KAUNAS UNIVERSITY OF TECHNOLOGY

DEPARTMENT OF MULTIMEDIA ENGINEERING

VINODH RAVICHANDRAN SENTHAMARAIKANI

SANJANA SURESH KUMAR

RESEARCH ON ANIMAL SOUND RECOGNITION

MASTER THESIS

SUPERVISOR

Doc.Dr.Rytis Maskeliunas

KAUNAS, 2016

KAUNAS UNIVERSITY OF TECHNOLOGY

DEPARTMENT OF MULTIMEDIA ENGINEERING

RESEARCH ON ANIMAL SOUND RECOGNITION

MASTER THESIS

P000M106

SUPERVISOR

Doc.Dr.Rytis Maskeliunas

Signature:

Date:

REVIEWER

Signature:

Date:

KAUNAS, 2016

KAUNAS UNIVERSITY OF TECHNOLOGY

DEPARTMENT OF MULTIMEDIA ENGINEERING

FACULTY OF INFORMATICS

<u>VINODH RAVICHANDRAN SENTHAMARAIKANI</u>

Final Thesis P000M106

<u>RESEARCH ON ANIMAL SOUND RECOGNITION</u>

DECLARATION OF ACADEMIC INTEGRITY

/ /2016

Kaunas

I confirm that I have produced the final project of mine, VINODH RAVICHANDRAN SENTHAMARAIKANI, on the subject "RESEARCH ON ANIMAL SOUND RECOGNITION" completely; all the provided data and research results are correct and have been obtained honestly. None of the parts of this thesis have been plagiarized from any printed, Internet-based or otherwise recorded sources. All direct and indirect quotations from external resources are indicated in the list of references. No monetary funds (unless required by law) have been paid to anyone for any contribution to this thesis.

I fully and completely understand that any discovery of any manifestations/case/ facts of dishonesty inevitably results in me incurring a penalty under procedure effective at Kaunas University of Technology.

-- ---------------------------

(*Name and Surname filled in by hand*) (*Signature*)

KAUNAS UNIVERSITY OF TECHNOLOGY

DEPARTMENT OF MULTIMEDIA ENGINEERING

FACULTY OF INFORMATICS

SANJANA SURESH KUMAR

Final Thesis P000M106

RESEARCH ON ANIMAL SOUND RECOGNITION

DECLARATION OF ACADEMIC INTEGRITY

/ /2016

(Kaunas)

I confirm that I have produced the final project of mine, SANJANA SURESH KUMAR, on the subject "RESEARCH ON ANIMAL SOUND RECOGNITION" completely; all the provided data and research results are correct and have been obtained honestly. None of the parts of this thesis have been plagiarized from any printed, Internet-based or otherwise recorded sources. All direct and indirect quotations from external resources are indicated in the list of references. No monetary funds (unless required by law) have been paid to anyone for any contribution to this thesis.

I fully and completely understand that any discovery of any manifestations/case/ facts of dishonesty inevitably results in me incurring a penalty under procedure effective at Kaunas University of Technology.

-- ---------------------------

(*Name and Surname filled in by hand*) (*Signature*)

ACKNOWLEDGEMENT

We extend our gratitude to our University for having provided us with an opportunity all facilities to complete our Master Thesis successfully.

We express our sincere thanks to our supervisor Doc.Dr. Rytis Maskeliunas for giving us valuable advises and guidance during our thesis. We also thank our beloved professors for guiding us during the entire course of the thesis.

We are committed to place our heartfelt thanks to all our teaching and non-teaching staff members of Faculty of Informatics, for providing us with the valuable suggestions and the kind of cooperation for completion of the Master thesis.

We also thank our family members without whose support and encouragement, it would not have been possible to complete the Master’s Degree successfully.

VINODH RAVIHANDRAN SENTHAMARAIKANI and SANJANA SURESH KUMAR, RESEARCH ON ANIMAL SOUND RECOGNITION SYSTEM: Master thesis Final Project. Supervisor Doc.Dr. Rytis Maskeliunas. The Faculty of Informatics, Kaunas University of Technology.

Research area and field: Research in Multimedia and Technology

Kaunas.

SUMMARY

The animal sound recognition system has been developed using the algorithms ZCR, MFCC, DTW and LPC along with SVM as a classification method for recognizing a particular animal's emotion. Among these algorithms used in the system, DTW and LPC show a higher accuracy rate compared to that of ZCR and MFCC. A Graphical User Interface has been developed, which enables the use of the animal sound recognition system a more convenient and easy method to detect the state of the animal and act according to that situation of the animal.

Keywords: Animal Sound Recognition System; ZCR; MFCC; DTW; LPC; SVM

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ZCR	Zero Cross Rate is used for end point detection of input voice such that the silence Voice are removed
MFCC	Mel Frequency Cepstrum Coefficients used for the process of feature extraction
DTW	Dynamic Time Wrapping is to get the optimal path between input voice and Reference voice
LPC	Linear Predictive Coding predicts the amount of voice added to the original noise
SVM	Support Vector Machine is a classification method for automatically detecting the Emotion
HMM	Hidden Markov Model
NN	Neural Network
GUI	Graphical User Interface
ASR	Automatic Speech Recognition
DNN	Deep Neural Network

ABSTRACT

Voice recognition systems in the modern world have become important for speech recognition technology. In this work, the animal voice recognition system is being developed, based on animal voice pattern recognition algorithm. The developed animal voice recognition system follows the Zero-Cross-Rate (ZCR), Mel-Frequency Cepstrum Coefficients (MFCC) and Dynamic Time Warping (DTW), Linear Predictive Coding (LPC) algorithms in common to detect certain animal voice and recognize their emotion. ZCR algorithm is used for the detection of the input voice in order to remove the silent part of the input sound. MFCC algorithm is used for feature extraction to implement a more compact and less redundant system from the given sounds. LPC analyzes the speech signal by estimating the formants in the input data and Linear Predictive Coding (LPC) is a predefined function, which predicts the amount of noise added to the original voice. The DTW Voice pattern classification module plays a very important role as it is used for getting optimal solution between the input voice and the reference voice in the database. Support Vector Machine (SVM) is the classification method used, which automatically detects the emotion of the animal. The obtained result with these algorithms provide a better animal voice recognition system and paves the way to develop a Graphical User Interface which eases the way of using the system.

