IOSR IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 2, Ver. II (Mar-Apr. 2014), PP 107-112
www.iosrjournals.org

Audio Denoising, Recognition and Retrieval by Using Feature Vectors

Shruti Vaidya[1], Dr. Kamal Shah[2]
[1]*MEIT-Student, TCET, Mumbai University, India*
[2]*MEIT- Professor, TCET, Mumbai University, India*

Abstract : *In this paper, we present the study of audio signal denoising, recognizing its audio type and retrieving similar signals from the database. Primarily the signal is denoised to remove any noise present in the unknown signal. Time Frequency Block Thresholding is used to achieve this. Removing the noise from audio signals requires processing of time-frequency coefficients in order to keep away from producing the musical noise. A block thresholding estimation procedure is introduced, which relates the parameters adaptively to signal property. Content Based Audio Retrieval system is very useful to identify the unknown audio signals. Audio signals are classified into music, speech and background sounds. This is done by the use of various feature vectors such as zero crossing rate, spectral centroid, spectral flux, spectral roll off. Use of multiple feature vectors improves the accuracy of the results being retrieved from the audio database. Various experiments demonstrate the performance and accurateness, providing good results for non standard signals.*

Keywords: *Audio Classification, Audio Retrieval, Distance Measure, Feature Vectors, Precision Recall, Vector Quantization*

I. INTRODUCTION

World Wide Web Applications have extensively grown since last few decades and it has become requisite tool for education, communication, industry, amusement etc. All these applications are multimedia-based applications consisting of images, audio and videos. Images/audio/videos require enormous volume of data items, creating a serious problem as they need higher channel bandwidth for efficient transmission. Further degrees of redundancies are observed. Thus the need for compression arises for resourceful storage and transmission. Compression is classified into two categories, lossless image compression and lossy image compression technique [1]. Human–computer interactions often involve multiple tasks such as routing, searching, discussion and data retrieving. In multi-tasking situations it is difficult to find the user. Vector quantization (VQ) is one of the lossy data compression techniques and has been used in number of applications, like pattern recognition, speech recognition and face detection, image segmentation, speech data compression , Content Based Image Retrieval (CBIR), Face recognition, iris recognition etc.

Audio Data is an integral part of many modern computer and multimedia applications. Numerous audio recordings are dealt with in audio and multimedia applications. The effectiveness of their deployment is greatly dependent on the ability to classify and retrieve the audio files in terms of their sound properties or content [2].

Audio, which includes voice, music, and different kinds of background sounds, is an important type of media, and also a significant part of audio visual data. Currently there are more number of digital audio databases in place and hence, people have realized the importance of audio database management relying on audio content analysis. There are also distributed audio libraries in the World Wide Web, and content-based audio retrieval could be an ideal approach for sound indexing and search. Existing research on content-based audio data management is very limited. There are in general two directions. One direction is audio segmentation and classification, where audio is classified into music, speech and background. The second direction is audio retrieval, where audio is retrieved from the database using different audio features. With the development of multimedia technologies, large amounts of multimedia information is transmitted and stored every day [3].

II. OVERVIEW OF AUDIO SYSTEM

The basic operation of the system is as follows. First, the feature vectors are estimated for each signal from the database. Then, for each input signal from the user, its feature vectors are estimated. The input signal from the user is compared to each signal form the database. Wherever the resemblance is achieved, the input signal form the user is classified into that particular category. Signals related to that category can be retrieved and heard.

As shown in the Fig.1, next step is comparing two samples. The signal is divided into frames and feature vector is calculated for each frame. The features describing the frequency as well as temporal characteristics are used. After the calculations, the values of the feature vectors of the input signal are compared to that of the database values. When resemblance is found between the input and database signals then the input signals can be classified appropriately into their domains [3].

