

Feature Selection and Extraction of Audio Signal

Jasleen ¹, Dawood Dilber ²P.G. Student, Department of Electronics and Communication Engineering, Amity University, Noida, U.P, India¹P.G. Student, Department of Electronics and Communication Engineering, Amity University, Noida, U.P, India²

ABSTRACT: Classification systems of the audio signals are used for analysis of the input signal and then were used to extract the different characteristics or features of the audio signal. Classification of audio signal is used to draw some sensory and physical characteristics like voice and is used to determine their characteristics. Extraction algorithms can be used vastly, depending on the field of classification Application. In this paper, features of audio signals and there extraction are discussed and how to select the optimal features from the selected features. A number of features such as MFCC, Pitch, fundamental frequency characteristics are discussed. The extracted features can be choosed using various algorithms such as genetic algorithm; greedy algorithms are explained which are used for getting the optimized output. The greedy algorithm is applicable only in some situations but not always but to get the optimized values Genetic algorithm always give the best results.

KEYWORDS: Audio Signal, Feature Selection, Feature Extraction, Pitch, MFCC, ZCR, Greedy algorithm, Genetic Algorithm.

I. INTRODUCTION

A signal is the physical representation of a positive knowledge. This knowledge can be in the form of data voice, picture and many more. The audio is any waveform whose frequencies range is in the human audible range.

Group of the audio signals are used to define different formats of input audio signals. These grouping have many benefits in the area of research like broadcasting, survey & retrieval of knowledge etc. To create a tag which outlines the output signal, also analyzes the input signal & audio signal, classification systems are used which are helpful for detection of whether the signal is music or any kind of a speech [10]. These signals can be presented based on melody, content, pitch & pace etc.

More publicly these absorb applications of Audio Signal Classification includes television and radio advertisement identification, for muting or VCR pausing, for receiving command from users. In this paper, we studied the various features of audio signals. The main aim of the feature extraction step was to encapsulate the most relevant and discriminate attributes of the signal to acknowledge these features. So, feature extraction was needed and the features selected were put into classifier. The features which were extracted from the input audio signals were greatly independent from each other. Feature ZCR was the easiest one and it counts wherever there was a change in waveform, when it cuts the zero axes.

II. RELATED WORK

For audio signal classification, we first extracted the features from the input signal. From these extracted features optimized features were then selected using various algorithms like genetic algorithm, greedy algorithm, the Sequential Forward Search, the Sequential Backward Search, mutative Algorithm etc. The audio signal was given as input to the feature extraction block in which various features like MFCC, Pitch, ZCR etc were extracted and then the output of this block was given to the feature Selection block from which optimized features were selected by using various

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

algorithms as shown in Fig.1. Finally, the output of this block was given to the classifier which applies some rules for the class.

Fig 1. Feature analyzing of audio signal in which audio input file carrying many features are needed to be extracted from the audio file. In the figure the audio file features are extracted like MFCC, ZCR etc and by selecting the best features we classify them by classifying model.

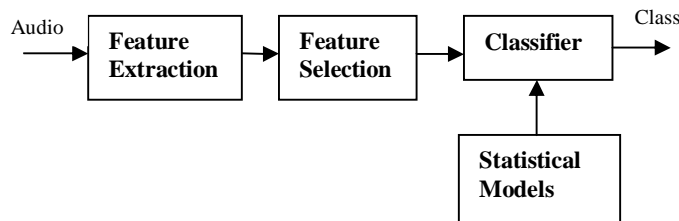


Fig 1

A. Extracting the Features

Before the grouping of audio signal, the features in that audio signal were first extracted and later on selected. Extraction of feature was done to reduce or to minimize the amount of data and to choose the various features from the mentioned features [4]. Feature extraction is the measure of competing a compact numerical representation that can be used to characterize a segment of audio.

The Valuable features can analyze the design of the classifier whereas lousy features can hardly be compensated with any classifier. The audio signal which is input signal was analyzed by feature extraction method in which various features are extracted like MFCC, Pitch, sampling frequency, loudness, volume etc.

