A Robust Music Retrieval System Using Salient Features

P. Dhanalakshmi(Associate professor)[1], R. Thiruvengatanadhan(Assistant professor)[2], S. Jayalakshmi(PG scholar)[3]
[1, 2, 3] Department of CSE
[1, 2, 3] Annamalai University

Abstract- *A wide variety of music audio ranging from CD's, MP3 Players to World Wide Web (WWW) are available online for public access. Managing and organizing such huge databases is a complex task. Hence music audio indexing and retrieval system is essential. Automatic music indexing and retrieval is a broad area of research. Acoustic features representing the music are extracted from music audio. The goal of music indexing and retrieval system is to provide the user with capabilities to index and retrieve the music archive in an efficient manner. Acoustic feature namely sonogram features are used to create the index. Retrieval is based on the highest probability density function. In this paper, we propose a method for indexing and retrieval of the music using Gaussian mixture models. Performance of the system is evaluated in terms of average number of clips retrieved and accuracy of retrieval. The system showed satisfactory results in terms of both the performance measures.*

Keywords- Acoustic Feature, Indexing and Retrieval, Gaussian Mixture Model and Probability density function.

I. INTRODUCTION

In today's world, Digital audio application is an important part in our everyday life. CDs, MP3 audio players, radio broadcast, TV, telephone, telephone answering machines and word recognition are popular examples. The voice, music and various kinds of environmental sound of audio is important type of media and also a significant part of video. The different kind of audio signals such as music, speech are indicated by term audio. File systems and multimedia databases have thousands of audio recordings. Audio is treated as opaque collection of bytes with primitive fields attached; namely, file format, name, sampling rate etc. In order to compare and classify the data efficiently the meaningful information are extracted from digital audio wave forms. The extracted information is stored as content description in a compact way [4]. These compact descriptors are not only used in audio storage and retrieval application but also in efficient content based segmentation, indexing, classification, and recognition and browsing of data. A data descriptor is often called as feature vector. The process of extracting such feature vector from audio is called audio feature extraction. To feature one piece of audio data a variety of more or less complex descriptors are extracted. Depends on the application, the particular efficient feature is used for compression and classification.

Music is form of art and entertainment that puts a sound together in way people like or finds interest to listen. The most form of music will be either instrumental or voice of people. The human voice was the first instrument used by the humans which produce a different kind of sound. The music audio retrieval is done by annotating the media with text. Since the music information is voluminous, retrieving of music in form of text is the tedious job. The text may not appropriate way to express huge information contained by audio [11, 12].

II. MUSIC INFORMATION RETRIEVAL

In the field of audio processing Music Information Retrieval (MIR) has a rapid growth [9]. Search engines like goggle, yahoo provide facilities to retrieve an audio for music fans and their reputations shown an exponential growth. In World Wide Web, to satisfy the requirements of music fans some of the websites are professionally maintained. Search engines are simple they were developed on the basis of text query.

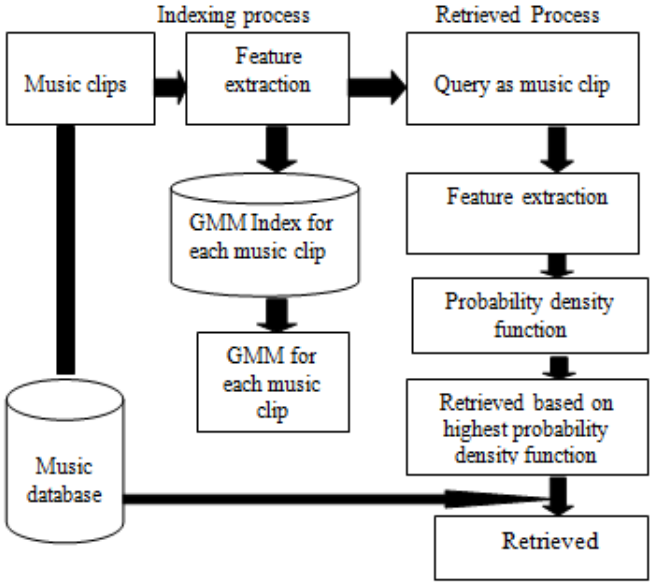

Figure 1: Proposed Methods for Music Indexing and Retrieval System.

