IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
e-ISSN: 2278-2834,p- ISSN: 2278-8735.
PP 30-34
www.iosrjournals.org

Audio Fingerprinting using Fractional Fourier Transform

Swati V. Sutar[1], D. G. Bhalke[2]
[1](Department of Electronics & Telecommunication, JSPM's RSCOE college of Engineering Pune, India)
[2](Department, of Electronics & Telecommunication, JSPM's RSCOE college of Engineering Pune, India)

ABSTRACT: *The An audio fingerprint is popular nowadays due to its compact size and faster identification of song. When we like the song being played on the radio and do not have information about the song, then capture the fraction of currently playing song and input this audio clip to music identification system. This system generates an audio fingerprint of input song and compares with the audio fingerprints stored in the database. After finding the best match, it displays the information regarding the song and plays back the song. Here, Fractional Fourier Transform is used to generate an audio fingerprint. Audio fingerprints are made using another two different methods: FFT and MFCC. For the generation of an audio fingerprint, an audio clip is given as an input query. These generated fingerprints compared with the fingerprint stored in the database. After finding the best match, it displays metadata of song and playback the song.*

Keywords *- Audio Fingerprint, Energy Differences, Fractional Fourier Transform*

I. INTRODUCTION

An audio fingerprint system is a content based compact signature that summarizes an audio recording. An audio fingerprint is useful for establishing the perceptual equality of two audio objects by comparing the associated audio fingerprints. These stored in the database along with their metadata [singer name, film] for a large number of songs. This system is useful to get the information regarding songs and plays back the song. An audio fingerprint system consists of different steps: pre-processing, feature extraction, fingerprint database and matching of fingerprints. The feature extraction plays a vital role, so these can be selected which is robust to signal degradation. Different methods have been used to generate an audio fingerprint like J. Haitsma, and T. Kalker proposed a highly robust audio fingerprint system in which 32- bit fingerprint extracted [1]. E. Unal et.al developed a system for audio fingerprint using query by humming [2].

In this paper, an audio fingerprint is generated using Fractional Fourier Transform. The rest of the document arranged as follows: Section 2 presents the proposed method. Results of audio fingerprinting discussed in Section 3. Section 4 followed by references.

1.1 Requirements of audio fingerprinting

1. **Invariance to distortion:** Audio fingerprint should identify an audio signal in the presence of noise, distortion and compression.
2. **Compactness:** Due to the smaller size of audio fingerprints a large number of audio fingerprints stored in the database.
3. **Computational Simplicity:** The extraction of fingerprint should be within less time.
4. **Faster Identification:** An audio fingerprint must identify the song within less time. It depends on how many seconds of audio is required to identify the song.

1.2 PARAMETERS OF AUDIO FINGERPRINTS

Robustness: The system must be robust to identify the audio correctly in the presence of noise, distortion. It can determine an audio clip accurately, regardless of the level of compression. Robustness is measured in terms of bit error rate.

BER=error bit/total bit

Reliability: This measures in terms of how often the song identified incorrectly.

Fingerprint Size: Size of fingerprint should be small and stored in RAM memory and expressed in the form of bits per second or bits per minute.

Granularity: This explains how many seconds of audio required for identifying the song. It depends on an application like giving a full audio signal or fragment of an audio clip.

Search Speed: It measures how much time required for finding the fingerprint with the fingerprint stored in the database. Search speed should be in the order of milliseconds [1].

1.3 APPLICATIONS OF AUDIO FINGERPRINT:

1. Broadcast Monitoring:

It refers to the automatic playlist generation of radio, television or web broadcasts for the purposes of royalty collection, program and advertisement verification and people metering. A large-scale broadcast monitoring system based on fingerprinting consists of several monitoring sites and a central location where the fingerprint server located. At the monitoring sites, fingerprints extracted from all the (local) broadcast channels. The primary site collects the fingerprints from the monitoring sites.

2. Connected Audio:

It used for consumer applications where audio is connected to additional and supporting information. An audio signal degraded due to FM/AM transmission, an acoustical path between loudspeaker and microphone of a mobile phone, speech coding and transmission over a cellular network. So, this is the challenging application.

3. Automatic Music Library Organization:

Many PC users have several songs library collection in hundreds or thousands. Music stored in a compressed format like MP3. These songs obtained from different sources like download from the internet, transfer using Bluetooth, ripping from a CD. So, songs library are not well organized. The metadata is always incomplete, inconsistent, so by using audio fingerprint the metadata of songs library arranged in a proper manner [1].