1. Zero Crossing Rate (ZCR)

ZCR is a measure of the number of the time the signal value crosses the zero axes. It is easy to measure the ZCR signals as they are easy to measure and are very famous. In this the cyclic sound has low value than the noisy sound which is having more values [1]. To roughly estimate the fundamental frequency, ZCR are used for the marked signals. While case is different for complex signals. ZCR can also be defined as a measure of how often the sound signal crosses from positive to negative or vice-versa. This feature is also used to separate the other features like the noise. It works for vector and matrix and the function is vectorized very fast.

2. Mel Frequency Cepstral Coefficient

MFCCs are a dense which presents the signals which are audio in nature are measured in units known as Mel scale [1]. These features are used for analysing speech signals and now recently are represented as melody signal. MFCCs calculated by defining the STFT crescents of individual frame into sets of 40 consents that use a set of the 40 weighting contours simulating the frequency sensing capability of humans. After this logarithm coefficients are taken into account and also a DCT is used so that it does not relate them. In normal case, the five first coefficients are taken as features.

The Mel scale relates the frequency which is pre received of a pure tone to its actual measured frequency.

Below the formula is used to convert frequency to Mel:

$$M(f) = 1125 \ln(1 + f/700) \quad (1)$$

and Mel to frequency:

$$M^{-1}(m) = 700(\exp(m/1125) - 1) \quad (2)$$

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

3. Pitch

The pitch determination is very important for many speech transforming algorithms. Pitch is the quality of a sound commanded by the rate of vibration generating it, the amount of highness or lowness of tone. The sound which is coming from the vocal cords begins at larynx & stops at mouth. Using brain's nerves, the shape of vocal tract and shaking of the vocal cord can be managed [6]. The produced sounds are categorized either as unvoiced sounds or voiced one. When unvoiced sounds are produced vocal cords don't shake and are open while when voiced sounds are producing, vocal cords vibrate and produce pulses which are also known as glottal pulses.

Pitch can be detected using:

1. Cepstral Method
2. Auto-correlation Method
3. Harmonic Product Spectrum (HPS)
4. Linear predictive Coding (LPC)

4. Fundamental Frequency

The fundamental frequency is the lowest most frequency in which the signal repeats. We can extract the signal if only the signal is periodic in nature [7]. By this, we can use periodic detector i.e., the extracted signal known as periodic signal.

This frequency can be changed from 40 Hz which is low pitched of the voices of the male to 600 Hz which is high pitched voice of female or children. To detect pitch, auto-correlation method technique can use to pitch periods for the detection purpose meaning, in order to detect 40Hz frequency, 50 ms of speech signal is analyzed for this.

B. Feature Selection

Feature selection is used to select various features from the extracted features so as to get the optimized values from that set of features. This is used to select features from the large set of available features which is extracted using feature extraction method and these features were used to determine the nature of the audio signal [1]. It is used to select the optimum values or features keeping accuracy and performance level by minimizing computational cost. So, it has drastic effect on the accuracy and has more computational cost if no features were developed.

Goals of Feature Selection Method:

- To maximize the performance of learning algorithms we have to choose feature subset.
- To decrease the need for computer storage and processing time needed to classify the data by not reducing the performance of algorithm.
- To detect a subset of features that is related to the natural problem being studied.
- Reduction of features can improve the quality of prediction and even can be a necessary, embedded, step of the prediction algorithm.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

1) Greedy Algorithm

A greedy algorithm all the time made some selection which looks perfect at that state. Greedy are those algorithms that are used for boosting the problems that occur [2]. If any difficulty happens, then it can be breaked using greedy method perhaps if it had following characteristics: *At individual step, "we can set the choice that looks perfect at the all the moment and we can get the optimal solution of the complete problem"*. By using this technique it becomes the best method to solve such situations such that the greedy algorithm is more efficient or reliable than the other methods.

Example:

Greedy algorithm can be explained by giving an example of a shopkeeper who wants to return minimum no of notes and coins to the customer.