Moreover, the difficulties in search of similar music or songs are realized by this system. If suppose the textual information, is incomplete the several difficulties been occur in satisfying the specific need of applications. In this paper the GMM is used for creating the index and retrieval is based on highest probability density function. The proposed method for music indexing and retrieval has been shown in Figure 1. The recorded music audio is taken as the input for music audio retrieval by query form then from a given music collection it automatically retrieves as a part or similar to it.

III. ACOUSTIC FEATURE EXTRACTION

A. Sonogram

Music follows certain rules that have been highly structured and provide strong regularities. In these work, acoustic features namely sonogram feature is extracted for indexing and retrieval.

The algorithm for extracting the Sonogram is as follows: Audio segment is transformed into spectrogram representation using Fast Fourier Transform (FFT) with hamming window (23 ms windows) and 50% frame overlap. Bark scale is applied by grouping frequency bands into 24 critical bands. Apply spreading function to account for spectral masking effects. Spectrum energy values on the critical bands transformed into decibel [db]. Calculate specific loudness levels through incorporating equal-loudness contours [phon]. Compute specific loudness sensation per critical band [sone]. For each segment the spectrogram of the audio is computed using the short time Fast Fourier Transform (STFT).The sonogram feature extraction has been shown in Figure 2.

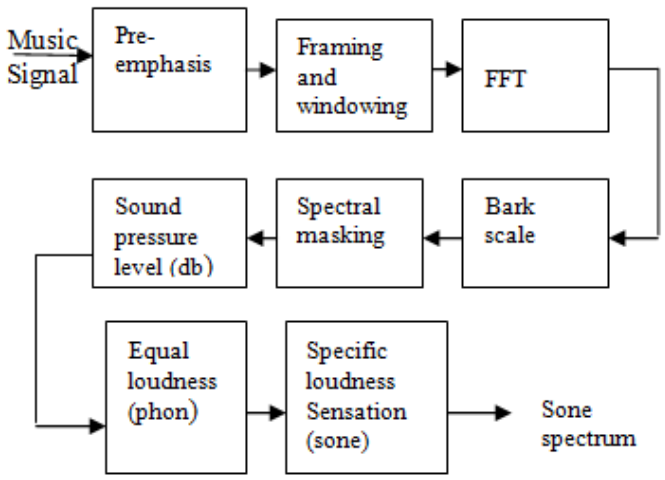

Figure 2: The sonogram feature Extraction.

The Bark scale, a perceptual scale which groups frequencies to critical bands according to perceptive pitch regions, is applied to the spectrogram, aggregating it to 24 frequency bands. A spectral Masking spreading function is applied to the signal, which models the occlusion of one by another sound. The Bark scale spectrogram is then transformed into decibel scale. Further psycho-acoustic transformations are applied: Computation of the Phon scale incorporates equal loudness curves, which account for the different perception of loudness at different frequencies. Subsequently, the values are transformed into unit sone, reflecting the specific loudness of the human auditory system. The sone scale reflects to the Phon scale in the way that a doubling on the Sone scale sounds to the human ear like a doubling of the loudness.

IV. TECHNIQUES FOR MUSIC INDEXING AND RETRIEVAL

A. Gaussian Mixture Model

The distribution of feature vectors is classified as two types, parametric and non-parametric. Model that assumes the shape of the probability density function is termed as parametric. In non parametric minimal or no assumption is made regarding probability density function.GMM basic concept is distribution of feature vectors that extracted from class can be modelled by a mixture of Gaussian densities[11].The Gaussian mixture model been shown in Figure 3.

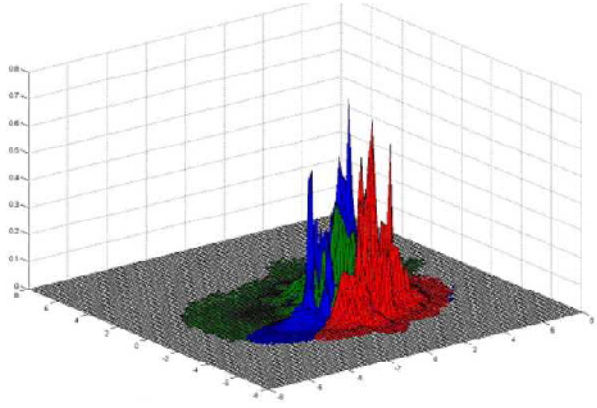

Figure 3: Gaussian mixture models.