II. PROPOSED METHOD

Audio fingerprint is generated using different methods like FFT, DCT, and Wavelet. In this paper, for generation of an audio fingerprint, Fractional Fourier Transform (FrFT) is used. An audio fingerprint system consists of fingerprint extraction and matching of fingerprints. The input to the system is an audio signal in the waveform. The output is metadata of song and plays back the song. The proposed audio fingerprint system consists of audio fingerprint extraction and matching of fingerprints. The audio fingerprint extraction consists of pre-processing, framing, overlapping, windowing, feature extraction and fingerprint modeling. Matching of fingerprints includes searching fingerprints in the database and when the best match occurs it displays the meta data and plays back the song.

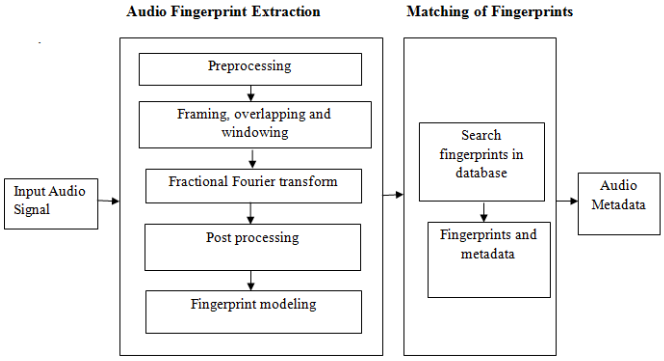

Fig.1] Proposed diagram for Audio Fingerprint Extraction

The following steps performed to generate an audio fingerprint.

2.1 Preprocessing:

This step is useful to remove silence part of the audio signal. The energy feature is used to remove silence part of an audio signal. This signal is used to down sample to a standard rate; here down sampling rate is used as 8 KHz so, that data rate becomes smaller which is useful for fingerprint generation.

2.2 Framing, Overlapping, and Windowing:

Framing is important to keep signal stationery as an audio signal is non-stationary. The signal is considered to be stationary for few milliseconds; here it is considered as 20 ms. Overlapping is used to preserve statistical properties, so for robustness to shifting considered more overlapping. Here, overlapping is of 90%. Each frame has been multiplied with Hamming window to keep continuity in first and last point in the frame. This window used because of its low sideband attenuation. The Hamming window W(n) is given as:

$$W(n)= 0.54-0.46 \cos (2\pi n/N) \qquad (1)$$

2.3 Fractional Fourier Transform:

It is the generalization of classical Fourier transform. FrFT belongs to the class of time-frequency representation. It is represented by using two orthogonal planes; one plane represented by time axis x(t), and another plane is frequency axis represented as X(w).Fourier transform F[x(t)]=X(w) is a counter-clockwise rotation of time axis to $\pi/2$ radian to the frequency axis. Like FT operator, FrFT is a linear operator that corresponds to rotation of a signal by an angle α that is not multiple of the $\pi/2$ radian. The signal rotated by an angle α. If $\alpha=0$, FrFT corresponds to zero rotation and $\alpha=1$, it corresponds to the $\pi/2$ rotation that is equivalent to Fourier transform. 'a' is the fractional order lies between 0 to 1 and signal represented by time and frequency plane.

FrFT is represented by F^{α}. FrFT has following properties:

1. F^{0}= I, zero rotation or identical operator
2. $F^{\pi/2}$ =F, Corresponds to Fourier transform operator.
3. $F^{2\pi}$ =I, 2π rotation
4. F^{α} F^{β} =$F^{\alpha+\beta}$, additivity of rotation

FrFT of a signal x (t) with order α given by $F\alpha(u)$ which represented by the plane,

$$F^{\alpha}(u) = \int_{-\infty}^{\infty} x(t) K_{\alpha}\ (t,u)\,dt \qquad (2)$$

Where K^{α} (t, u) is a transformation kernel and is represented by,

If α is not multiple of 2π (3)

$$\sqrt{1 - j\cot\alpha / \sqrt{2\pi}}\, e^{jt^2 + u^2/2\cot\alpha - jut\csc\alpha}$$

The 'α' makes an angle with a time axis and is called fractional order. As 'a' changes from 0 to 1, FrFT changes the signal from time domain (α=0) to the frequency domain ($\alpha=\pi/2$) [3]. So, α provides an additional degree of freedom and flexibility for processing of audio signals. The orthonormal basis function of FrFT is linear chirps, so it is more suitable for non-stationary signal analysis. It decomposes a signal in terms of chirp signals and more suitable for analysis of audio signals. FrFT is useful to represent the signal in the multiple domain that is time and frequency domain, so it captures the dynamic behavior of audio signals. Results checked from α=0.90 to 0.99. The better outcomes observed at angle α= 0.95 which selected for consideration because of better accuracy.

