

# A CONTENT-BASED MUSIC SEARCH SYSTEM USING QUERY BY MULTI-TAGS WITH MULTI-LEVELS OF PREFERENCE

Ju-Chiang Wang<sup>1,2</sup>, Meng-Sung Wu<sup>1</sup>, Hsin-Min Wang<sup>1</sup> and Shyh-Kang Jeng<sup>2</sup>

<sup>1</sup> Institute of Information Science, Academia Sinica, Taipei, Taiwan

<sup>2</sup> Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan

## ABSTRACT

This demonstration paper presents a novel content-based music search system that accepts a query containing multiple tags with multiple levels of preference (denoted as an MTML query) to search music from an untagged music database. We select a limited number of most frequently used music tags to form the tag space and design an interface for users to input queries by operating the scroll bars. To effect MTML content-based music retrieval, we introduce a tag-based music aspect model that jointly models the auditory features and tag labels of a song. Two indexing methods and their corresponding matching methods, namely pseudo song-based matching and tag affinity-based matching, are incorporated into the pre-learned tag-based music aspect model. The content-based music search system is implemented on the MajorMiner dataset, which consists of 2,472 10-second music clips and their associated human labeled tags crawled from the MajorMiner website. The MTML query interface contains 36 top tags used in the dataset. We randomly select 1,648 music clips with their tag labels for training the tag-based music aspect model and 824 clips without using their tag labels for building the untagged music database for content-based retrieval.

## 1. INTRODUCTION

Music tags are generated from a folksonomy, which is a full-scale taxonomy of music that reflects the current usage among Internet users [1]. Several web-based music tagging games have been created with the purpose of collecting useful tags, e.g., MajorMiner [2] and Tag A Tune [3]. These applications provide tagged music resources for researchers to investigate users' tagging behavior among different kinds of music tracks [1]. However, the collected

This work was supported in part by the Taiwan e-Learning and Digital Archives Program (TELDAP) sponsored by the National Science Council of Taiwan under Grant: NSC 100-2631-H-001-013.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

© 2011 International Society for Music Information Retrieval

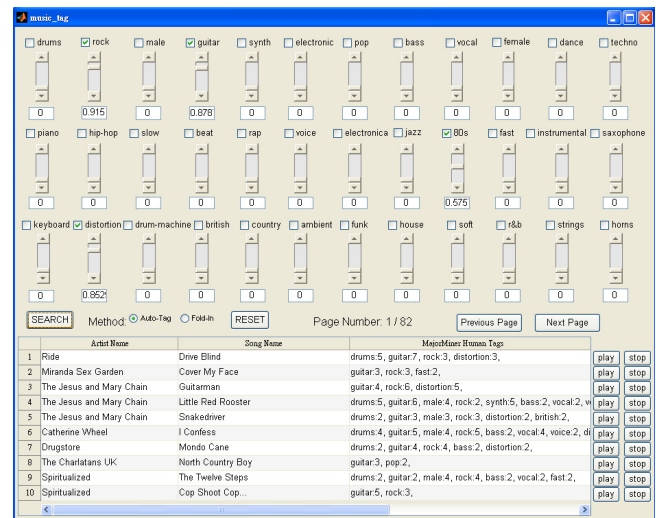


Figure 1. The query interface of the MTML content-based music search system. We take the top 36 most frequently used tags to form the query interface, in which users can input a query containing multi-tags with different preference levels (between 0 and 1) through the scroll bars.

music tags have only been assigned to existing or popular music tracks, i.e., there is no tag available for new or unpopular tracks. The cold-start issue has motivated research into a number of interesting topics, such as automatic music tag annotation and content-based music retrieval in a query-by-tag manner.

In this demonstration, we propose a novel content-based music search system that accepts a query comprised of multiple tags with multiple levels of preference (denoted as an MTML query), derived from a predefined music tag set, to search for music in an untagged music database. As the screenshot shown in Figure 1, users can input queries by simply operating the scroll bars to give a preference weight (between 0 and 1) for each tag in a predefined tag set. It is believed that a tag will enter the common musical vocabulary once it is adopted by a large number of users. Therefore, it is reasonable to select a limited number of most frequently used music tags to form the semantic space of possible tag queries for use in the MTML interface.

