Audio Signature for Song Identification

Dipali D. Bendale
PG Student: Department of Electronics and Telecommunication, Rajarshi Shahu College of Engineering, Tathawade, Pune
dipalibendale@gmail.com

Prof. Dr. Mrs. S. D. Apte
Professor: Department of Electronics and Telecommunication, Rajarshi Shahu College of Engineering, Tathawade, Pune
sdapte@rediffmail.com

Abstract- With the widespread use of internet and easy availability of mobiles, laptops, tablets etc. information is at finger tips to the users. Also with the developments in Audio, Video Engineering, the need for fast content identification from huge database, controlling illegal distribution of contents, broadcast monitoring etc. is demanding more and more attention. Growing illegal distribution of auditory content makes music copy-right owners and song distributors extremely concerned about their digital rights. The automatic detection of copy-righted songs has become a building block of many multimedia content sharing web-sites. Audio Signature plays an important role in these scenarios. Audio Fingerprinting is a form of audio signature that serves the purpose of content identification successfully.
Fingerprints are compact signatures of audio signals (or any other media) that can be used to distinguish between different songs based on their musical content.
This paper discuss about the implementation of audio fingerprinting using Fourier Transform. A hash value is computed based on the differences in energy of different frequency bands for every window. The block of these hash values is used to uniquely identify the song.

Keywords- *Audio Fingerprint, Energy differences, Bit derivation*

*****

I. INTRODUCTION

Imagine a situation you are traveling in a car and listening to Vividh Bharati. You have just missed the announcement about the song you are listening. You liked the song and you are curious to know about the details like the movie, singers etc. With your smart phone you can identify the song by simply giving the query of song for few seconds to the music identification service. This is possible with the use of audio fingerprinting.
An audio fingerprint refers to the unique, compact signature that summarizes an audio and is based on the contents of an audio. Short and unlabeled audio clips can be identified using the audio fingerprint in fast and reliable way. With audio fingerprint we can compare the two larger objects by comparing their small representations. Also audio fingerprinting makes identification possible even though the audio contents are corrupted with noise.

II. AUDIO FINGERPRINTING

A. Audio Fingerprint

Audio fingerprint is based on acoustic relevant characteristics of audio content. The fingerprint function maps the audio object that contains large number of bits to fingerprint with less number of bits.

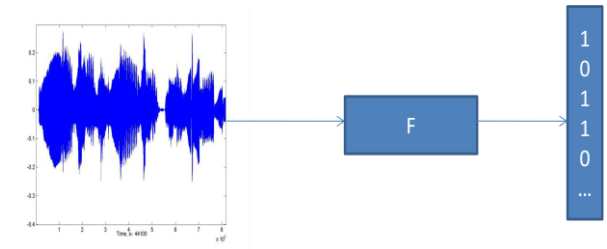

Figure1. Audio Fingerprint

B. Requirements of audio fingerprint

1. Accuracy
2. Invariance to distortions/Robustness
3. Discrimination power
4. Compactness
5. Computational Simplicity
6. Fast identification

C. General Framework

There are two basic processes:

1. Fingerprint extraction
2. Matching algorithm

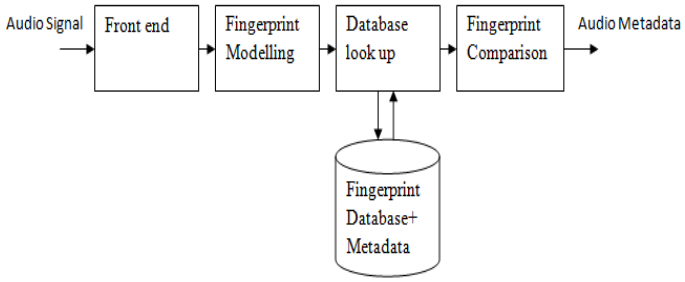

Figure2. Audio fingerprinting system

The front end computes different features from signal. The signal features are then modeled to represent the fingerprint in compact, unique way. A huge database is created of the extracted fingerprints. These fingerprints are stored with metadata information such as title, artist etc. When the query is played the matching algorithm compares the fingerprint with the fingerprints stored in database and identifies the audio and the corresponding metadata is displayed.