2.4 Band Division and Energy Computation:

The power spectrum of the signal is multiplied by magnitude response of the set of 33 triangular band pass filters and in the range 300Hz-2000Hz. Sub-bands are formed by using the logarithmic spacing. These band pass filters equally spaced along the Mel frequency, which is related to the standard linear frequency f by the following formula:

$$Mel\ (f) = 2595 * \ln(1+f/700) \qquad (4)$$

Mel frequency is proportional to the logarithm of linear frequency and which is close to the human perceptual system [4].

2.5 Sub fingerprint generation using Fractional MFCC:

The steps included for Fractional MFCC are Pre-processing, Framing, Overlapping, Windowing, Fractional Fourier Transform, Mel filter bank, Log, DCT and Fractional MFCC.

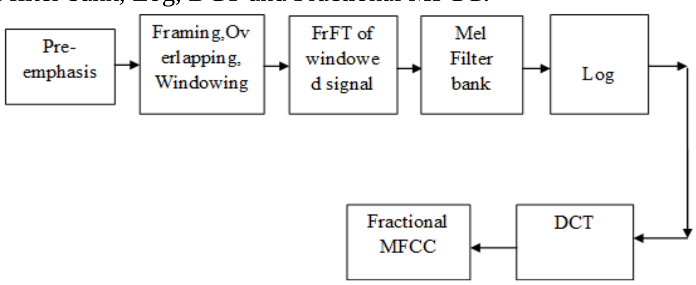

Fig.2] Calculation of Fractional MFCC coefficients

These frequency bands lie in the range from 300 Hz-2000 Hz and have logarithmic spacing. By selecting 33 MFCC coefficients, 32- bit sub-fingerprint value is extracted for every frame. The 32- bit sub-fingerprint obtained by taking the energy band differences. The equation is:

$$F(n,m)=\begin{cases} 1 & \text{if } E(n,m)-E(n,m+1)-(E(n-1,m)-E(n-1,m+1))>0 \\ 0 & \text{if } E(n,m)-E(n,m+1)-(E(n-1,m)-E(n-1,m+1))\leq 0 \end{cases} \quad (5)$$

Where,

$E(n,m)$ denotes band m-th energy of frame n, and $F(n,m)$ denotes the m-th bit of the sub-fingerprint of frame n. So, for 33 coefficients, 32 bit sub-fingerprint is generated for each frame. A fingerprint block consisting of 150 subsequent fingerprints have been taken. The particular song has many fingerprint blocks based on length of the song.

2.6 Matching of Fingerprints:

For matching of fingerprints, first give a query signal to a system that generates an audio fingerprint. The generated fingerprint matched with the fingerprint stored in the database. Matching is done based on Hamming distance and when best match found it displays the corresponding metadata and plays back the song.

III. Experimental Results:

For the generation of an audio fingerprint, 20 Hindi wave format songs are used. An audio clip of 3 seconds for each song taken into consideration. These segments are depending on the length of a song. So, 2075 fingerprints have been generated for 20 songs using Fractional MFCC.

