ISSN:2229-6093

Shweta Kharat et al, Int.J.Computer Technology & Applications,Vol 5 (2),662-668

Audio Fingerprinting and Review of its Algorithms

Saurabh Modani
Saurabh Pujari
Shweta Kharat
Mrs. Mugdha Shah
Department of Computer Engineering
Vishwakarma Insitute of Technology, Pune

Abstract

An audio fingerprint is a fixed format representation of an audio signal, which provides a digital summary of the signal. This fingerprint can then be used to identify features of an audio signal by matching it with existing database. Thus, metadata can be obtained from a small clip of a parent audio file. The audio fingerprinting process is basically a 2-stage process viz. identification stage and matching stage. The main aim is to use a robust algorithm to use certain level of distortion and compression to generate a compact fingerprint for efficient memory use and simplification of process.

KEYWORDS

Audio Fingerprint, Philips Robust Hashing Algorithm

1.0 Introduction

With the advent of the computer age, people are always on the look-out for new and efficient solutions to their basic computer problems. Nowadays, our personal computers have huge collections of multimedia files which may be downloaded from the internet, CDs or P2P networks. The media on most PCs is haphazardly maintained ,i.e., some files don't have appropriate file names , some files have metadata associated to them while some do not, some files have compressed or distorted audio information, etc. This is where audio fingerprinting provides us with a solution by identifying audio directly from the signal and not from associated metadata or file information. This technique enables tagging of unlabelled audio by providing just a part of that audio file.

Owners of multimedia content don't want unauthorized upload of their guarded content online. This content-verification can be realized by audio fingerprinting. Fingerprinting can be used to check if the content being processed is originally derived from the same source material. Audio, images as well as video can be used to obtain fingerprints. For protecting the piracy in contents, one can save the fingerprint of the required audio in database as the blacklisted content and thus the content cannot be shared. Fingerprint extraction and matching is used for that purpose.

1.1. Basic Requirements

For audio fingerprinting, the system has to meet following three requirements as follows [3]:

- **Robustness**: The fingerprint of a distorted piece of music has to be sufficiently close to the fingerprint of the undistorted recording.
- **Collision-resistance**: The fingerprints of two different pieces of music should be sufficiently different.

- **Database search efficiency**: In order to keep the database scalable, the fingerprint representation has to allow for efficient database search.

These requirements are primarily concerned with identification. To use fingerprints for indicating the quality (SNR) of compressed music, the fingerprinting system has to meet a fourth criterion: the distance between the fingerprints of the original and compressed version should also reflect the amount of compression.

2.0. Different Stages of Audio Fingerprinting

There are various stages in audio fingerprinting which are as follows[3]

1) Pre-processing: The audio file is given as the input to the algorithm. The pre-processing involves the conversion of audio signal into mono, filtering using a low-pass filter, and down sampled to a standard sample rate.

2) Framing and Overlap: Division of signal into strongly overlapped frames is done in this step. Overlapping is applied in order to ensure robustness to shifting.

3) Time-Frequency Transform: A particular set of characteristics is converted into new set of features using linear transform. Application of proper transform will reduce the redundancy. The most common transforms used are Fast Fourier Transform (FFT), Discrete Cosine Transform (DCT) etc.

4) Feature Extraction: This can be performed using various fingerprint algorithms. The audio signal is segmented into frames. For each frame a set of features is calculated.

5) Post-Processing: Each feature is then represented by a number of bits in the post-processing step. The compact representation of the time–frequency features of a single frame is called a sub-fingerprint. Due to the large overlap, subsequent sub-fingerprints are (strongly) correlated and vary slowly in time. The fingerprint of a song consists of a sequence of sub-fingerprints, which are stored in a database. A song-fragment is identified by matching a sequence of sub-fingerprints, called a fingerprint block, to the items in the database. A fingerprint block usually corresponds to several seconds of music

2.1. Properties of Audio Fingerprinting

The properties of audio fingerprinting algorithm are as follows [1],

- **Accuracy:** It is a measure of correctness of the results. The number of correct, wrong (false positives) and missed identifications.
- **Complexity**: It is the computational cost of fingerprint extraction which includes the size of the fingerprint, the complexity of the search algorithm used, the complexity of the fingerprint comparison or matching, the cost of adding new items to the database, etc.
- **Fingerprint rate (size)**: It is the amount of elements or bits extracted per second. The fingerprint size is directly related to the number of fingerprints that can be represented, and to the granularity. Larger the fingerprint rate, finer the granularity. It should be kept as small as possible.
- **Granularity**: The minimum length of the audio clip required for a dependable identification. The audio track is identified using a small fragment. Fine granular system should be approached for reliable audio identification using small excerpts.
- **Reliability**: This is of huge importance in role of fingerprint in copyright enforcement organizations. It is a measure of what level a fingerprint can be depended upon for its accurate matching.
- **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**: It is a measure of the performance with very large databases of titles or a large number of concurrent identifications. This affects the accuracy and the complexity of the system.
- **Security**: For applications where risk is high, fingerprint extraction should be dependent on

key. One then should not be able to change the content without changing the fingerprint.