1. INTRODUCTION

The Voice recognition system plays an important role in the modern world and used for security purpose on a wide scale. The voice of human beings and animals can be analysed and it leads to identify a person /animal /music by the uploading or recording the sound. This technique is quite common and has lot of algorithms associated with it. Unlike others, animals represent a major group of the mostly multicellular organisms of the animal kingdom. Animals are very dynamic, since they can move independently as well as spontaneously.

Mammals, fishes, amphibians, birds, invertebrates, and reptiles are the six groups, grouped by the scientists among 30 million species of animals, as estimated. In animals, mammals have high frequency sound and sound recognition system can be implemented for both animals and mammals. Voice Recognition can be carried out for animals like lion, tiger, cat, bear, dog, etc.

A security system can be improved using the sound recognition technique. The sound recognition system can analyse whether the recorded or uploaded musical note belongs to any of the categories like human beings, animals, or an object. The sound recognition system became most widely used technology and human voices are analysed or recognized by patterns using different algorithms. In the same way, the recognition system can also be applied to animals by getting their voice in pattern and analysing their voice. Currently we have many animal recognition systems developed in order to apply in certain areas, such as security, research, education, etc. The most techniques are achieved through the pattern recognition, vocal recognition, vision recognition, and animal fibre recognition system.

The goal of this animal voice recognition system is to take a single or multiple animals' voices with different emotions and to develop a system that can help the human to recognize a particular animal's emotion. Vocal frequencies vary from one animal to another animal and the accuracy of detecting a particular animal's emotion is higher. The developed animal Voice Recognition System for detecting an animal's emotion is very useful for the security purpose in zoo and many places. This system can also be used in research areas where researchers and scientists study about animals to understand their state (e.g. in veterinary hospitals). The owner of the pet dogs and domestic animals to study or analyse the mood of the animal can also use the sound recognition system; their emotional and physical state can be judged by their emotions.

This system can be widely expanded from pet animals like dogs to wild animals with the help of which people living near forest areas can understand the emotion or state of the animal and can act according to that situation.

AIM

The aim of this work is to develop and experimentally evaluate an algorithm’s capability to recognize the emotion of an animal like anger, crying, happiness or howling state using ZCR, MCCF, LPC, DTW algorithms.

OBJECTIVES

1. Overview the system and analyse the problem;
2. Analyse the database that is necessary for implementation of the method;
3. Design and compare the time efficiency, precision, complexity, standard deviation, variants, mean value;
4. Obtain results, provide recommendations for method implementation and expansion of the functionality
5. Compare which method is most suitable for obtaining the voice recognition system.

2. ANALYSIS

INTRODUTION OF THE SYTEM

Speech recognition (SR) plays as interdisciplinary and acts as a sub-field of computational linguistics, which incorporates knowledge and research in the field of computer science, artificial intelligence and electrical engineering for developing methodologies and technologies. These methods enable the recognition and translation of language or voice recorded in the form of text by computers and other devices such as smart technologies and robotics, which is widely known as an Automatic Speech Recognition system (ASR).

ASR is the most preferable technology, since the user cannot go each time, uploads the voice or patterns in a system, and analyzes the emotion of the sound. SR system uses training where an individual speaker reads the text or isolated vocabulary into the system. A speaker independent system does not use training the system. Speaker dependent systems use training the system.

Speech recognition applications include user interfaces such as voice dial-up, call routing, domestic appliance control, search, and data entry. In addition, this system is used for handicapped people who cannot type, from speech to text or word order etc.: This feature has already been integrated into Google.

The sound recognition system helps in identifying the emotion of a particular speaker or an animal. Identifying the sound simplifies the task of translating the voice or sound in the system that are trained for a specific person’s voice and can also be used for authenticating a security region in the form of voice commands.

From the technology point of view, speech recognition has major innovations with different algorithms, methodologies and technologies applied throughout the different systems in different times, since the evolution of speech recognition system. Most recently, the field has focused on advances in deep learning and big data. This technology is greater in demand. This system has become user friendly to recognize a particular emotion.

2.1. EXISTING MODELS, ALGORITHMS AND METHODS

The speech recognition is developed using many major algorithms and methodologies. The system is developed to understand the voice of an animal. Most commonly applied methodologies are Hidden Markov Model (HMM), Recurrent Neural Network (RNN), Dynamic Time Wrapping (DTW), Neural Network, and Deep Learning. These models and algorithms together successfully provide a better sound recognition system. These experiments were carried out with different human and animal sounds and emotions. The previous research methods and algorithms are explained below.

2.1.1. HIDDEN MARKOV MODELS

The modern sound recognition systems in the modern world are mostly based on Hidden Markov Models. HMM is a statistical model that gives a sequence of quantities or it can be symbols or patterns. HMMs are used in speech recognition because a signal of sound e.g. static signals or a short-time static signal. Speech can be approximated as a static. Speech can be thought as a Markov model for many stochastic purposes [15].

