as dance theatre, shadow puppet theatre and other related genres would be classified as microtonal.

2.2 Musical Document Collections

Most of the Malaysian music archives are in the form of vinyl, tapes, manuscripts and scores. A few collections of vinyl recordings are in mint condition, and the rest simply not well preserved. A number of rare recordings exist in the cassette format. However, due to the way these were recorded, most are not very clear audibly and requires audio cleanup and enhancements to be useful for purposes of MIR. As for music manuscripts, the collections in certain agencies have been through fire and floods.

The military has its own musical tradition inherited from the British army who once colonised the country. Recently a collection of approximately 400 titles from the Boosey military band library was discovered in neglected state in a storage facility. Restoration of these is being planned and the first phase includes fumigation and storage in air-conditioned facilities. Follow up processes are vacuuming and scanning of the scores forming digital versions.

For genres that are lacking in documents, the Ministry of Culture, Arts and Heritage under its various agencies have been documenting various traditional art forms via video and audio. For music passed on aurally from generation to generation, some of the singings of the 'masters' would have to be recorded and transcribed manually. All these would be converted into the appropriate digital format (if not already in the suitable format) for MIR purposes.

3. System Integration

In general, the development of MIR systems requires integration of many technologies, most of which are multi-disciplinary. Computing technologies include Information Retrieval (IR), Digital Libraries (DLs), DBMS (Database Management Systems), MMIRSs (Multimedia Indexing and Retrieval Systems). This paper looks closely at three computing technologies and presents the issues that would have to be addressed towards the integration and development of the MIR system for Malaysian Music.

3.1 Information Retrieval

The study by Doraisamy [3] investigated the use of text IR systems for MIR. Musical sequences are converted as text strings. These strings are then subdivided into overlapping sub-strings of *n*-grams and encoded with text alphabets. The musical *n*-grams are then used as terms for the indexing and retrieval of the musical document. The data-driven encoding formula based on tonal music in this study would have to be reformulated. Micro-tonality would have to be addressed and one approach suggested for computer representations of Microtonal Music by Suyoto [2] is to alter or approximate the micro-tonality using the MIDI pitch bend events. This would have to be tested.

3.2 Feature Extraction and Data Mining

Automatic genre classification has been investigated by [2] using Traditional Malaysian Music genres. This would be important where there may be problems in manual classification of these genres for retrieval systems due to it being only recognized by a small group of people in the future. Testing the system against manual classification is currently being conducted.

3.3 Digital library

Amongst the large music digital library projects is Meldex [7]. This was developed using Greenstone, a suite of software for building and distributing digital library collections available at http://www.greenstone.org/. It has also been shown to be useful for the retrieval of online scores available at http://chopin.lib.uchicago.edu/ [6]. This approach would be adapted for the retrieval of images of our scores as most of it would have to remain in the archives due to the delicate conditions of the source and also ownership and copyright issues.

4. Conclusion

Based on the enormous amount of data on Malaysian music available through government agencies such as the National Archives, Radio Television Malaysia (the Malaysian broadcasting agency) and private collections, there is a need the retrieval or cataloguing system to be updated based on state-of-the-art MIR technologies. Ongoing work includes outlining the system architecture and an early prototype for a MIR system for Malaysian Music.

References

[1] N.M. Norowi, S. Doraisamy and R.W.O.K. Rahmat, "Factors Affecting Automatic Genre Classification: An Investigation Incorporating Non-Western Musical Forms", in *ISMIR 2005 Sixth Int. Conf. on Music Inf. Retr. Proc.*, 2005, pp. 13-20.

[2] I. S.H. Suyoto, and A.L. Uitdenbogerd. "Exploring Microtonal Matching," in *ISMIR 2004 Fifth Int. Conf. on Music Inf. Retr. Proc*., 2004, pp. 158-163.

[3] S. Doraisamy, "Polyphonic Music Retrieval: The N-Gram Approach", *PhD Thesis, Imperial College London*, 2004.

[4] M.G. Nasuruddin, "Muzik Tradisional Malaysia", Dewan Bahasa dan Pustaka, Kuala Lumpur, 2003.

[5] J.H. Lee, J.S. Downie and S.J. Cunningham, "Challenges in Cross-Cultural/Multilingual Music Information Seeking", in *ISMIR 2005 Sixth Int. Conf. on Music Inf. Retr. Proc.*, 2005, pp. 153-160.

[6] T.A. Olson and S.J. Downie, "Chopin early editions: Construction and usage of online digital scores, in *ISMIR 2005 Sixth Int. Conf. on Music Inf. Retr. Proc.*, 2003, pp. 247-248.

[7] R. McNab, L.A. Smith, D. Bainbridge and I.H. Witten, "The New Zealand digital library MELody index, D-Lib Magazine, May 1997.