Suppose, the bill of customer is Rs 753 and the customer is giving a note of 1000rs to the shopkeeper and the shopkeeper has to return minimum no of notes to customer. We can solve the above problem i.e., making coin change in MATLAB by creating graphical user interface (GUI) application.

Fig 2. Greedy Algorithm is explained by creating GUI application. In the application shown below the greedy algorithm can be explained by entering the amount of 247 in the text box opposite to MONEY label and this will generate how many minimum notes needed to be returned to the customer. This is how the greedy algorithm works.

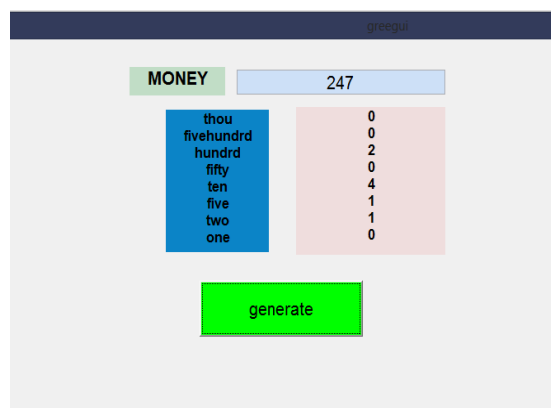


Fig 2

So from this we analysed that shopkeeper returned two notes of five hundred rupee, four notes of ten rupee and two coins of five and two rupees and this is how the greedy algorithm works.

2) Genetic Algorithm

Genetic algorithm are based on Darwin's theory of evolution. Genetic Algorithms are used to develop optimal solution by method of evolution-inspired search and optimization. From the population to make next generation, Genetic algorithm uses following rules which are:

- Selection rules used for selecting the individuals, known as parents, contributing at next generation.
- Crossover rules used to make children from the 2 parents for next generation.
- Mutation rules used to form children by applying irregular changes to each parents shown in Fig.3.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

Fig 3. Flow diagram of Genetic Algorithm: In this the initial population is either taken or has been generated which is then evaluated on the basis of optimality. To create new population genetic algorithm goes under following three stages which are Selection, Crossover and Mutation and from them the optimality for genetic algorithm will be computed.

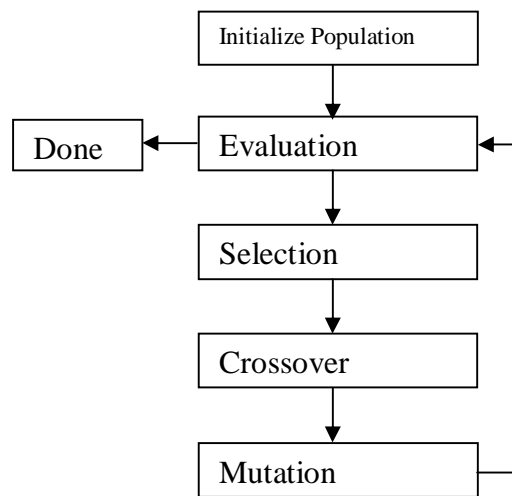


Fig 3

We can use Genetic Algorithm by calling GA function in command window or by using GA Toolbox.

III. EXPERIMENTAL RESULTS

The audio signal containing features were processed so as to extract those features. After successful extraction of these features, the optimal features were selected from the extracted features. In this, we have an audio file e.g. 'flute.wav' and we extracted features from this file.

The features extracted were:

- a. MFCC: The Mel scale relates the frequency which is pre-received of a pure tone to its actual measured frequency. The sampling frequency is being calculated from this which is 22050. Also frame size is calculated which is [1323 79] where, 79 is number of frames and 1323 is sampling frequency of each frame. The results were shown in Fig 4.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

Fig 4. MFCC analysis of sampling frequency and number of frames. The number of times in which we sampled the audio signal per unit time is sampling frequency and this sampled audio is stored as a number. The sampled frequency of the audio signal computed is 22050 and 79 are the total number of frames being calculated.