HMMs are very popular because they can be automatically trained and are feasible to use in the automatic sound recognition system. HMM plays a wide role. In sound recognition, the hidden Markov model normally outputs a sequence of n-dimensional real-valued vectors where n being small integer, outputting one of every 10 milliseconds. The vectors in the system consists of cepstral coefficients obtained by Fourier Transforms and then cosine Transforms respectively. The Hidden Markov Model has a statistical distribution, which is a diagonal covariance Gaussians mixture for each vector. Each pattern or word will have a unique output distribution. HMM trains the words or patterns separately by concatenating the individual trained HMMs and analyze the data [18].

Cepstral normalization is used to normalize different recording conditions. HMM parameter estimation technique is used as discriminative training technique, which optimizes some of the training data, for instance Minimum Mutual Information (MMI), Minimum Classification Error (MCE), and Minimum Phone Error (MPE). HMM is the model created by using Hidden Markov Model Toolkit (HTK), in which the data is trained and segmented into files with each state assigned with feature vector $^{n}/_{m}$: n being training file, m being amount of states. The groups in HMM are assigned one state at a time from left to right as the syllables are modeled to the fact that only left to right models have been used[18].

The Viterbi algorithm is used to find the best path to the audio decoding system. Hidden Markov Model includes audio and language model information: Finite Said Transducer (FST) method. Decryption can be improved while maintaining good training to a maximum of the best training set and use a better scoring function to select the best according refined result.

2.1.2. NEURAL NETWORKS

Neural networks have emerged as an attractive acoustic modeling method ASA end of the 1980s. Since then, neural networks are used in many speech recognition systems, such as phoneme classification, and isolated word recognition and speaker adaptation.

Neural networks never take an assumption of the statistical properties of the future and have many properties, which result in attractive models for sound recognition. Neural networks support few assumptions about input functions of statistics. Probabilities of Neural networks allow exceptional physical training efficient [17]. Neural networks are rarely successful continuous recognition task, mainly due to their ability to simulate the lack of time dependencies.

Due to the inability of the Neural Networks to model time dependencies, an alternative approach is to use neural networks, pre-processing functions such as features transformation, dimensionality reduction, for the HMM based recognition.

2.1.3. DEEP FEED FORWARD AND RECURRENT NEURAL NETWORKS

Deep Neural Networks (DNN) is an artificial neural network with multiple hidden layers of units between the input and output layers. Neural networks are alike, which DNNs simulate complex nonlinear relationship. Similar to the shallow neural networks, DNNs can simulate complex nonlinear relationship. DNNs generate compositional models, where extra layers allow the lower layers of the composition of functions, providing a great learning ability and intricate patterns of language data modeling potential.

The success of DNNs large vocabulary speech recognition took place in 2010. The industry researchers, in collaboration with academic researchers, decided to build dependent HMM decision trees based on the high output layers of DNN.

DNN's fundamental principles use the raw features of the recognition system. The principle was explored successfully in the deep auto encoder on the raw spectrogram or linear filter bank features, showing its superiority over the Mel Cepstral features containing few stages of fixed transformation from spectrograms.

Hochreiter &Schmidhuber in 1997 published a recurrent neural network, which included many aspects of sound recognition and have been over taken by a deep learning method called Long Short Term Memory (LSTM). Recurrent gates called forget gates are augmented in LSTM. LSTM RNNs prevent the extinction of the gradient problem and can learn very deep life tasks involving memories of events that occurred thousands of discrete time steps before, which is very important for speech recognition. Many applications use stacks of LSTM RNNs and teach the Connectionist Temporal classification (CTC) to find RNN weight matrix that maximizes the label sequences of the training set probability by taking relevant input sequences. CTC achieves both alignment and recognition of a system. In 2007, CTC trained LSTM outperformed the first traditional sound recognition in certain applications. LSTM also improves large vocabulary speech recognition and text-to-speech synthesis for Google, Android and photo real talking head. Google's speech recognition experienced a dramatic performance jump of 49% with the help of CTC trained LSTM, which is now available through Google voice to all smartphone users.

Since the initial successful debut of DNNs speech recognition around 2009-2011 and LSTM of 2007 there was a huge new progress made. This progress (as well as future trends) has been divided into the following eight key areas:

- Scaling and speeding up the DNN training and decoding
- Discriminative training of a DNN in a sequence manner
- Processing of features by deep models with solid mechanisms
- Adaption of DNNs and related deep models
- Multi task and transfer learning by DNNs

- Designing a convolution neural networks to exploit domain knowledge of speech
- Recurrent Neural network and its LSTM variants
- Tensor based models and integrated deep generative/ discriminative models

The automatic sound recognition technique is the first and the most successful case of deep learning techniques in the present time and is embraced by both industry and academic boards. Major sound recognition systems are based on deep learning methods (e.g. Microsoft Cortona, Xbox, Skype translator, Google now, Apple siri, etc.)

2.1.4. CONCLUSION

The above given algorithms like HMM, DTW, Neural Networks , Deep feed forward neural Network (DNN) join together and successfully provide sound recognition of Human being, animals and birds. Many successful products are available in the market based on the user of using all the algorithms together, but it does not recognize animal sounds with high frequency, and also does not support the outside environment with respect to noise, external sounds and it is very difficult to recognize the emotion of an animal. As our goal is to attain the emotions of animal by recognizing there sound, these algorithms can stand as base for the Animal voice recognition product.

1. In HMMs, training the models is slightly harder because the feature vectors belonging to the particular state of the training data is not known;
2. ANNs have a complex structure, which makes the determination of the input variables and training of the system more complex;
3. DNNs possess complex structures involving high level abstraction in data and use multiple processing layers.