 ISSN: 2321-8169
1545 – 1549

III. PROPOSED TECHNIQUE

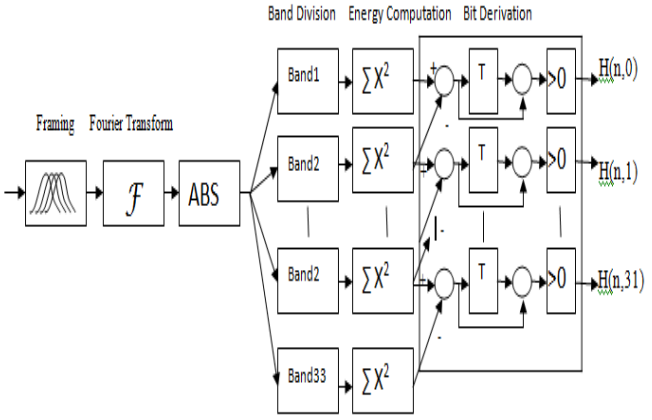

Figure3. Fingerprint generation

Preprocessing:

In this step, if the audio signal is stereo, it is converted to mono by taking an average of channels. The signal is down sampled to 5.5125 KHz and the silence part is removed. The signal is further preemhasized.

Framing and Windowing:

An audio signal is divided into frames so that to consider it as stationary for few milliseconds. A frame has a length of 2048 samples. We are using Hanning window function for minimizing the discontinuities at the beginning and end of frames. To assure robustness to shifting 90% overlap is applied.

$$w(n)=0.5\{1-\cos(2\pi n/N)\},\ 0 \le n \le N$$

Fourier Transform:

To transform the set of measurements to a new set of features the linear transform we have used is Fast Fourier Transform. We are using Fourier transform to find the frequency components of a signal in a noisy domain.

Band Division:

Sub-bands are formed by using the logarithmic spacing. We have used Mel scale and obtained 33 non-overlapping frequency bands in the range 300 Hz-2000Hz. The Mel scale uses triangular-shaped filters. The filter bank mimics the critical band, which represents different perceptual effects at different frequency bands. The edges are so placed that they coincide with the centre frequencies in adjacent filters.

Mel frequency= 2595× log (1+linear frequency/700)

Sub-fingerprint generation:

The energy differences among a set of sub-bands is taken to derive the bits of sub-fingerprint as given by the formula

$$F(n,m)=\begin{cases} 1 \text{ if } E(n,m)-E(n,m+1)-(E(n-1,m)-E(n-1,m+1))>0 \\ 0 \text{ if } E(n,m)-E(n,m+1)-(E(n-1,m)-E(n-1,m+1))<0 \end{cases}$$

Where F (n,m) denotes m-th bit of the sub-fingerprint of frame n and E (n,m) denotes energy of band m of frame n. Thus for 33 bands, 32bit sub fingerprint is generated for each frame.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A database is created of 40 mp3, Hindi songs, with stereo format, 16 bit depth and 44.1 KHz sampling rate for experimentation. Five segments of length 12s are cut to function as query in our audio fingerprinting system. Thus there are total 200 segments. The fingerprints are generated for these song segments. Each Fingerprint block contains 256 subsequent 32-bit sub-fingerprints. All these song segments are recorded using a microphone and total 200 fingerprints are generated for these recorded segments.

