AUDIO FINGERPRINTING - WHAT, WHY AND HOW

K. Banuroopa[1], Dr. D. Shanmuga Priyaa[2]

Abstract

An Audio fingerprint is a small digest of an audio file computed from its many attributes. Like fingerprints of human, the fingerprint of an audio file identifies an audio file in a database but does not allow retrieving any other characteristics of the files. The primary application of Audio fingerprinting is to label unlabeled audio regardless of the audio file format. Applications of Audio fingerprint include creating and indexing large audio library, the protection of author's copyrights within a Digital Right Management (DRM) system, automatic tagging of audio files and even using it for audio stegnography for a secure communication.

Keywords : Audio fingerprinting, MIR, QbH, CBID,

I. Introduction

Data mining is a process of extracting useful information and pattern from a large database. The information retrieved using the nontrivial data mining techniques will usually be implicit, previously unknown and profoundly useful. In the recent decade due to availability of internet to many people can access and use music which distributed in large amount in the internet. As the availability of music and music related information increases the need to organize and analyze this information also increases. The data mining methods when applied to music database it can do genre classification, emotion and mood detection, playlist generation and feature extraction to automatically tag the audio file. Feature extraction means reducing the resources needed to describe a large database. The features extracted can be acoustic features like pitch, frequency, length of the audio or musical genre, artist information. For music data mining , feature extraction from acoustic signal plays an important role.

II Music Information Retrieval

Music Information Retrieval (MIR) is an active research area devoted to the information needs of the music listeners. MIR includes many different approaches which aim at music management, easy and correct access to the music for the enjoyment of the listeners. Most of the approaches of the MIR are content-based. The aim of the content-based approaches is that an audio file can identified and described by the features extracted from its content rather than its meta data. Content –based information retrieval is an alternative approach to the traditional text-based searching process.

Query by Humming (QbH) is a content-based information retrieval approach which uses a small

[1]Research Scholar, Department of Computer Science, Karpagam University.

[2]Professor Department of information technology, Karpagam University

piece of audio as a query instead using meta such as name, artist, album etc. QbH can be used when meta data information is not available. The music information retrieval systems the piece of audio used can be sample of original file, a hummed tune of the music or even whistling the tune. [1]. QbH uses complex algorithms to convert the hummed tune to query and search the music database for a match.

Another Content-based audio identification (CBID) system called Audio Fingerprinting is also used for audio file identification in a large database. This approach calcutes a small fingerprint or digest of the each and every audio file in the database and uses it as an index to identify the audio file searching. This method is used to automatically label audio data when presented with an unlabelled audio file. The audio fingerprint of the query audio file is calculated and matched against the fingerprints of audio files in the database instead of searching the entire file for a match.

III Audio Fingerprinting

An audio finger print is a small set of features that uniquely identifies a song. An audio fingerprint can be used for broadcast monitoring, audience measurement, meta-data collection.

Following are the desired properties of a good, robust and efficient audio fingerprinting system:[3]

- Accuracy: measure of correctness
- Complexity : computational overhead and cost involved
- Fingerprint rate (size): amount of elements or bits extracted per second or song.
- Granularity: The minimum length of the audio clip required for a dependable identification
- Reliability: measure of to what level a fingerprint can be depended upon for its accuracy in returning matches correctly.
- Robustness: Ability to accurately identify an item, regardless of the level of compression and distortion or interference in the transmission channel and withstand the effect of signal processing operations
- Scalability: to hold a large number of fingerprints
- Security: Accessibility through key only.

The audio fingerprinting system has two basic processes: (1) calculation of Fingerprint (2) matching fingerprints. [2]. In his paper Pedro Cano depicts a frame work for content-based audio identification using audio fingerprinting.