The feature vector of GMM's is represented by Gaussian components and they are characterized by the mean vector and covariance matrix. GMM models have the capability to from an arbitrarily shaped observation density. For a D dimensional feature vector x, the mixture function for category s is defined as,

$$p\left(\frac{x}{\lambda^s}\right)^n = \sum_{i=1}^{M} \alpha_i^s f_i^s(x) \qquad (1)$$

The mixture density function is a weighted linear combination of M component uni-modal Gaussian densities $f_i^s(.)$. The Gaussian densities function $f_i^s(.)$ is categorized by mean vector μ_i^s and covariance matrix $\sum_i^s$ using,

$$f_i^s(x) = \frac{1}{\sqrt{(2\pi)^d \left|\sum_i^s\right|}} \exp\left(-\frac{1}{2}\left(x-\mu_i^s\right)^T \left(\sum_i^s\right)^{-1}\left(x-\mu_i^s\right)\right) \qquad (2)$$

Where, $\left(\sum_i^s\right)^{-1}$ and $\left|\sum_i^s\right|$ denote the inverse and determinant of the covariance matrix $\sum_i^s$, respectively. The iterative Expectation Maximization (EM) algorithm is used to estimate the parameters of GMM [10].

Most popular clustering algorithms called EM algorithm used to estimate the probabilistic models for each Gaussian component. The Expectation (E-step) and Maximization (M-step) are iterated till the convergence of the parameter. EM algorithm finds out maximum EM algorithm finds out Maximum likelihood of parameter.

V. PROPOSED METHOD FOR INDEXING AND RETRIEVAL OF MUSIC CLIPS

The algorithm for indexing and retrieval of Music clips is described below:

A. Algorithm for Indexing of Music

Step 1: Collect 100 music clips $m_1, m_2, \ldots\ldots\ldots m_{100}$ each of 5 seconds duration from songs collected from musicbrainz database.
Step 2: 13-dimensional Sonogram features are extracted from all 100 music clips to form the Music index.
Step 3: A GMM is fit for all 100 music clips using the Sonogram features.
Step 4: The GMMs of the extracted features form the index.

B. Algorithm for Retrieval of Music Using Index

Step 1: A music query clip of 2 seconds duration is extracted from the music wave file.
Step 2: Sonogram features are extracted from the music query.
Step 3: The probability of the query feature vectors belonging to the 100 GMMs is computed.
Step 4: The maximum probability density function corresponds to the music query audio is computed.
Step 5: The music clips which have the maximum probability density function is retrieved.

5.1 Performance Measures

A. Accuracy of Retrieval

Performance measure for music audio indexing system is based on the accuracy of retrieval. It is measured using Equation (3).

$$Accuracy\ of\ retrieval = \frac{No\ of\ query\ fv\ correctly\ retieved}{Total no\ of\ fvuesd\ for\ testing} X\,100 \qquad (3)$$

Where fv denotes the feature vectors.

B. Average Number of Clips Retrieved for Each Query Based on a Threshold

The music audio retrieval performance is measured by average number of clips retrieved for each query based on a predefined threshold as shown in the Equation (4).

$$Average\ no.of\ clips\ retrieved = \frac{No.of\ clips\ retrieved\ for\ ecah\ query}{Total\ no.of\ qureies} \qquad (4)$$

VI. EXPERIMENTAL RESULTS

A. Database for Music Indexing

The experiments are conducted for indexing the music audio using the database collected from the online called music brainz. Music audio of 5 minutes duration is taken from online. 100 different complete song form a total dataset, 5 seconds of duration is extracted from each song, which is sampled at 22 KHz and encoded by 16 bit.