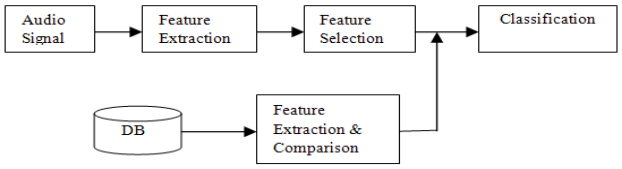

Figure 1: Overview of Audio System

III. FEATURE EXTRACTION

In order to improve the correctness of segmentation and classification of an audio signal, it is important to select good features that can capture the characteristics of audio signal. Numerous feature vectors are available which provide help for discrimination. These features will be described in detail in this section [3, 4].

- **Zero Crossing Rate**

The zero crossing rate counts the number of times that the signal amplitude changes signs in the time domain during one frame of length N,

$$ZCR = \frac{1}{2}\sum_{n-1}^{N} |sgn(X[n]) - sgn(X[n-1])| \quad (1)$$

Where the sign function is defined by
Sgn(x) = 1, x >=0
= -1, x <0

- **Spectral Centroid**

Centroid is the gravity of the spectrum, where the sign function is defined by

$$Cr = \frac{\sum_{K=1}^{N/2} f[k]|X_r[k]|}{\sum_{K=1}^{N/2} |X_r[k]|} \quad (2)$$

Where *N* is a number of FFT points, *Xr* [*k*] is the STFT of frame *xr*, and *f* [*k*] is a frequency at bin *k*. Centroid models the sound sharpness. Sharpness is related to the high frequency content of the spectrum.

- **Spectral Roll-Off**

The roll-off is a measure of spectral shape useful for distinguishing voiced from unvoiced speech. The frequency below which 85% of the magnitude distribution of the spectrum is concentrated is known as Roll-Off. That is, if K is the largest bin that fulfils,

$$\sum_{K=1}^{N/2} |X_r[k]| \leq 0.85 \sum_{K=1}^{N/2} |X_r[k]| \quad (3)$$

- **Spectral Flux**

The spectral flux is defined as the squared difference between the normalized magnitudes of successive spectral distributions that correspond to successive signal frames. That means, a measure of the spectral rate of change, which is given by the sum across one analysis window of the squared difference between the magnitude spectra corresponding to successive signal Frames,

$$F_r = \sum_{K=1}^{N/2} (|X_r[k]| - |X_{r-1}[k]|)^2 \quad (4)$$

Flux has been found to be a suitable feature for the separation of music from speech, yielding higher values for music samples.

- **Beat Strength**

Statistical measures of the histogram namely mean, standard deviation, mean of the derivative, standard deviation of the derivative, entropy and so on are evaluated to obtain an overall measure of beat strength. All the events are processed in the beat domain.

- **Rhythmic Regularity**

A beat histogram in which there is periodic spacing, the peaks denote high rhythmic regularity. This can be measured by the normalized autocorrelation function of the beat histogram. It contains clear peaks for rhythmically regular music examples and it will be more linear if the regularity is weaker. In order to ease this to a scalar measure of rhythm regularity, the mean across the lags of the difference between the autocorrelations and the linear function is computed. Since, the computation is performed on a frame-by-frame basis; histograms are obtained in long-term intervals given by the texture windows. Therefore all the features related to the beat histogram are single-valued features wherein the time domain mean and standard deviation sub-features will not be applicable.

- **Audio Spectrum Centroid**

A perceptually adapted definition of the centroid, wherein a logarithmic frequency scaling centered at 1 kHz is introduced,

$$ASC_r = \frac{\sum_{k=1}^{N/2}\left(\frac{f[k]}{1000}\right)P_r[k]}{\sum_{k=1}^{N/2}P_r[k]} \quad (5)$$

Where P_r represents the power spectrum of the frame r.

- **Audio Spectrum Spread**

Audio Spectrum Spread tells about the concentration of the spectrum around the centroid and is defined as,

$$ASS_r = \sqrt{\frac{\sum_{k=1}^{N/2}[log_2(\frac{f[k]}{1000})-ASC_r]^2 P_r[k]}{\sum_{k=1}^{N/2}P_r[k]}} \quad (6)$$

Lower spread values mean that the spectrum is highly concentrated near the centroid and higher values mean that it is distributed across a wider range at both sides of the centroid.