3. ANIMAL VOICE RECOGNITION SYSTEM

Difference between Animal Calls and Human Speech

Recently, recognizing a sound has gained much importance. Audio and video database indicates a rising effectiveness in sound recognition system [11].

There is a lot of difference between human sound and an animal sound, emotions and mood. If the word is pronounced with a thin voice, we realize that it is a child or a person is

speaking politely. Unlike human words, animal calls do not represent exact values and are not meaningless. This is because human speech and animal calls contain information on the mood and intentions. It is based on the call functions like how the pitch (low or high), volume and repetition rate is executed. In addition, most of the species have a rich vocal repertoire: for example, they can choose between the growling, barking, howling, whining, whimpering, squealing and sometimes have particularly exotic sounds such as echolocation clicks [12].

People and animals produce sounds using the vocal folds couple, located in the larynx and vocal cords vibrates with a frequency of several hundreds or thousands of times per second. The frequency is measured in hertz (Hz), where 1 Hz = 1 cycle per second.

3.1. THE ZERO CROSSING RATE

The Zero Cross Rate (ZCR) is one of the end point detection techniques in which the silent part of the sound is removed and the processing is done only on the main part of the input sound. Zero Cross Rate (ZCR) is mathematically represented as:

$$Z(n) = \frac{1}{2}\sum_{m=1}^{N} |sgn[x(m+1)] - sgn[x(m)] \text{ ---------- } (1)$$

Where

$$sgn[x(m)] = \begin{cases} +1 & x(m \geq 0) \\ -1 & x(m < 0) \end{cases} \text{ ---------- } (2)$$

Here sgn is a spectrogram function, which returns the short-time Fourier Transforms of the input signal [20].

Zero Cross Rate is also an important parameter for voice/ unvoiced classification. In the Zero Cross count, the energy is concentrated in the signal spectrum. The technique of low pass filter is used in ZCR. A low pass filter passes signals with frequency that are lower than the cutoff frequency and weaker signals that have high frequencies than the cutoff frequencies. In audio application, the filter is called high-cut filter or treble filter.

ZCR is formally defined as

$$zcr = \frac{1}{T-1} \sum_{t=1}^{T-1} II \{s_t s_{t-1} < 0\} \text{ ---------- } (3)$$

Where

S represents the signal of length T

$\pi(A)$ is an indicator function, which indicates 1 when its argument A is true and 0 when its argument A is false.

In the developed system, ZCR calculates the zero value in the variables resulting in the removal of empty space during a dog's bark.

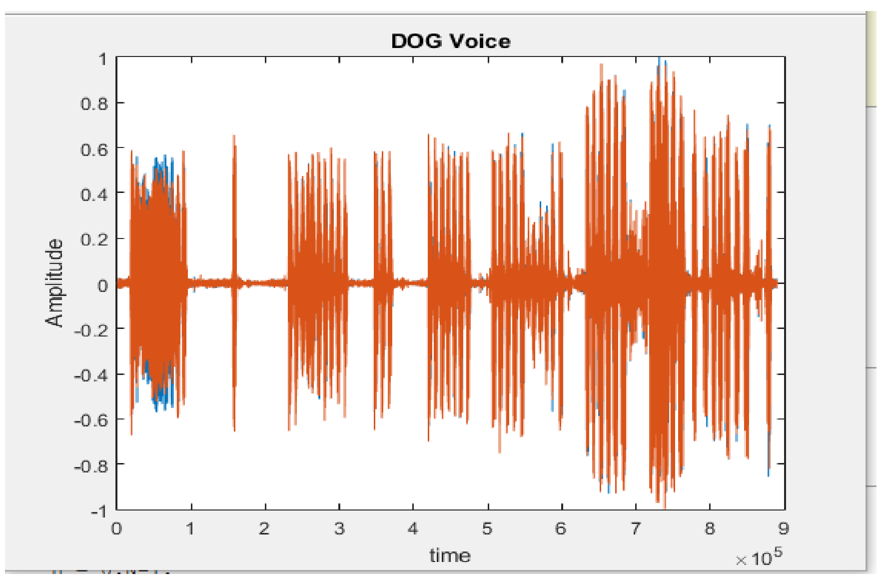

Figure 3.1: Original Dog Voice

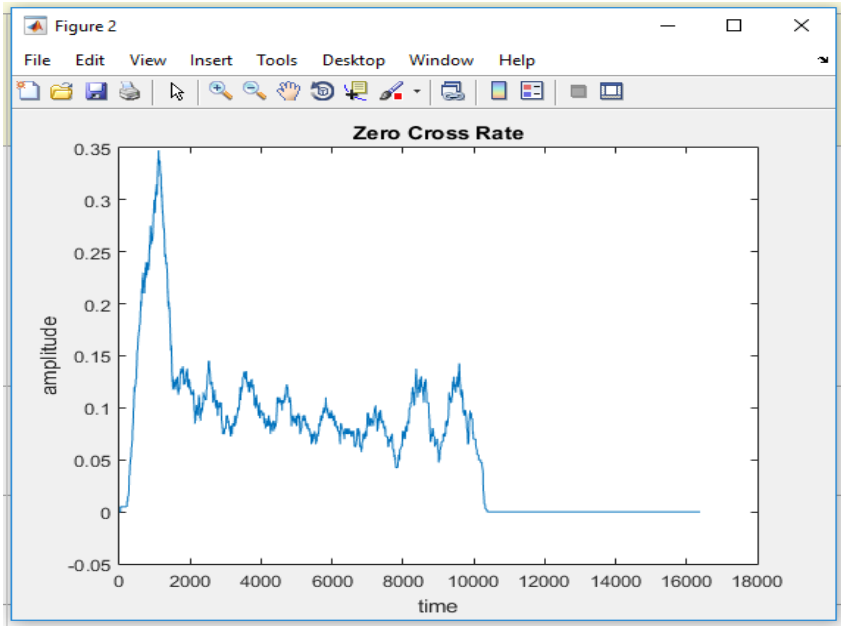

Figure 3.2: Zero Cross Rate Of One Of The Dog's Emotion

In ZCR Windowing define certain functions, such as window type, window amplitude and the length of the window is used. All of these features are used to calculate the short-term time series x. The low-pass filter method is used ZCR.