- **Versatility**: It is the ability to identify audio regardless of the audio format and to use the same database for different applications.

3.0. Different Algorithms used for Audio Fingerprinting

1: Philips Robust Hashing Algorithm (PRH)
2: Multiple Hashing Algorithms (MLH)
3: Spectral Shape Descriptors (SSD)
4: Robust Audio Recognition System (RARE)

3.1. Philips Robust Hashing Algorithm

The Philips Robust Hash (PRH) algorithm is a well content-based audio identification technique.[1] The ability of the algorithm to give correct output in critical situations has been mathematically verified by taking into consideration the probability for bit error[7] or bit error rate(BER).The algorithm basically works in two steps[2,8]:-

1. Extraction of fingerprints: The input audio signal given to the algorithm is partitioned into overlapping frames having length of about 370ms with the frame shift equal to 1/32 of the frame length. The Fast Fourier Transform (FFT) [1, 9] is applied in order to get the power spectrum. After this the energy of 33 non-overlapping logarithmically spaced sub-bands whose frequency ranges from 300Hz to 2000Hz are calculated. The sub-fingerprints are computed from the sub-band energies in each frame.

2. Searching in the database: The computed sub-fingerprints of each audio file in the database are registered in hash table with the sub-fingerprints treated as the keys. In this phase of database searching, 256 sub-fingerprints each of which is approximately of 3 seconds are extracted from the query audio, and each sub fingerprint is compared to find match with the hash table contents to find the candidate position where it may come from. The existing fingerprint block which is of the same size as the query block from the candidate position obtained, BER between two blocks is found out and compared with a threshold which is given a fixed value of 0.35.Now the result depends on the BER. If the BER comes out to be less than the threshold, the two comparing signals are considered similar and the candidate audio is declared as the result. Diagrammatically PHR can be represented as [4]

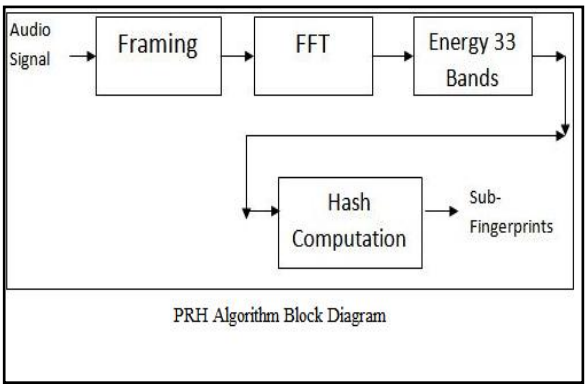


Fig.1 PRH Algorithm Block Diagram

3.2. Multiple Hashing Algorithms

In multiple hashing (MLH) method, the temporal sequence of energies in each sub-band is applied the DCT, resulting in the construction of sub-fingerprint for each DCT coefficient stream [4]. The application of DCT in MLH method has two reasons. First, taking into consideration all the orthogonal transforms, the de-correlation performance of DCT is closest to the Karhune-Loeve transform. Second, DCT has a strong energy compaction property stating that most of the signal energy tends to be concentrated in a few low-frequency components.

The MLH algorithm works in the following way:-

1. Division of audio into overlapping frames.
2. Application of Fast Fourier Transform.
3. Computation of energy.

The next step however, is performing DCT before determining the hash strings.

4. The temporal sequence of energies in each sub-band is applied DCT and for each DCT coefficient stream, a sub-fingerprint is generated.
5. The further computations of sub-fingerprints are performed only on the lower-ordered K values from the DCT coefficients.
6. The hash table records and stores the sub-fingerprint derived from the audio files in the database. Since K sub-fingerprints are computed for each frame, K hash

tables are constructed. Searching in database is performed in three steps:

a. The input or query audio is divided into 256 frames, and for each frame, K sub-fingerprints are obtained using the fingerprint extraction method.
b. The position of candidate is generated in each hash table, with the creation of the candidate list by compilation of all the search results in all included hash tables.
c. The BER is computed by comparing the query fingerprint block with those stored at the candidate positions in the database.