B. Acoustic Feature Extraction

The process of converting an audio signal into a sequence of feature vectors is termed as feature extraction; it carries characteristic information about the signal. These vectors are used as basis for various types of audio analysis algorithms. In this paper, sonogram feature are extracted for each music audio clip of 5 seconds. A frame size of 20 ms and a frame shift of 10 ms are used. The 22 sonogram features are extracted for each music audio clips of 5 seconds. Thereby 500 features vector are obtained for a clip of 5 seconds

duration. Hence, 500 x 22 feature vectors are arrived at for each of the 5 second music clip and this procedure is repeated for all 100 clips.

C. Creation of Index

For 100 music clips GMMs are constructed using sonogram feature it form the index. . Experiments are also conducted with GMMs using sonogram features to create index.

D. Retrieval of a Music Using Index

For retrieval, from music clip the last 2 seconds is used as query. For every frame in the query the probability density function that the query feature vector belongs to the first Gaussian is computed. The same process is repeated for all the feature vectors.

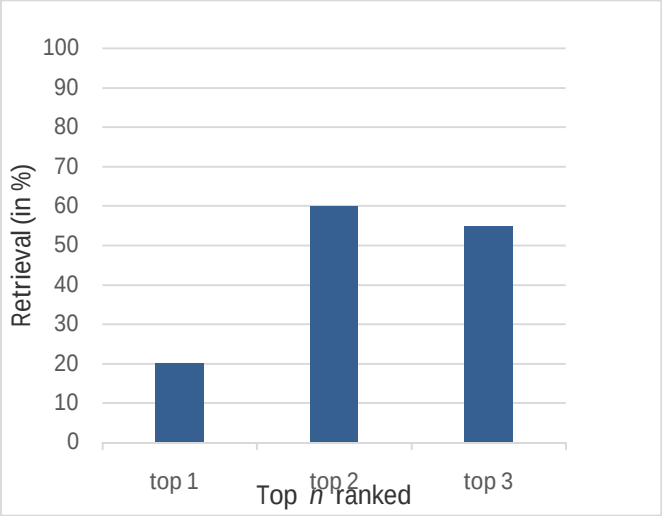

Figure 4: Accuracy of Retrieval of Music Clips in the Top *n* Ranked List

The average probability density function is computed for the first Gaussian. Similarly the average probability density function for all the Gaussians is computed. Retrieval is based on the maximum probability density function. Figure 4 shows the percentage of retrieval in top ranked list.

The GMM is used to capture the distribution of features namely sonogram features. The performance of GMM for different mixtures is shown in Figure 5 Various experiments for different mixtures are carried out and the retrieval is based on the highest probability density function. With GMM, the performance is found to increase as the mixture increased from 1 to 3 and the optimal performance is achieved with 3 mixtures. When mixtures increased from 3 to 5, the performance remained stable. After 5 mixtures, the performance deteriorated.

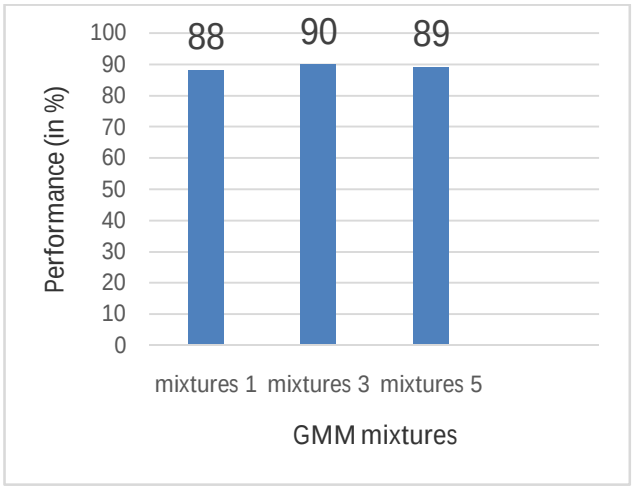

Figure 5: Performance of GMM for Different Mixtures.