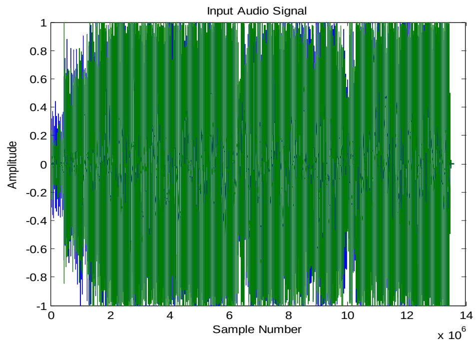

Fig.3] Plot of Original signal of song Mashallah

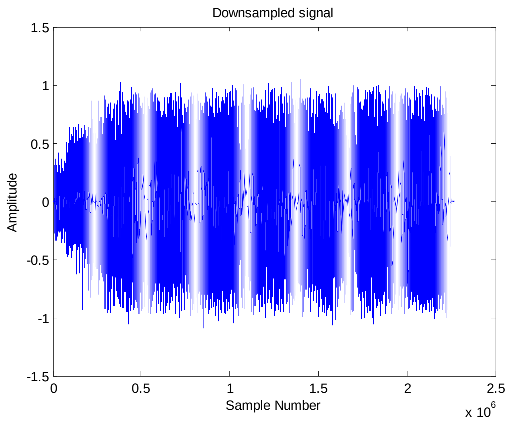

Fig.4] Plot of Down sampled signal

	20	21	22	23	24	25	26	27	28	29	30	31	32
133	1	1	0	1	1	0	1	0	1	0	1	0	0
134	1	0	1	0	1	1	1	0	0	0	0	1	1
135	0	1	0	1	0	0	0	1	1	1	1	0	1
136	1	0	1	1	0	0	0	1	1	1	0	0	1
137	0	1	0	0	1	1	0	1	0	0	1	1	0
138	1	0	0	1	0	0	1	0	1	0	1	1	0
139	1	1	1	0	1	1	0	1	0	0	1	0	1
140	0	0	0	1	0	0	1	0	1	1	1	1	0
141	1	1	1	0	1	1	0	1	0	0	1	0	1
142	1	0	0	0	1	1	0	1	0	0	1	1	0
143	0	0	1	1	0	0	0	1	1	1	0	1	1
144	1	1	0	0	1	1	1	0	0	0	1	0	1
145	0	0	1	1	0	0	0	1	1	1	0	1	0
146	0	0	0	1	0	0	0	1	1	0	0	1	0
147	1	1	1	0	1	0	0	1	1	0	1	0	1
148	1	1	1	0	1	1	1	0	0	0	0	0	1
149	0	1	1	1	1	0	1	0	1	0	0	1	0
150	1	0	0	0	1	1	1	0	0	0	1	0	0

Fig.5] Fingerprint bits (150*32)

Fig.6] Fingerprint of song segment of Mashallah (150*32)

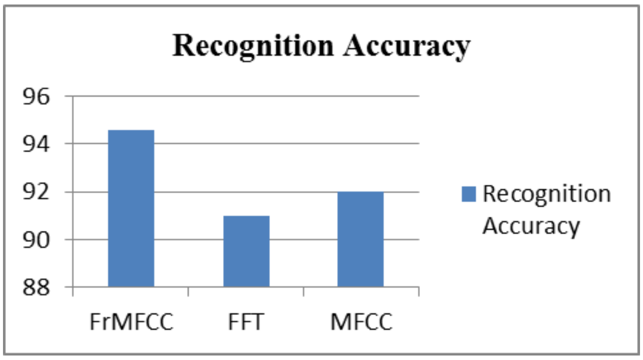

Fig.7] Comparison between proposed method and conventional method

IV. CONCLUSION

It has observed that Fractional Fourier Transform is highly useful for non-stationary signals since it provides a great degree of freedom and flexibility. FrFT captures dynamic time-varying nature of audio signals because of chirp functions, so it gives better accuracy and lesser size of the fingerprint.

V. Future Work

The future work can include, making a robust system by adding noise to query signal for song identification, which would closely approximate real-world situations.

REFERENCES

[1]. J.Haitsma and T. Kalker, "A highly robust audio fingerprinting system," in Proc. Int. Conf. on Music Information Retrieval, pp. 107-115, 2002.
[2]. E.Unal et al., "Challenging uncertainty in query by humming systems: a fingerprinting approach," IEEE Transactions on Audio, Speech and Language Processing, vol. 16, No. 2, pp. 359-371, 2008.
[3]. D.G.Bhalke, C.B.Ramarao and D. S. Bormane, "Automatic musical instrument classification using Fractional Fourier Transform based-MFCC features and counter propagation neural network," Journal of Intelligent Information System, Springer International Publishing, DOI: 10.1007/S10844-015-0360-9.
[4]. Dr.S.D.Apte, Speech and Audio Processing, Wiley-India, 2012.
[5]. P. Doets and R. Lagendijk, "Distortion estimation in compressed music using only audio fingerprints," IEEE Transactions on Audio, Speech and Language Processing, vol.16, No. 2, Feb. 2008.
[6]. W. Son, H. Cho, K. Yoon and S.Lee, "Sub-fingerprint Masking for a Robust Audio fingerprinting System in a Real noise Environment for Portable Consumer Devices, "IEEE Transactions on Consumer Electronics, vol.56, No.1, February 2010.
[7]. S.Lee, D.Yook and S.Chang, "An efficient audio fingerprint search algorithm for music retrieval," IEEE Transactions on Consumer Electronics, vol.59, No.3, August 2013.
[8]. Y. Liu et.al, "DCT based multiple hashing technique for robust audio fingerprinting," IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 61-64, 2009.
[9]. C. Belletini and Mazzini, "A framework for robust audio fingerprinting," Journal of Communications, vol.5, No.5, May 2010.
[10]. H. Schreiber and M.Muller, "Accelerating index based audio identification," IEEE Transactions on Multimedia, vol.16, no.6, pp. 1654-1664, Oct. 2014.