The output is the most hit candidate with BER less than the specified constant value say threshold.

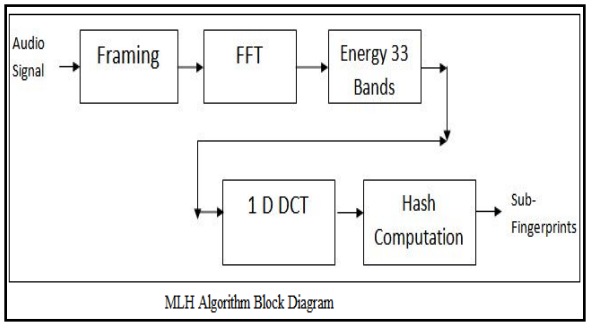


Fig.2 MLH Algorithm Block Diagram

3.3. Spectral Shape Descriptors (SSD)

Mapelli proposed an algorithm named SSD fingerprinting. In this algorithm, the power spectral density i.e. PSD's periodogram is used to extract the features [3]. This is determined from length windowed Fourier transform of the corresponding frame. The features USED are namely Mean energy, the SFM and the SCF. Feature is extracted separately for each sub-band. For this purpose, the arithmetic and the geometric mean of energies is used. Arithmetic and geometric mean is determined. Then the feature mentioned above like ME, SFM and SCF are determined periodogram. Each sub-band has a set of frequency bin indices. A 4 bit non uniform quantizer i.e. NUQ is used to quantize each feature in each band. The fingerprint is a result of the quantized level index of each feature of the above mentioned 3 features. MSE i.e. mean square error is used to find the distance between two fingerprint blocks.

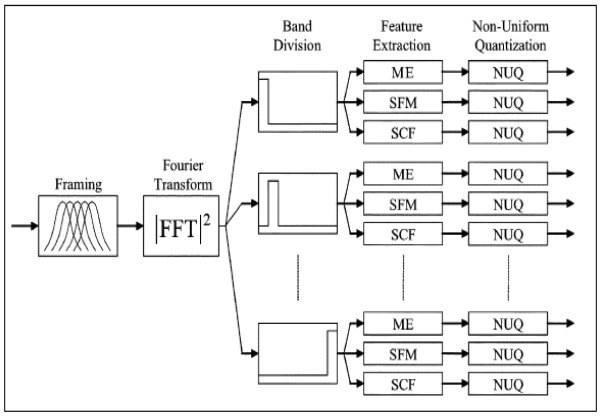


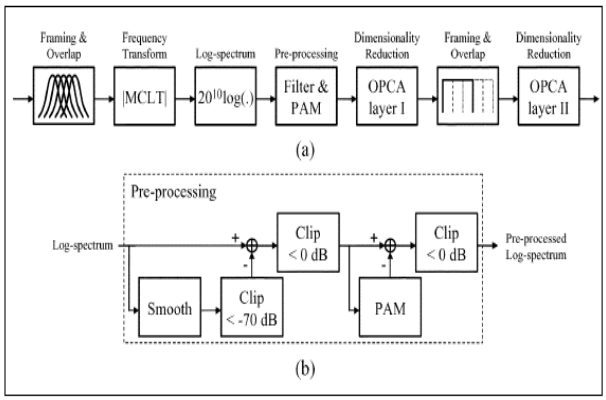
Fig.3 SSD Block Diagram

3.4. Robust Audio Recognition Engine (RARE)

For fingerprint extraction in RARE, log power spectrum of modulated complex lapped transform i.e. MCLT is used to represent data in time frequency domain [4]. Equalization effects removal and volume adjustment is done using the log power spectra. Non-audible frequency components from spectrum are removed using PAM i.e. psycho acoustic model in 2nd step.