- **Audio Spectral Flatness**

It can be defined as the deviation of the spectral form from that of a flat spectrum. Flat spectra correspond to noise or impulse-like signals hence high flatness values indicate noisiness. Low flatness values generally indicate the presence of harmonic components. Instead of calculating one flatness value for the whole spectrum, a separation in frequency bands is performed, ensuring in a vector of flatness values per time frame. The flatness of a band is defined as the ratio of the geometric and the arithmetic means of the power spectrum coefficients within that band. Each vector is eased to a scalar by calculating the mean value across the bands for each given frame, hence obtaining a scalar feature that describes the overall flatness.

- **Low Energy Rate**

It can be expressed as the percentage of frames within a file that have root mean square (rms) energy lower than the mean rms energy in that file. Apart from the beat-histogram based features, this is the feature which is computed on a texture window basis rather than on a frame-by-frame basis.

- **Loudness**

The previously discussed features were dynamic-related and are based on physical measures such as amplitude or energy. The loudness measurement provides for a better understanding that is to the human ear perception of the various sound dynamics.

IV. CLASSIFICATION

After the feature extraction and selection process is accomplished the next important thing is to classify the signal. Classification is the process by which a particular label is assigned to a particular input audio signal. This label defines the signal. A classifier defines boundaries for assessment in the feature space, which separates different classes of signals from each other. Classifiers are categorized by their real time capabilities, on the basis of the approach and their character.

4.1 k-Nearest Neighbor

The K nearest neighbors classifier (KNN) is a non-parametric classifier. Let us consider Dn=x1…. xn as a set of n labeled prototypes then the nearest neighbor rule for classifying an unknown vector x is to assign it, the label of its nearest point in the set of labeled prototypes Dn. The KNN rule classifies x by assigning the label most frequently represented among the k nearest samples. Generally k is considered odd to avoid ties in selection. This algorithm requires heavy time and space other methods can be used to make faster computations and reduce storage requirements [5].

4.2 Fast Fourier Transform (FFT)

A Fast Fourier Transform (FFT) is an algorithm to compute the discrete Fourier transform (DFT) and it's inverse. A Fourier transform converts time (or space) to frequency and vice versa; an FFT rapidly computes such transformations. The FFT is obtained by decomposing a sequence of values into components of different frequencies more quickly. This operation is useful in many fields [6]. After the conversion into the frequency domain feature vectors can be applied for further processing.

4.3 Distance Metrics

Once the appropriate domain is obtained, the different signals belonging to the domain can be retrieved by the use of any of the following distance calculation methods [7]:-

- **Euclidean Distance**

Deriving the distance between two data points involves computing the square root of the sum of the squares of the differences between corresponding values.

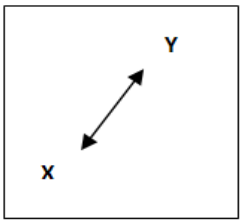

Figure 2: Euclidean Distance

- **Manhattan Distance**

The Manhattan distance function computes the distance that would be traveled to get from one data point to the other if a grid-like path is lead. The Manhattan distance between two entities is the sum of the differences of their corresponding components.

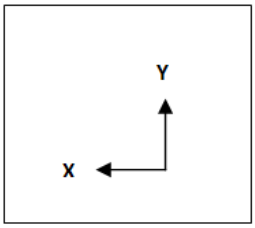

Figure 3: Manhattan Distance

V. DENOISING

Audio signals are often contaminated by environmental noise and buzzing or humming noise from audio equipments. Aim of Audio Denoising is to gradually reduce the noise while retaining the underlying signals. There are numerous applications namely, music and speech restoration [8].

Time Frequency Audio Denoising procedures compute a short time Fourier transform or a wavelet transform or a wavelet packet transform of the noisy signal, and process the resulting coefficients to reduce the intensity of noise. These representations disclose the time frequency signal structures that can be distinguished from the noise.