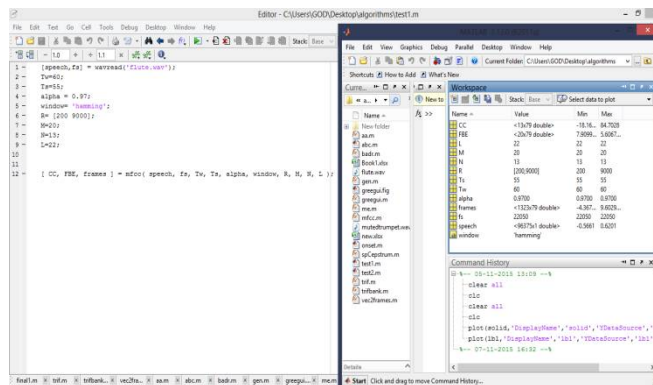


Fig 4

b. PITCH: Pitch is important for many speech transforming algorithms. To obtain cepstrum coefficients of a signal, below function is called:

$$\text{function [c, y] = sp Cepstrum (fs, window, show)}$$

Where, c (size Nx1) contains cepstrum coefficients and y (size Nx1) contains Fourier response. The waveform of Amplitude vs. Time (s) and frequency (Hz) is shown in Fig.4.

Fig 5. Amplitude vs. Time and Frequency: The amplitude which the peak frequency of the audio signals is computed with respect to Time in seconds and Frequency in Hertz which will the waveform and cepstrum of the audio signal.

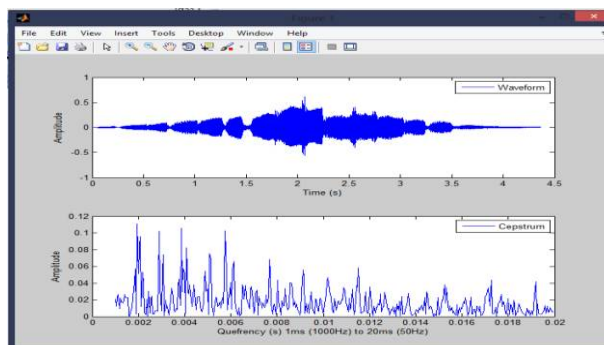


Fig 5

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

c. Genetic Algorithm:

To optimize an equation (e.g. $y=2*x^2-\cos(x) +8*x^3$) we use genetic algorithm (GA). For this we have to call a fitness function and then evaluate the best point.

Taking example, we have made fitness function "gen(x)" of equation. The best point and value of fitness function is shown in Fig.5.

Fig 6. Working of Genetic algorithm in which we compute the values of x and fval which in returns give the best optimal value of the audio file carrying features. The GA compute the best point from the final population and also it calculates the value of function at the best point.

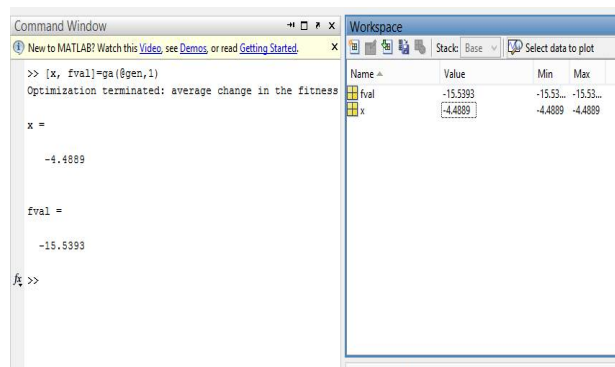


Fig 6

Where, x is the best point in the final population computed by GA and fval is the value of function (@gen) evaluated at point x.

Fig 7. "GAPLOTBESTF" is the first plot used which plots mean and the best score of population at each and every stage or generation. "GAPLOTSTOPPING" is the second plot describing why the optimization is stopped and giving the total percentage of that criterion giving the best fitness value and the mean fitness value.

