Content-based Music Audio Recommendation

Pedro Cano
Music Technology Group
IUA-Universitat Pompeu Fabra
08003, Barcelona, Spain
pcano@iua.upf.es

Markus Koppenberger
Music Technology Group
IUA-Universitat Pompeu Fabra
08003, Barcelona, Spain
koppi@iua.upf.es

Nicolas Wack
Music Technology Group
IUA-Universitat Pompeu Fabra
08003, Barcelona, Spain
nwack@iua.upf.es

ABSTRACT
We present the MusicSurfer, a metadata free system for the interaction with massive collections of music. MusicSurfer automatically extracts descriptions related to instrumentation, rhythm and harmony from music audio signals. Together with efficient similarity metrics, the descriptions allow navigation of multimillion track music collections in a flexible and efficient way without the need for metadata nor human ratings.

Categories and Subject Descriptors
H.5.5 [**Information Interfaces and Presentation**]: Sound and Music Computing – *signal analysis, synthesis, and processing, systems.*

General Terms: Algorithms

Keywords: Music content management, music information retrieval, content-based audio retrieval, music recommendation.

1. INTRODUCTION
The standardization of personal computers, the ubiquity of high-storage portable devices, the proliferation of Peer2Peer networks and world-wide low-latency networks have dramatically increased digital audio dissemination and access. Major music labels now provide their music catalogs in near-CD quality formats through online distributors such as Apple iTunes, Amazon or Yahoo! Music. Additionally, the world-wide web is also extremely rich in music.

Currently, the common ways to access music are to query by artist or song names (or other types of editorial data) or to browse recommendations generated by collaborative filtering, that is, to browse information of the type "users that bought this album also bought this album". A clear drawback of the first approach is that consumers need to know the name of the song or the artist beforehand. The second approach is only suitable when a fairly large number of consumers have heard and rated the music. This situation makes it difficult for music lovers to access and discover music whose editorial data is unknown and which remains outside the main commercial streams. Another issue with collaborative filtering methods is that the similarity recommendations created by analyzing the behavior and ratings of users do not necessarily correspond to actual music similarity but may be biased by popularity [1].

Audio Fingerprinting is the first audio content-based technology that has found its place in the industry. Fingerprinting enables the identification of a musical piece from short excerpts (possibly distorted) of acoustic signals. This technology can recognize unknown music by holding a cell phone close to loudspeakers in a pub, monitor what is broadcast on the radio or match the mp3 files on a hard disk to proper metadata. Fingerprinting systems use low-level analysis of audio signals to extract compact signatures of recordings (the fingerprints) which are stored on a database. When the system is presented with an unknown recording, its fingerprint is extracted and compared to those stored on the database [2]. They often permit robust and efficient music identification but cannot address the problem of music similarity since they do not represent conveniently the abstract dimensions humans consider when rating musical pieces as similar or dissimilar. To give an example, fingerprinting technologies cannot identify covers of a music title, nor tell its musical genre, nor recognize the artist unless it has a recording of that very music title analyzed in the fingerprint database.

Recently, much effort has been put in computational modeling of music similarity, including works by Blum et al. [3] or Berenzweig et al. [4]. Only recently, world-wide cross-validations of music similarity systems have been conducted, in the form of a public competition, during the International Symposium on Music Information Retrieval (ISMIR) 2004 (see http://ismir2004.ismir.net/ISMIR_Contest.html). A detailed analysis of the results [5] shows that state-of-the-art technology, based on standard machine learning techniques associated to relatively low-level descriptions of musical items, such as Mel-Frequency Cepstrum Coefficients or Spectrum Histograms, yield promising results. They are able to classify 729 music tracks into 6 different genres with an accuracy of 78.8%. They are also able to identify artists from a collection of 120 music titles out of a list of 40 artists with an accuracy of 24%.

Despite these promising results, no system has yet been deployed for industrial exploitations. There are several reasons for this. A typical online music provider classifies among over a hundred genres and sub-genres, it deals with tens of thousands different artists and collections of over a million music titles. Current technologies have not dealt with real world problems. Moreover, state-of-the-art solutions are based on low-level representations that may conceal the truly relevant aspects of music. Higher-level musically meaningful representations may hold the key towards efficient, effective and human understandable music recommendation systems.

Research at the crossroad between music psychology and neuroscience provides definitions for "truly relevant aspects of music". According to [6], music cognition can be pictured as several distinct and interconnected modules. They propose a functional cognitive architecture resting on a series of neurophysiological experiments (e.g. analysis of music-related deficits, case studies of specific music impairments). Subsequent to the low-level analysis performed by the outer and middle

Copyright is held by the author/owner(s).
MM'05, November 6–11, 2005, Singapore.
ACM 1-59593-044-2/05/0011.

auditory system, two types of processes would take place: "temporal organization" vs. "pitch organization". While listening to music, our auditory cortex would therefore represent distinctly, among other percepts, rhythm and meter (temporal organization) as well as tonality, intervals and melodic contours (pitch organization).