A time-frequency block thresholding estimator regularizes power subtraction estimation by computing a single attenuation factor over time-frequency blocks. The time-frequency plane {l, k} is segmented in I blocks B_i wherein the shapes are selected randomly. The signal estimator f is calculated from the noisy data y with a constant attenuation factor a_i over each block B_i

$$f[n] = \sum_{i=1}^{I} \sum_{(L,k)\in B_i} a_i Y[l,k] g_{l,k}[n] \quad (7)$$

A block thresholding estimator can thus be interpreted as an estimator derived from averaged signal to noise ratio estimations over blocks. Every attenuation factor is processed from all coefficients in each block, which regularizes the time-frequency coefficient estimation.

VI. PRECISION AND RECALL

Precision and recall are greatly used parameters in evaluating the correctness of a pattern recognition algorithm [9].Precision is a measure of fidelity whereas recall is a measure of completeness. Precision basically is a measure of the number of retrieved documents that are relevant to the search. Precision can also be evaluated at a given cut off n.

Recall as mentioned earlier is a measure of completeness i.e. it is basically the probability of a relevant document being returned by the query. In binary classification, recall is also referred as sensitivity. The mathematical computation of precision and recall is as follows [10]:

Precision = |Relevant Documents ∩ Retrieved Documents| / | Retrieved Documents | (8)

Recall = | Relevant Documents ∩ Retrieved Documents | / | Relevant Documents |

Crossover point in precision recall is the point on the graph where both the precision and recall curves meet. Crossover point in itself can be used a way to measure the correctness of an algorithm. A higher crossover point indicates a better performance for a particular method.

VII. RESULT

The study is presented on the basis of different audio input signals taken in wave format [10, 11]. The received incoming regular (non standard) signals are analyzed by using the various feature vectors as studied above, to determine the class it belongs to. The graphical representation supports the result

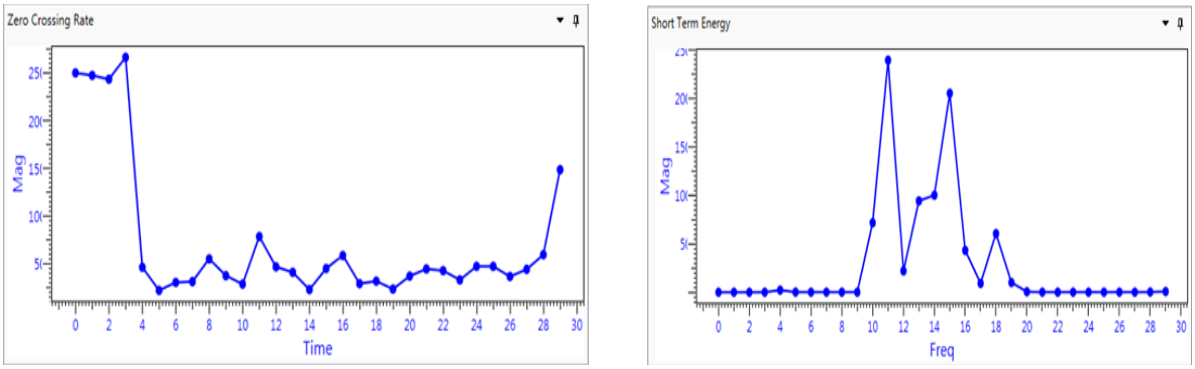

Figure 4: Analyzing the input signal

Fig. 4 shows the signal analysis. When an unknown signal is analyzed, it gives the desired outcome, to which class it belongs. Graphical representation for zero crossing rate and short time energy feature vectors is shown.