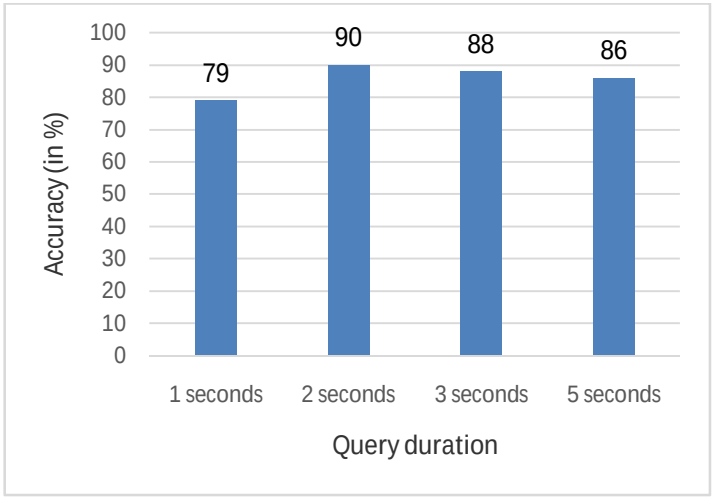

Figure 6: Performance of Music Retrieval for Different Durations of Query Clips.

For studying the performance, several experiments have been performed on music indexing for different durations of query music clips at 1, 2, 3 and 4 seconds respectively. Accuracy of retrieval is shown in Figure 6.

Table 1 shows performance of indexing and retrieval of query music clips using various feature sets.

Table 1: Performance of Music Indexing system using sonogram features.

Features	Accuracy of retrieval (in %)
Sonogram	90

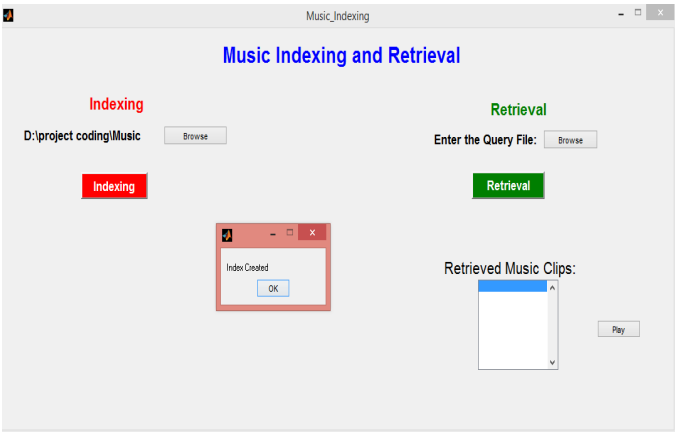

Figure 7: Snapshot for music audio indexing system

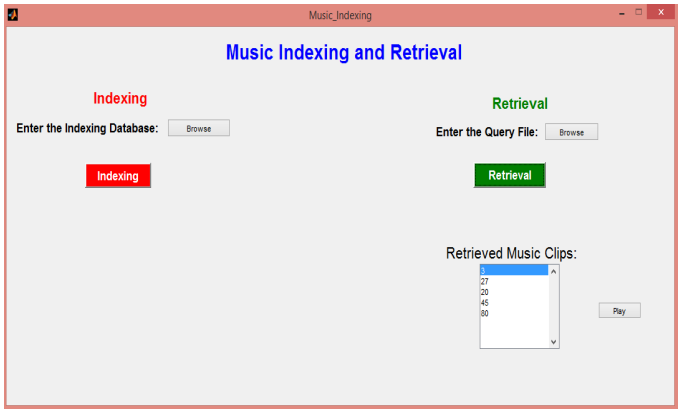

Figure 8: Snapshot for Music audio retrieval system

VII. CONCLUSION

In this work, methods are proposed for indexing and retrieval of music. In the music indexing for music audio clips the index is created for each song in the database. Sonogram features are extracted from each of the individual song. Retrieval is done for all the music query audio clip using Gaussian mixture model (GMM) models, based on the features extracted. The probability that the indexing feature vector belongs to the Gaussian is computed. The average Probability density function is computed for each of the feature vectors in the database and the retrieval is based on the highest probability. The query feature vectors were tested and the retrieval performance was studied. Performance of music audio indexing system was evaluated for 100 of songs, and the method achieves about overall 90.0% accuracy rate.