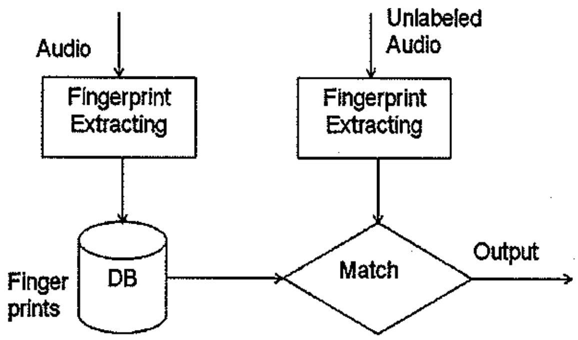

Figure1 : Content–based Audio Identification Framework.

The first process of fingerprint extraction involves a frontend and fingerprint modeling for the entire database of the music. The modeling process can use any one of the fingerprinting methods to compute the fingerprints of the audio signals by extracting its features. Then the audio fingerprints are stored in the database. The Fig-2 gives the sequence of various stages of fingerprint modeling.

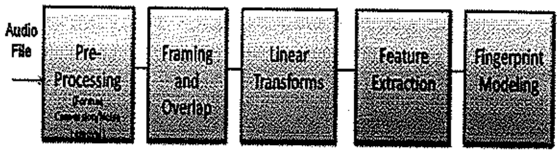

Figure 2 : Audio Finger Print Extraction

A : Preprocessing

The first step is to pre-process the audio file. In this step one or more of the following processes are done[2] :

1. Digitalize audio if necessary
2. Conversion to a general format
3. Averaging left and right channels
4. Bandpass filtering
5. GSM coder/decoder in a mobile phone identification system
6. Pre-emphasis
7. Amplitude normalization.

B. Framing and Overlap

In this step the signal is divided into frames. The number of frames computed per second is called frame rate. A tapered window function is applied to each block to minimize the discontinuities at the beginning and end. Overlap must be applied to assure robustness to shifting.

C. Linear Transforms

In the next step of applying linear transforms, the transformation of the set of measurements to a new set of features is done. By choosing the correct transform the redundancy is reduced profoundly. The most common transformation is the Fast Fourier Transform (FFT) and the Discrete Cosine Transform (DCT). DCT is said to exhibit shift invariance property [4].

D. Feature Extraction

Once the audio file is made as time-frequency representation, additional transformations are applied in order to generate the final acoustic vectors. In this step the objective is again to reduce the dimensionality and, at the same time, to increase the invariance to distortions and to extract more perceptually meaningful parameters. The many methods used for feature extraction are as given below [2] :

- Mel-Frequency Cepstrum Coefficients (MFCC)
- Spectral Flatness
- High-level descriptors
- Pitch
- Bass
- Robust Hash
- Frequency Modulation

E. Post Processing

In the post processing step after features are extracted the following processes are done to make the signal a higher order derivative for better calculations [2]. Normalization is done to ease hardware implementations, reduce the memory requirements and for convenience in subsequent parts of the system. Decorrelation, differentiation Quantization is done to gain robustness against distortions.

IV Audio Fingerprint Methods

The Shazam algorithm was patented by Wang and Smith in the year 2002. In this method discrete-time Fourier transform is used to create a time-frequency spectrogram. Points with higher intensity than the neighbors called peaks are marked and the rest of the information is discarded. This is used to form a constellation. While matching the fingerprints, sample constellation is matched with database collection of constellations by using pattern matching algorithms. The disadvantage is large number of database entries should be searched. [5]

The SHAZAM system was developed for capturing cover audio pieces from millions of Western songs stored in the database, with the objective of tapping into the melodic construct of the song (devoid of other forms of embellishments). When applied to the Indian database the system was found less effective, due to subtle changes in both rhythm and melody mainly due to the semiclassical nature of Indian film songs. The retrieval accuracy was found to be 85 %. [6]

The method of Kalker and Haitsma [7] follows the Kurth's[8] approach of a subfingerprint design based on the global energy of each interval and uses a decomposition of the spectrum of each frame into bands using a logarithmic spacing and Hamming distance between the sequences of sub fingerprint of two audio files. However, the corruption of subfingerprints by noise and alterations corrupts the Hamming distances and reduces the amount of information that an indexation algorithm may deduce from such distances.