Low-pass filter is a filter that passes signals with a frequency of less than a certain cutoff frequency and reduces the signals with frequencies higher than the cutoff frequency. The attenuation for each frequency is dependent on the filter design. Low pass filter high-pass filter is the opposite. A band-pass filter is a low-pass and high-pass filter combination.

3.1.1. PROCESSING

The process of obtaining the acoustic characteristics of sound signal is known as feature extraction. Feature extraction is used for training, testing and recognition process. The features extracted in the developed system are ZCR, MFCC, DTW and LPC. The process of feature extraction involves the following steps:

1. Frame Blocking
2. Windowing
3. FFT(Fast Fourier Transform)
4. Mel-Frequency Wrapping
5. Cepstrum (Mel Frequency Cepstral Coefficients)

3.1.2. FEATURE EXTRACTION

Feature Extraction simplifies the work by summarizing large amount of data without losing its property that defines the sound.

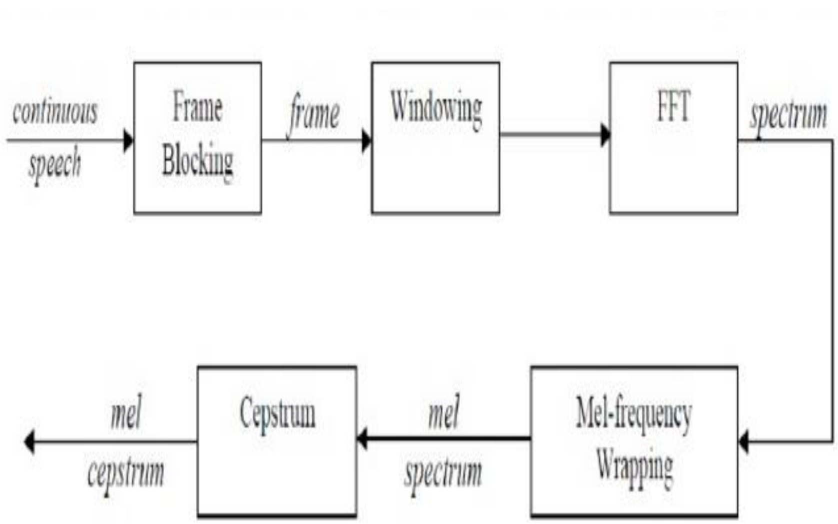

Figure 3.3: Feature Extraction

[20]

There are 200 different input sounds in the system and features extracted vary accordingly to those input sounds and classified separately for ZCR, DTW, MFCC and LPC.

3.1.2.1. FRAME BLOCKING

Research shows that characteristics of the speech signal remains stationary over a relatively short period of time interval. As a result, the speech signal is processed at short time intervals. It is divided into frames with sizes typically between 30 to 100 milliseconds. Each frame overlaps its previous frame in accordance with a predetermined size. In the developed animal sound recognition system, the sampling frequency is fixed to 10 frames per second, which is also used for training the system.

3.1.2.2. WINDOWING

Windowing is a process of removing the discontinuities at the frame edges. The windowing function is defined as W (n) , 0 < n < N- 1 ,

Where N is the sample number of each frame, the received signal is defined as;

Y (n) = x (n) w (n).

Normally Hamming windows are used. For windowing, hamming window technique is used to find high noise. The technique of FFT analysis is used for hamming.

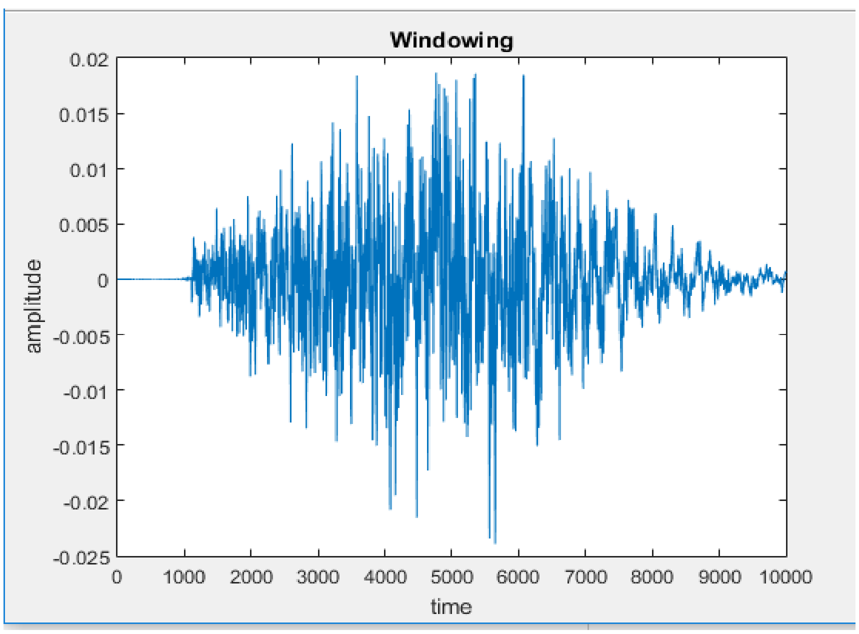

Figure 3.4: Windowing

For FFT, spectrogram function is used for feature vector extraction. Cost gives the size of the matrix.