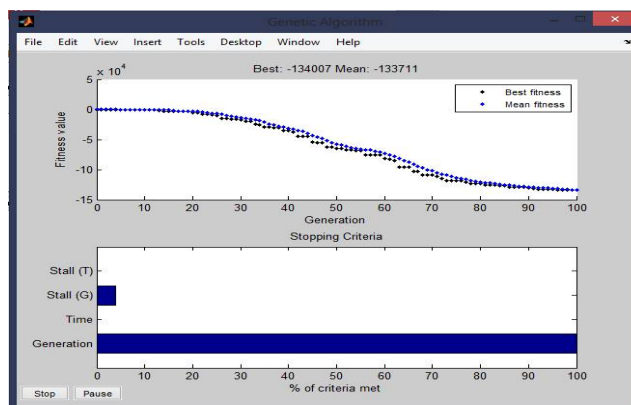


Fig 7

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

IV. CONCLUSION

Various features of audio signals and the methods to extract them from the audio signal were studied. Features like MFCC, ZCR, pitch, sampling frequency etc were used and various methods to select the best optimal feature from the extracted features were investigated. The genetic algorithm proved better than the greedy algorithm because it gave more optimised results compared to greedy algorithm. Genetic algorithm gave more robust solution as compared to greedy algorithm. The features giving good description of the signal were given the priority from the different features. MFCCs were used as they define the information related to pitch & rhythmic contents which helps in grading or classifying that gives better classification results than the other features.

Genetic algorithm explores in a highly and efficient way, so the space of all the possible subsets to obtain the set of features that maximises the predictive accuracy of learned rules. The reason for the termination of optimization can also be checked and visualisation of the result can be obtained but the greedy algorithm can only give the best result in certain cases. So the selected features by using the above algorithms reduce the complexity of the system and thus reduce the cost. Hence, the feature selection was run to achieve acceptable high recognition rate and also for the reduction of the running time of a given system.

In order to classify the incoming audio signal, either speech or musical by their nature, this would be achieved by determining the properties of the input signal. Most effective algorithms could have their performance level to be raised. As the FFT that is fast Fourier transform is relied on the periodic or cyclic function we can analyse some undesired effects occurring. The data being analysed of any "windows" is not actually periodic as they are having different size of the data of window. Importantly, there are more other "effects" which can be implemented like compression methods, noise reduction or removal methods and more.

REFERENCES

- [1] G. Tzanetakis and P. Cook, "Musical Genre Classification of Audio Signals", IEEE Trans. Speech and AudioProcess, vol. 10, pp. 293-302, 2002 July.
- [2] Yong, M., Falzon, B.G. and Iannucci, L. (2008). "On the application of genetic algorithms for optimising composites against impact loading", Elsevier, International Journal of Impact Engineering, vol. 7, 2010 May.
- [3] Hafner, C. and Frohlich, J., "Generalized Genetic Programming for Solving Engineering Problems", Proc. PIERS Symposium, (Boston): 672.
- [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, November 2004.
- [5] M. K. Lee*, W. Leung**, T. L. Pun and H. L. Cheung, "EDGE DETECTION BY GENETIC ALGORITHM", 0-7803-6297-7/100/\$10.00 © 2000 IEEE.
- [6] J. Foote, "A Similarity Measure for Automatic Audio Classification", 1997 Spring Symp. on Intelligent Integration and Use of Text, Image, Video, and Proc. AAAI Audio Corpora, Stanford, CA, 1997
- [7] J. J. Burred and A. Lerch, "Hierarchical Automatic Audio Signal Classification", J. Audio Eng. Soc, Vol. 52, pp. 724-739, July/August 2004.
- [8] G. Tzanetakis and P. Cook, "Human Perception And Computer Extraction Of Musical Beat Strength", Proc. of the 5th Int. Conference on Digital Audio Effects (DAFx-02), Hanburg, Germany, 2002 September.
- [9] Cho H.J., Wang B.H., S., "Automatic rule generation for fuzzy controllers using genetic algorithms: a study on representation scheme and mutation rate", IEEE World Congress on Computational Intelligence Fuzzy Systems, 1998
- [10] Specht, D.F., "Probabilistic Neural Networks and the Polynomial Adaline as Complementary Techniques for Classification", IEEE Transactions on Neural Networks, vol. 1, 1990, pp. 111-121.