Philips Robust Hashing Algorithm uses Haitsma and Kalker's algorithm. In PRH the input, i.e. the audio signal is first divided into overlapping frames. FFT function is then applied to successfully obtain the power spectrum. The next step is the computation of the energies logarithmically spaced sub bands. From each frame, subfingerprints, or hash strings are then calculated.

In another method called Waveprint[9], a novel method for audio identification. Waveprint uses a combination of computer-vision techniques and large-scale data-stream processing algorithms to create compact _fingerprints of audio data that can be efficiently matched. The resulting system has excellent identification capabilities for small snippets of audio that have been degraded in a variety of manners, including competing noise, poor recording quality and cell-phone playback.

Another method is proposed in [10] as a spectro-temporal land marking approach to audio fingerprinting with the addition of a rank-ordering of local maxima that is being submitted to MIREX 2015 for the Audio Fingerprinting task. This system is named STELLAR (Spectro-TEmporaL LAndmarking with Rank ordering).This task involves database of ten thousand songs into a database of no more than 2GB taking no more than 24 hours. Having created a database, the system should identify approximately 6000 noisy queries within 24 hours.

V Conclusion

In this paper the concept of audio fingerprinting for content-based audio retrieval is introduced. The general framework for CBID and some of the methods for finger print extraction are discussed.

References :

[1] Santha Sinanan, " The Future is here: Query by Hummung as an example of content-based music information retrieval", Canada Library Student Journal, April 2010.

[2] Pedro Cano And Eloi Batlle, "A Review of Audio Fingerprinting", Journal of VLSI Signal Processing 41, 271–284, 2005.

[3] H.B. Kekre, PhD., Nikita Bhandari, Nisha Nair," A Review of Audio Fingerprinting and Comparison of Algorithms", International Journal of Computer Applications (0975 – 8887) Volume 70– No.13, May 2013

[4] G. Richly, L. Varga, F. Kov'acs, and G. Hossz'u, "Short-term sound stream characterisation for reliable, real-time occurrence monitoring of given sound-prints," in Proc. 10th Mediterranean Electrotechnical Conference, MEleCon, 2000.

[5] A. Wang, "The shazam music recognition service," Comm. of the ACM, vol. 49(8), 2006.

[6] S. Sri Ranjani , V. Abdulkareem, K. Karthik, P. K. Bora "Application of SHAZAM-Based Audio Fingerprinting for Multilingual Indian Song Retrieval", Advances in Communication and Computing Volume 347 of the series Lecture Notes in Electrical Engineering pp 81-92, 2015

[7] T. Kalker and J. Haitsma. "A highly robust audio fingerprinting system", In Proceedings of ISMIR'2002, pages 144–148, 2002.

[8] F. Kurth. "A ranking technique for fast audio identification". In Proceedings of the International Workshop on Multimedia Signal Processing, 2002.

[9] Michele Covell,Shumeet Baluja," Waveprint: Efficient Waveletbased Audio Fingerprinting", Pattern Recognition • November 2008.

[10] Toby Stokes. "Spectro-temporal landmarking with rank-ordered local maxima for audio fingerprinting",16th International Society for Music Information Retrieval Conference, 2015.

Author's Biography

K.Banuroopa, works as an Assistant Professor at Karpgam University. She has 10 years teaching experience. Currently, she is pursuing research degree in Karpagam University.

Dr. D. Shanmuga Priyaa, working as a Professor in the department of Information Technology at Karpagam University Coimbatore. She had completed Ph.D in Computer Science at Karpagam University in August2014 and she got 16 yrs of experience. At present she is guiding 8 Ph.D and 1 M.Phil scholars. She has published more than 15 international journals and more than 2 international conferences.