3.1.2.3. Fast Fourier Transforms (FFT)

In the developed system, FFT identifies the least energetic parts having non-overlapping windows. FFT varies with time and amplitude of the given input sound or emotion of a dog. A term called ‘cost’ is included the system which fixes the size of the matrix (may be a 12*12 matrix or a 8*8 matrix).

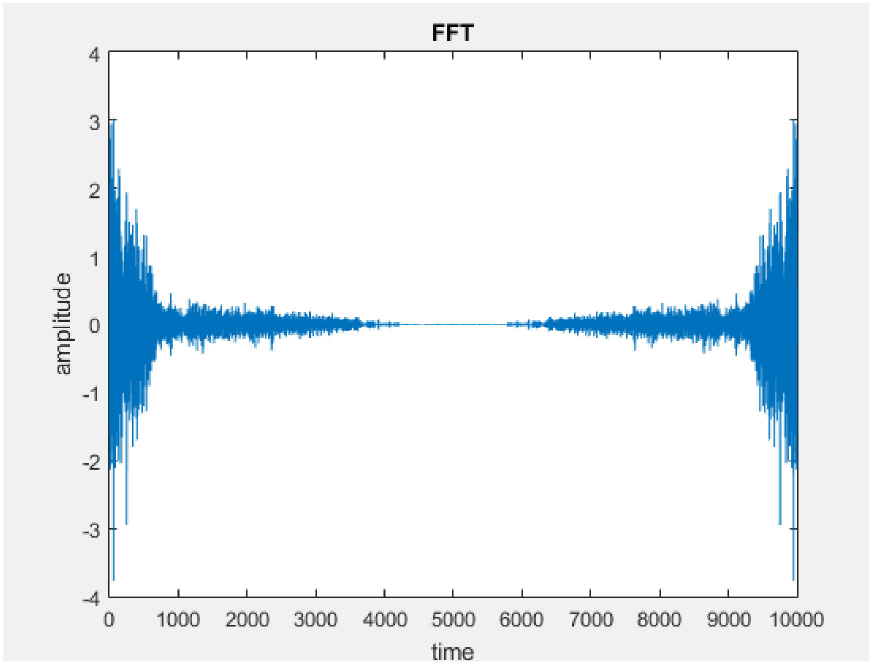

Figure 3.5: Fast Fourier Transform Varying With Time and Amplitude

3.1.2.4. Mel Frequency Wrapping

The frequency contents of sound for sound signal do not follow a linear scale. In such cases, for each interval of sound with frequency F, a scale called the Mel – scale measures the intuitive pitch.

The frequency perceived by human ear will be nonlinear. The Mel scale band pass filter passes the signals for each frame, which can mimic the sound to human ear. The Mel frequency scale ranges below 1000Hz for linear frequency spacing and above 1000Hz for logarithmic spacing. Mel frequency can be calculated by using the formula:

$$mel(f) = 2595 * log_{10}\left(1 + \frac{f}{700}\right) \quad \text{----------} \ (4)$$

3.1.2.5. CEPSTRUM

Cepstrum is the Fourier Transform of log with the unwrapped phase of Fourier Transforms. Cepstrum is mathematically given as

$$Cepstrum\ of\ a\ signal = FT\big(log\big(FT(the\ signal)\big) + j2\pi m\big) \text{ --------- (5)}$$

Where

m is the integer to unwrap the angle or the imaginary part of the complex log function.

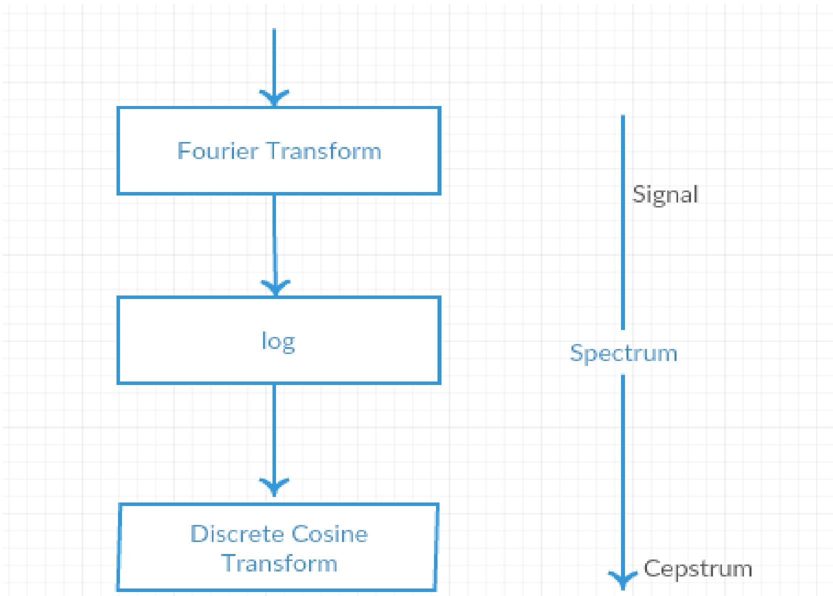

Figure 3.6: Block Diagram Of Cepstrum

3.2. MEL FREQUENCY CEPSTRUM COEFFICIENTS

Mel Frequency Cepstrum Coefficients is mainly used for feature extraction. MFCC analyzes the frame duration, frame shift, pre emphasis coefficient high pass filter range, number of filter channels and number of cepstral coefficients (IFFT). MFCCs are derived from cepstral representation of an audio clip. MFCC converts logarithmic Fourier Coefficients to Mel scale. The sequence of computation of MFCC is illustrated as:

Signal →FT→log→Mel→DCT→MFCC

MFCC is the formation of human peripheral auditory system. The frequency content of sound signals does not follow a linear scale. In such cases, Mel scale measures the pitch or tone for each interval of sound with frequency f [21].