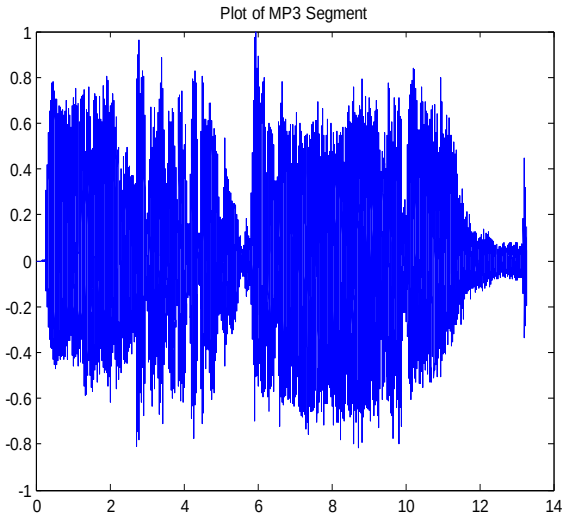

Figure4. Plot of MP3 segment of song

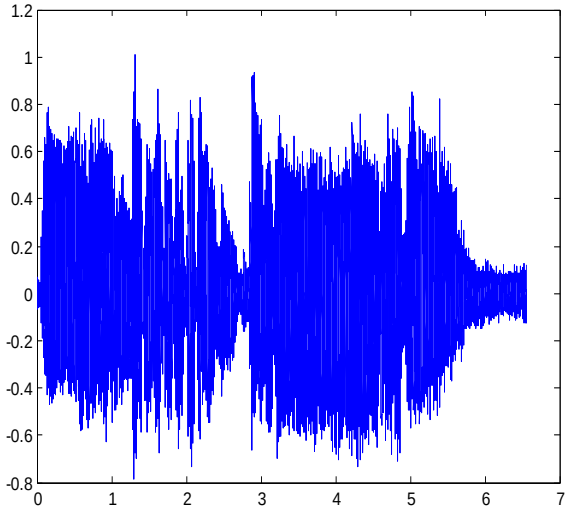

Figure5. Input audio signal with AWGN

 ISSN: 2321-8169
1545 – 1549

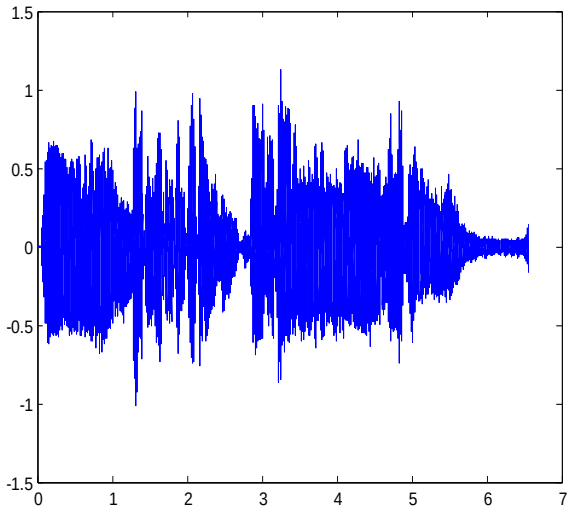

Figure6. Input audio signal with preemphasis

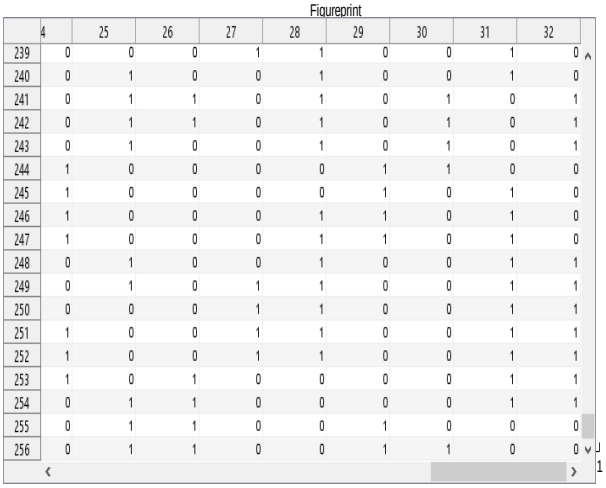

Figure7. Fingerprint bits (256*32)

Ranjhanna hua song segment1 Fingerprints

1. Fingerprint of song segment 2. Fingerprint for Recorded segment

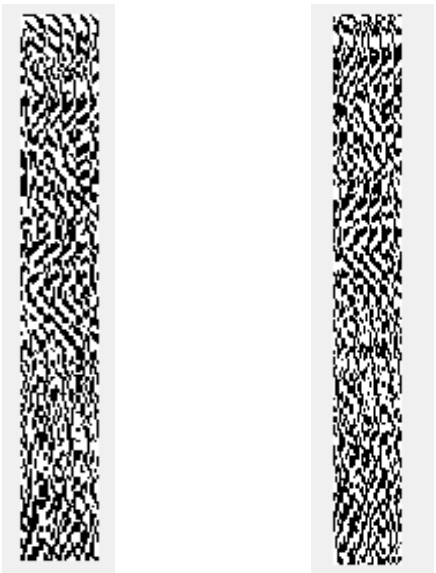

Figure8. Fingerprints

Hamming distance is calculated by comparing the query fingerprint with the fingerprints in database for analysis purpose.