We will demonstrate a music browsing and recommendation system based on a high-level music similarity metric and computed directly from audio data. This metric accounts for several perceptually and musically meaningful dimensions as those evidenced in music psychology and neuroscience research.

Rhythm is of course an important musical aspect, we represent it with several descriptors, automatically computed from audio signals: Tempo, Meter, Rhythm patterns (characteristic for instance of a Waltz pattern or a Cha Cha pattern) and finally Swing (typical for instance of Jazz pieces) [7]. Complementary to rhythm, our similarity metric also accounts for tonal aspects of musical pieces [8]. It implements the Tonal Strength, i.e. the degree of tonality of a piece (consider for instance the difference between a Beatles song and a piece by Alban Berg, or to a less extent, a Jazz piece, likely to account for tonality modulations). We also make use of the KeyNote and the KeyMode. The former is, to put it simply, the main chord of a song, taking values among the twelve semitones of a chromatic scale (A, A#/Bb, B, C, C#/Db and so on) while the latter indicates the type of scale that is used in the composition: either Major or Minor. Minor modes are sometimes associated with 'sad' mood, and major modes with 'happy' moods.

In addition, in our system, musical similarity also accounts for the Timbre. Based on spectral characteristics of the audio signals, this dimension represents aspects of the instrumentation [9] as well as post-production and sound quality characteristics. Our system tackles another important perceptual dimension, the Dynamics of audio signals, that is, the variations of loudness/amplitude with time. For instance, Classical music usually shows important variations with respect to this dimension unlike many highly-compressed Heavy-Metal pieces. The last dimension implemented in our similarity measure concerns the Genre probability which gives an estimate of the membership of a given piece to several established musical genres (as e.g. Pop, Classical, Jazz, etc.).

By default, a global measure of similarity is defined by specific weights assigned to these diverse musical dimensions. This results in a specific representation of the musical space that users can explore. As an additional feature, users can adjust the similarity metric at will, giving particular emphasis to a specific musical dimension, and thus exploring a different, personalized musical space. Evaluating the relevance of the "default" similarity metrics is a difficult task that can easily become highly subjective. In order to avoid this pitfall, we will report in details on an evaluation of our system with respect to the evaluation framework set up for the ISMIR 2004 [5]. The current state-of-the-art system for artist identification yields an accuracy of 24% for a 40 artist classification. Our system outperforms it by a factor of 2 (60%). Moreover, it reaches comparable performances (24% of correct artist identification also with respect to the same metrics) on a much bigger dataset (around 400 times bigger) consisting in 273,751 songs from 11,257 different artists.

We will demonstrate the system online performance in query-by-example tasks on a musical repository of over a million songs: retrieval takes tenths of a second while extraction of the aforementioned set of musical features runs 20 times faster than playing time, both operations carried on using an off-the-shelf PC.

The system functionalities can be accessed and evaluated on a database of 5000 items from Magnatune (only this smaller legal database from http://www.magnatune.com is publicly available for copyright reasons). See http://musicsurfer.iua.upf.edu (user: acm and passwd: acm321.

Additional authors: José Pedro Garcia, Thomas Aussenac, Ricard Marxer, Jaume Masip, Òscar Celma, David García, Emilia Gómez, Fabien Gouyon, Enric Guaus, Perfecto Herrera, Jordi Masseguer, Beesuan Ong, Miguel Ramirez, Sebastian Streich and Xavier Serra.

2. REFERENCES

[1] Cano, P., Celma, O., Koppenberger, M., and Buldu, J.M. "Music Artist Recommendation Networks", submitted.

[2] Cano, P., Batlle, E., Kalker, T., and Haitsma, J. (2002a). A review of algorithms for audio fingerprinting. In Proc. of the IEEE MMSP, St. Thomas, Virgin Islands.

[3] Blum, T., Keislar, D., Wheaton, J., and Wold, E. (1999). US Patent 5,918,223.

[4] Berenzweig, A., Logan, B., Ellis, D. P.W., and Whitman, B. (2004). A large-scale evaluation of acoustic and subjective music-similarity measures. Computer Music Journal, 28(2):63–76.

[5] Cano, P. et al, "ISMIR 2004 Audio Description Contest", submitted.

[6] Peretz, I., Coltheart, M. "Modularity of music processing." Nature Neuroscience 2003. 688-691.

[7] Gouyon, F. and Dixon, S. (2005). A review of automatic rhythmic description systems. Computer Music Journal, 29(1).

[8] Gómez, E. (2005). Tonal description of polyphonic audio for music content processing. INFORMS Journal on Computing, 17(11).

[9] Herrera, P., Peeters, G., and Dubnov, S. (2003). Automatic classification of musical instrument sounds. Journal of New Music Research, 32(1).