MFCC is proved to, be more efficient and is calculated as in the following steps

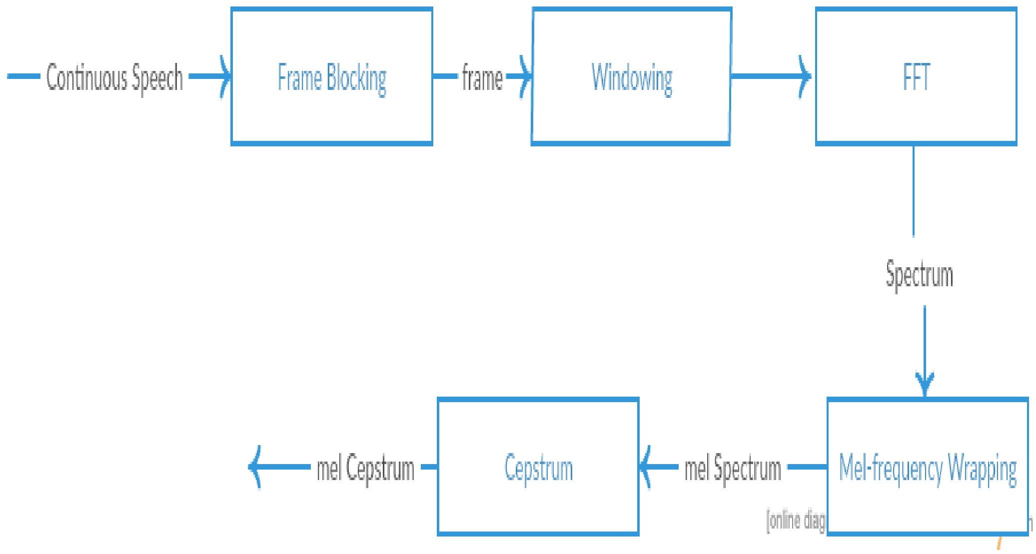

Figure 3.7: Sequence of Steps Involved in MFCC [20]

MFCC uses the Tri filter bank technique, which does not consider the zero values in the system. In the final process, the log Mel spectrum is converted back to the time domain resulting in MFCC.

3.3. PERCEPTUAL LINEAR PREDICTION

In contrast to pure linear predictive analysis of sound, perceptual linear prediction (PLP) modifies the short-term spectrum of sound by several psychophysically based transformations. Perceptual linear prediction, similar to LPC analysis, is based on the short-term spectrum of sound. This technique uses the concepts from the psychophysics of hearing to derive and estimate an auditory spectrum:

1. The critical-band spectral resolution,
2. The equal-loudness curve

3.4. SPEECH CLASSIFIER

The ASR problem depends on much broader theme of science and engineering so-called recognition. The recognition is to classify the objects of interest to one of the number of categories or classes and is called models and in our case, it is a sequence of acoustic vectors that are extracted from the input sound techniques in the previous section. Classes here refer to the individual sound generators.

3.5. DYNAMIC TIME WRAPPING

Dynamic Time Wrapping calculates the distance between the input sound and the reference sound inside the database. DTW algorithm is used for measuring similarity between two temporal sequences, which may vary with time and feature. DTW is used to find the minimum wrapping path by using the feature vectors such as mean value, standard deviation, variants and elapse time. Dynamic Time Wrapping is a measurement of the similarity between two sequences, which may vary in time or speed of the algorithm. DTW is applied to audio, video and graphics in which data can be converted into a linear representation system. Generally, it is a method that allows the computer to find an optimal match between the two sequences mentioned, within certain limits, e.g. sequences are wrapped in a non – linear manner to match each other [20].

In the designed animal voice recognition system, DTW plays the role of finding the minimum wrapping path, which enables us to analyze the feature vectors resulting the ease of recognizing the emotion of an animal.

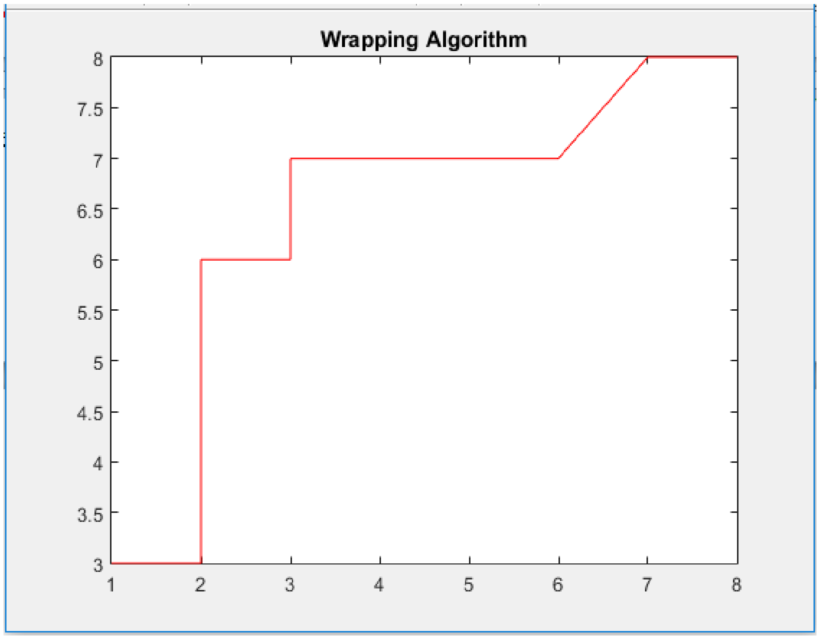

Figure 3.8: Dynamic Time Wrapping of a Dog's Emotion

DTW is a well-known application used in automatic sound recognition system, which helps to cope up with different sound input speeds. In DTW, frame blocking is used to separate frames to bocks i.e., blocks are allotted by taking 10 seconds of each frame.