0.0313	0.0625	0.125	0.125	0.1563
0.0625	0.0313	0.125	0.125	0.0938
0.125	0.0625	0	0.0625	0.0313
0.125	0.125	0.0625	0.0313	0.0938
0.125	0.0625	0.0313	0.125	0.0313

Table 1. Hamming distances between query fingerprints and song fingerprints in database for 5 Segments

Hamming distances for 100 song segments are analyzed further and Accuracy, Sensitivity, Specificity and Precision are calculated on the basis of threshold.

Sr. No.	Parameter	Percentage
1	Accuracy	99.81
2	Sensitivity	95
3	Specificity	99.86
4	Precision	87.15

Table 2. Recognition performance analysis

The query is played from the mobile and given to the microphone. Fingerprint is calculated for this query and it is compared with the stored fingerprints in database. For the matching fingerprint the minimum Hamming distance is less than the threshold and the metadata is displayed for that song. It includes Title, Album, Singer, Lyrics and Music information.

CONCLUSION

In this paper we have discussed the general framework of audio fingerprinting system and the implementation based on Fast Fourier Transform. Since HAS operates on logarithmic scale we have taken the logarithmic spacing for band division and computed the hash values based on energy difference of bands to ensure the robustness. Further work can be done to improve the robustness of the system for song identification adding the noise data to the query.

V. REFERENCES

[1] J. Haitsma, and T. Kalker, “A highly robust audio fingerprinting system”, International Symposium on Music Information Retrieval (ISMIR), pp. 107-115, 2002.

[2] W. Son, H. Cho, K. Yoon and S. Lee, "Sub-fingerprint Masking System in a Real-noise Environment for Portable Consumer Devices " IEEE Transactions on Consumer Electronics,Vol.56,No. 1, February 2010.
[3] Avery Li-Chun Wang, "An Industrial-Strength Audio Search Algorithm", ISMIR, 2003.
[4] S. Baluja and M. Covell, "Audio fingerprinting: Combining computer vision and data-stream processing", IEEE International Conference on Acoustics, Speech and Signal Processing, 2007
[5] M. Davidson Kamaladas and M. Maxina Dialin, "Fingerprint Extraction of Audio Signal using Wavelet Transform', ICSIPR, 2013.
[6] Dr. S. D. Apte, "Speech and Audio Processing", Wiley-India, 2012.
[7] Bachu R.G., Kopparthi S., Adapa B. and Barkana B. D., "Separation of Voiced and Unvoiced using Zero crossing rate and Energy of the Speech Signal", ASEE 2008.
[8] Carlo Bellettini and Mazzini, "A Framework for Robust Audio Fingerprinting", Journal of Communications, Vol. 5, No. 5, May 2010.
[9] Vijay Chandrashekhar, Matt Sharifi and David Ross, "Survey and evaluation of audio fingerprinting schemes for mobile query-by-example applications", ISMIR, 2011.
[10] H. B. Kekre, N. Bhandari and N. Nair, P. Padmanabhan and S. Bhandari, "A Review of Audio Fingerprinting and Comparision of Algorithms", International Journal of Computer Applications, vol. 70-No.13, may 2103.
[11] Pedro Cano, Eloi Ballle, Ton Kalker and Jaap Haitsma," A Review of Algorithm for Audio Fingerprinting", Multimedia Signal Processing, 2002 IEEE workshop.
[12] P. Cano, E. Batlle, T. Kalker, and J. Haitsma, "A review of audio fingerprinting," J. VLSI Signal Process. Syst., vol. 41, pp. 271–284, November 2005.
[13] http://www.ee.columbia.edu/ln/rosa/matlab/mp3read.m