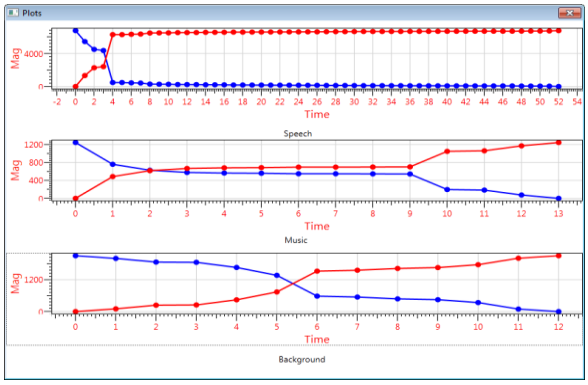

Figure 5: Precision Recall Graph for Non-standard

VIII. CONCLUSION

In this paper a search method is used which will first denoise the signal after which the signal is recognized. Once the signal is recognized similar signals can be retrieved. To achieve this multiple feature vectors are used to obtain more accurate results. An audio block-thresholding algorithm is used for denoising purpose, which adapts all parameters to the time-frequency regularity of the audio signal. Number of experiments have revealed satisfactory results for non standard signals, through objective and subjective evaluations.

REFERENCES

[1] Dr. H. B. Kekre, Tanuja K. Sarode, "New Clustering Algorithm for Vector Quantization using Rotation of Error Vector ", *(IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, No. 3, 2010*

[2] Guodong Guo and Stan Z. Li, "Content-Based Audio Classification and Retrieval by Support Vector Machines", *IEEE Transactions On Neural Networks, Vol. 14, No. 1, January 2003*

[3] Vaishali Nandedkar, "Audio Retrieval Using Multiple Feature Vectors", *International Journal of Electrical and Electronics Engineering (IJEEE) ,Volume-1 ,Issue-1 ,2011*

[4] Hariharan Subramanian, Prof. Preeti Rao and Dr. Sumantra. D. Roy, "Audio Signal Classification", *M.Tech. Credit Seminar Report, Electronic Systems Group, EE. Dept, IIT Bombay, Submitted November2004*

[5] R. Duda, P. Hart, and D. Stork., "Pattern classification" , John Wiley & Sons, New York, 2000.

[6] Gokul P, Karthikeyan T , KrishnaKumar R. Malini S., Final Year ECE students, Department of Electronics and Communication Engineering, Assistant Professor, Department of Electronics and Communication Engineering, " Computational Time Analysis of Signal Processing Algorithm-An Analysis" *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)* e-ISSN: Gokul P, Karthikeyan T , KrishnaKumar R. Malini S., Final Year ECE students, Department of Electronics and Communication Engineering, Assistant Professor, Department of Electronics and Communication Engineering, " Computational Time Analysis of Signal Processing Algorithm-An Analysis" *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)*

[7] T. Soni Madhulatha, Associate Professor, Alluri Institute of Management Sciences, Warangal, "An Overview On Clustering Methods", *IOSR Journal of Engineering Apr. 2012, Vol. 2(4) pp: 719-72*

[8] Guoshen Yu, Stéphane Mallat, Fellow, IEEE, and Emmanuel Bacry, "Audio Denoising by Time-Frequency Block Thresholding", *IEEE Transactions On Signal Processing, Vol. 56, No. 5, May 2008*

[9] Dr. H. B. Kekre Sr. Prof, Department of Computer Science, NMIMS University, Mr. Dhirendra Mishra Associate Prof, Department of Computer Science, NMIMS University, Mr. Anirudh Kariwala Student, Department of Computer Science, NMIMS University," A Survey Of CBIR Techniques And Semantics" , Dr. H.B. Kekre et al. / *International Journal of Engineering Science and Technology (IJEST)*

[10] Shruti Vaidya, Dr. Kamal Shah, " Application of Vector Quantization for Audio Retrieval", *International Journal of Computer Applications (0975 – 8887) Volume 88 – No.17, February 2014*

[11] Shruti Vaidya, Dr. Kamal Shah, "Audio Classification and Retrieval by Using Vector Quantization" , *International Journal Of Scientific & Engineering Research, Volume 5, Issue 2, February-2014, ISSN 2229-5518*