We believe that the MTML query can help facilitate content-based music retrieval for two reasons. First, unlike images, which often contain only a few clearly identifiable

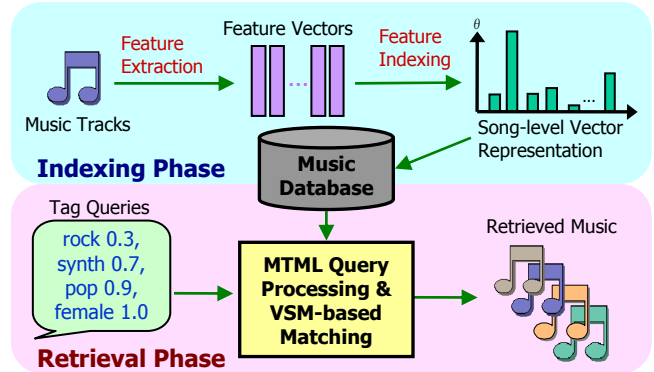
objects [3], a piece of music can be described in nature by multiple music tags. Therefore, retrieving music with multi-tags is more fulfilled and effective for users' information needs than a single tag. Second, a song in a tagged music database has multiple tags with different counts. The tag count, which indicates the tag popularity, corresponds to the number of users who have annotated the song with the tag. Therefore, the MTML query can be directly matched with the *tag count distribution* of each song in a *tagged music database* to search relevant music. The MTML query can also be matched with the *tag affinity distribution* of each song in an *untagged music database*. The tag affinity distribution of an untagged song consists of the confidence degree given by each tag predictor. The MTML query actually gives a more precise query than a single tag query and a simple binary multi-tags query.

## 2. METHODOLOGY OVERVIEW

The proposed MTML content-based music search system is implemented in two phases: the *feature indexing phase* and the *music retrieval phase*, as shown in Figure 2. In the feature indexing phase, each music track in the untagged music database is extracted into a set of 70-dimensional frame-based feature vectors by MIRToolbox [4], and then the track is indexed as a fixed-dimensional vector based on its audio features. We use two approaches to index track-level audio features: indexing based on an auditory feature reference (implemented by a global Gaussian mixture model) or indexing based on automatic music tagging. In the retrieval phase, given an MTML query, the music search system will return a ranked list of music tracks. We apply two matching methods that correspond to the abovementioned indexing approaches, namely, the *pseudo song-based matching* method (denoted by the "Fold-In" method in the query interface) and the *tag affinity-based matching* method (denoted by the "Auto-Tag" method in the query interface) using the well-known vector space model (VSM). Users can switch between the two methods in the interface. To achieve content-based music retrieval using MTML queries, we jointly model the tag labels with counts and the auditory features of each track in a tagged training music database. Then, the resulting tag-based aspect model is applied on each track in an untagged music database. The methodology of the system is detailed in [5].

## 3. SYSTEM OVERVIEW

The content-based music search system is implemented on the MajorMiner dataset, which contains 2,472 10-second music clips and their associated human labeled tags, crawled from the MajorMiner website [2] during March 2011. The raw human tag set includes about 1,031 unique



**Figure 2.** The flowchart of the MTML content-based music search system.

tags. The top 36 most frequently used tags are used in the MTML query interface. We randomly select 1,648 music clips with their tag labels for training the tag-based music aspect model based on the 36 tags. The remaining 824 clips without using their tag labels form the untagged music database for the content-based music search system. The system automatically indexes each of the 824 clips in the untagged music database based on its audio features. Then, given an MTML query, the system will return a ranked list of music clips, in which each clip is described by its artist name, song name, audio waveform (for online listening) and the MajorMiner human labeled tags with counts. Note that the MajorMiner human labeled tags are shown as a reference for subjective evaluation. These tags are considered as "weak labels" because some tags could be missing. In other words, if a tag does not appear in the human labeled tags of a song, it does not mean that the song does not have the tag. This is the nature of social tagging since, when tagging a piece of music, people create tags for it rather than verify tags from a predefined tag set. Users can subjectively evaluate the quality of the ranked list based on their listening experiences.

## 4. REFERENCES

- [1] P. Lamere, "Social tagging and music information retrieval," *J. New Mus. Res.*, 37(2), pp. 101-114, 2008.
- [2] M. Mandel and D. Ellis: "A web-based game for collecting music metadata," *J. New Mus. Res.*, 37( 2), pp. 151-165, 2008.
- [3] E. Law, B. Settles, and T. Mitchell: "Learning to tag from open vocabulary labels," *ECML*, 2010.
- [4] O. Lartillot and P. Toivainen, "A Matlab toolbox for musical feature extraction from audio," *DAFx*, 2007.
- [5] J.-C. Wang, M.-S. Wu, H.-M. Wang and S.-K. Jeng: "Query by multi-tags with multi-level preferences for content-based music retrieval," *IEEE ICME*, 2011.