Two stage projections of log power spectra is used for feature extraction. OPCA i.e. oriented principle component analysis is used for projection. It uses both undistorted and distorted data for a one-time, offline training purpose. Data is projected in those directions where the ratio of signal energy and distortion energy in the training data is maximized. Eigen values of covariance matrix are used to determine the directions. First OPCA projection uses preprocessed log-power spectra of the on training data. The 2nd one uses a number of concatenated, projected spectra from the first OPCA projection. The floating point representation of the trace of projected spectra makes the fingerprints. Euclidean distance i.e. root mean square is used to determine the distance between fingerprints.

a) Fingerprint extraction
b) Preprocessing



RARE Block Diagram

Fig.4 RARE Block Diagram

4.0. Comparison of PRH with MLH, SSD & RARE Algorithms

The overall recognition rate using PRH algorithm is 97.75% while it is 99.75 for MLH algorithm as mentioned by author in [4]. Thus using multiple hash tables, we get very accurate results in case of MLH. Although it is accurate result than PRH, its accuracy increases drastically with increase in number of hash tables. For that purpose, it requires much greater memory. But this can be avoided in PRH algorithm. Using PRH, we get better result without much wastage of memory. Thus it is not justifiable to use multiple hash table in MLH because we get very good results using PRH and not compromising.

The different features are used by PRH, SSD and RARE like sampling rate, frequency range, window length, frame overlap ratio, bit per feature, frequency bands, features and frames per segment[3]. Although all these features for all three algorithms are different, the fingerprints act as a function of SNR or compressed bit rate. The performance of identification phase changes with change in above mentioned features. This difference reflects changes in original recording and compressed version. The main hurdle for it is the variance in fingerprint difference. This variance is reduced in PRH by discarding certain unreliable bits in computing between two fingerprints as mentioned by author in [3]. For SSD and RARE algorithms, this problem is still an issue to be solved [3].

4.1. Audio fingerprinting applications.

Audio Content Monitoring and Tracking: Content distributors need to know if they have the rights to broadcast the content to the customers. In usage-policy monitoring applications, the goal is to avoid misuse of audio by consumer. A piece of audio is identified by the system by means of fingerprint and database is contacted for the rights' information.

Added Value Services: Content information can be provided to the user on the basis of user's preference. A musician will want to know the instruments being played in the recorded audio. Another user might want to know the name of the song, album name, year, live/playback, etc. A sound engineer would be more interested in the recording process information. This can be identified from the matching database entry of the extracted fingerprint, which is associated with the metadata. Different user profiles can be created providing dedicated information as and when required.

Duplicate identification can be done on large unsorted playlists, thereby, reducing memory requirement and unwanted cluttering.

Integrity Verification Systems: The integrity of audio recordings must be verified before usage, meaning the recording hasn't been modified or distorted too much. Commercials can be checked to be broadcast with the required length and quality. The verification that the suspected infringed recording is same as the recording whose ownership is known can be done using audio fingerprinting.

Watermarking Support: The system can assist in watermarking by deriving secret keys from the actual content. By using the fingerprint, the detector is able to locate anchor points and resynchronize at insertion/deletion attacks in perceptual hashing.

5.0. Conclusion

Different algorithms used for audio fingerprinting are PRH, MLH, SSD, RARE etc. are discussed above. The performances of algorithms are discussed in detail. PRH gives better results when memory cannot be

compromised as compared to MLH. Also the variance in fingerprint is reduced in PRH.

6.0. References

[1]. A Review of Audio Fingerprinting by PEDRO CANO AND ELOI BATLLE and TON KALKER AND JAAP HAITSMA in Journal of VLSI Signal Processing 41, 271–284, 2005

[2]. Modeling Audio Fingerprints: Structure, Distortion, Capacity [PhD thesis].

[3]. Distortion Estimation in Compressed Music Using Only Audio Fingerprints by Peter Jan O. Doets, Student Member, IEEE, and Reginald L. Lagendijk, Fellow, IEEE in IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 16, NO. 2, FEBRUARY 2008.

[4].Audio Fingerprinting Based on Multiple Hashing in DCT Domain by Yu Liu, Hwan Sik Yun, and Nam Soo Kim, Member, IEEE in IEEE SIGNAL PROCESSING LETTERS, VOL. 16, and NO. 6, JUNE 2009.

[5]. Comparison of Algorithms for Audio Fingerprinting by Heinrich A. van Nieuwenhuizen, Willie C. Venter and Leenta M.J. Grobler.

[7]. F. Balado, N. Hurley, E. McCarthy, and G. Silvestre, "Performance analysis of robust audio hashing," IEEE Trans. Inform. Forensics Security, vol. 2, no. 2, pp. 254–266, June 2007.

[8]. A Review of Algorithms for Audio Fingerprinting by Pedro Cano and EloiBatlle and TonKalker and JaapHaitsma.

[9]. Fast Fourier Transform and MATLAB Implementation by Wanjun Huang for Dr. Duncan L. MacFarlane.