REFERENCES

[1] M. Adam Tindal, K. Ajaykapur and A. George Tzanetakis., "Training Surrogate in Musical Gesture Acquisition Systems", IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 13, NO. 1, FEBRUARY 2011.

[2] D. Alfonso Perez-Carrillo and M. Marcelo Wanderley., "Indirect Acquisition of Violin Instrumental Controls from Audio Signal with Hidden Markov Models", 932 IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 23, NO. 5, MAY 2015.

[3] A. Chung-Che Wang, J. Jyh-Shing Roger Jang., "Improving Query-by-Singing/Humming by Combining Melody and Lyric Information", IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 23, NO. 4, APRIL 2015.

[4] Dmitry Bogdanov, Martín Haro, Ferdin and Fuhrman, Anna Xambo, Emilia Gomez, A. Perfecto Herrera., "Semantic audio content-based music recommendation and visualization based on user preference examples", Information Processing and Management49 (2013) 1333.

[5] Francisco Raposo, Ricardo Ribeiro, and David Martins de Matos., "Using Generic Summarization to Improve Music Information Retrieval Tasks", IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 24, NO. 6, JUNE 2016.

[6] Mark Levy, Student Member, IEEE, and Mark Sandler., "Music Information Retrieval Using Social Tags and Audio", IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 11, NO. 3, APRIL 2009.

[7] Sebastian Ewert, MeinardMüller, VerenaKonz, Daniel Mullensiefen, and Geraint A. Wiggins., "Towards Cross-Version Harmonic Analysis of Music", IEEE TRANSACTIONS ON MULTIMEDIA, VO L.14, NO.3, JUNE 2012.

[8] Sunhyung Lee, DongsukYook, and Sukmoon Chang., "An Efficient Audio Fingerprint Search Algorithm for Music Retrieval", IEEE Transactions on Consumer Electronics, Vol. 59, No. 3, August 2013.

[9] Thomas Lidy, Carlos N.Silla Jr. OlmoCornelis, FabienGouyon, Andreas Rauber,CelsoA. A.Kaestner and Alessandro L.Koerich., "On the suitability of state-of-the-art music information retrieval methods for analysing, categorizing and accessing non-Western and ethnic music collections", Signal Processing 90 (2010) 1032–1048.

[10] R. Thiruvengatanadhan and P. Dhanalakshmi., "A Novel Method for Indexing and Retrieval of Speech using

Gaussian Mixture Model Techniques", International Journal of Computer Applications (0975 – 8887) Volume 148 – No.3, August 2016.

[11] R.Thiruvengatanadhan and P. Dhanalakshmi., "Indexing and Retrieval of Music using Gaussian Mixture Model Techniques", International Journal of Computer Applications (0975 – 8887) Volume 148 – No.3, August 2016.

[12] R.Thiruvengatanadhan and P. Dhanalakshmi., "A Fuzzy C-Means based GMM for Classifying Speech and Music Signals", International Journal of Computer Applications (0975 – 8887) Volume 102– No.5, September 2014.

[13] R. Thiruvengatanadhan and P. Dhanalakshmi., "Indexing and Retrieval of Speech using Perceptual Linear Prediction and Sonogram", Advances in Natural and Applied Sciences, 9(6) Special 2015, Pages: 117-122.

[14] R.Thiruvengatanadhan and P.Dhanalakshmi., "Speech/Music Classification using SVM" International Journal of Computer Applications (0975 – 8887) Volume 65– No.6, March 2013.

[15] Yi Yu, Roger Zimmermann, YeWang, and Vincent Oria., "Scalable Content-Based Music Retrieval Using Chord Progression Histogram and Tree-Structure LSH", IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 15, NO. 8, DECEMBER 2013.

[16] YonatanVaizman, BrianMcFee, and GertLanckriet., "Codebook-Based Audio Feature Representation for Music Information Retrieval", IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 22, NO. 10, OCTOBER 2014.

[17] Zhouyu Fu, Guojun Lu, Kai Ming Ting, and Dengsheng Zhang., "A Survey of Audio-Based Music Classification and Annotation", IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 13, NO. 2, APRIL 2011.