3.6. LINEAR PREDICTIVE CODING

Linear Predictive Coding is a feature based extraction method widely used by developers for sound/ speech recognition system. The main reason of using LPC is that in LPC, the sound production can be modelled by using linear predictive analysis. In LPC, the vocal tract parameters are extracted from a given sound. LPC analysis the signal of a sound by calculating the formants and removing the effects from the sound signal, estimates the intensity and frequency of the remaining sound [21]. Formants are the vocal tracts (mouth/throat) of a sound generator by its resonance.

Sound signals in Linear Predictive Coding vary with time; this means that the process is divided into small parts of the sound signal, which are called frames – 30 to 50 frames per second. These results showing the extra noise added to the original dog's sound. LPC algorithm gives a vector of coefficients that represent a smooth spectral envelope. The input and output relationship of the filter is given by the linear differential equation:

$$u(n) = s(n) + \sum_{i=1}^{10} a_i\, s(n-1) \text{ ---------- } (6)$$

Where

u (n) is an innovation of the signal

s (n) is the original signal

a_i is the coefficient of the filter

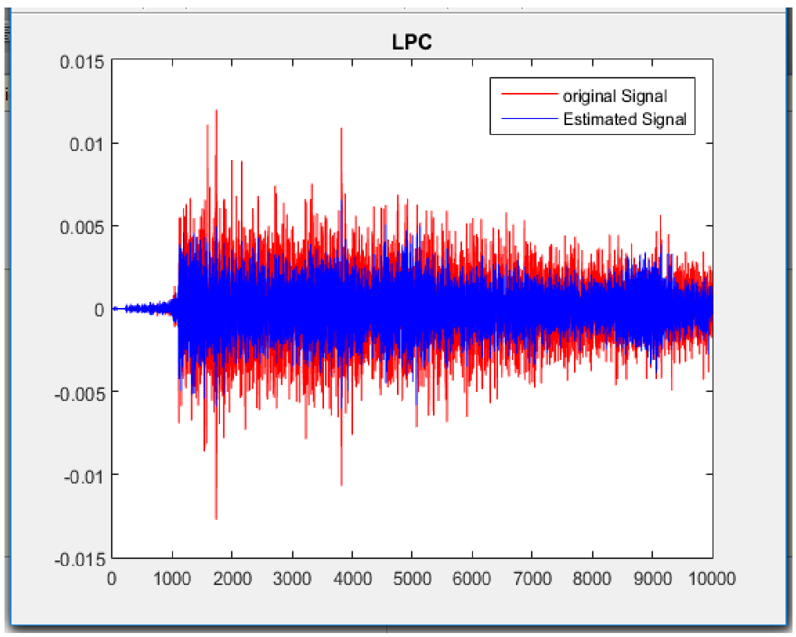

Figure 3.9: Linear Predictive Coding Showing the Original and the Estimated Signal

The above diagram shows the original signal and estimated signal of a dog's sound. The optimal values of the filter coefficients are obtained by minimizing the Mean Square Error (MSE):

$$e[n] = s[n] - \hat{s}[n] \rightarrow \min(E(e^2[n])) \text{ ---------- } (7)$$

Where

E [n] is the mean square error.

LPC uses the method of correlation to generate a pattern from the features.

3.7. SUPPORT VECTOR MACHINE (SVM)

Support Vector Machines (SVM) analysis the data used for classification and regression. Training examples are categorized into two parts: one in which the SVM training algorithm builds a new model and the other makes a non-probabilistic binary linear classifier. In the developed system, the training set is give as dog's emotion like angry, crying, happy and howling.

In this system, the classifier takes the input of N- dimensional vectors and then maps the training data into N-dimensional space. SVM constructs a hyper plane in which the data points of the two classes are separated linearly and are represented by both the classes. The technique of Kernel function or RBF Kernel is applied as a solution, which maps the data from the N-dimensional space and transforms it into N+1 dimensional space. During this process, some of the data that is not consistent will be ignored [18].

The RBF Kernel of sample x and x' that are represented as feature vectors in an input space is defined as:

$$K(x, x') = \exp(-\frac{||x-x'||^2}{2\sigma^2}) \quad \text{----------} \ (8)$$

Where $||x-x'||^2$ is a squared Euclidean distance between the two feature vectors and σ represents a free parameter. An equivalent and a simpler definition includes a parameter $\gamma = \frac{1}{2\sigma^2}$:

$$K(x, x') = \exp(-\gamma||x - x'||^2)$$

RBF kernel value decreases with distance and ranges between zero and one (when x = x'); as an interpretation of similarity measure. Kernel function in feature space has an infinite number of dimensions; for $\sigma = 1$:

$$exp\left(-\frac{1}{2}\ ||x - x'||^2\right) = \sum_{j=0}^{\infty} \frac{(x^T x')^j}{j'}\ exp\left(-\frac{1}{2}\ ||x||^2\right)\ exp\left(-\frac{1}{2}\ ||x'||^2\right) \quad \text{--------} \ (9)$$

3.8. USER INTERFACE WITH MATLAB

Graphical User Interface created with Matlab, allows the user to upload a directory, an audio file or record an audio file. Once the input sound is given, the system will analyze the sound using automatic voice recognition system developed using the ZCR, MFCC, DTW and LPC algorithms; and gives the output on accuracy of the algorithms and emotion of the animal.

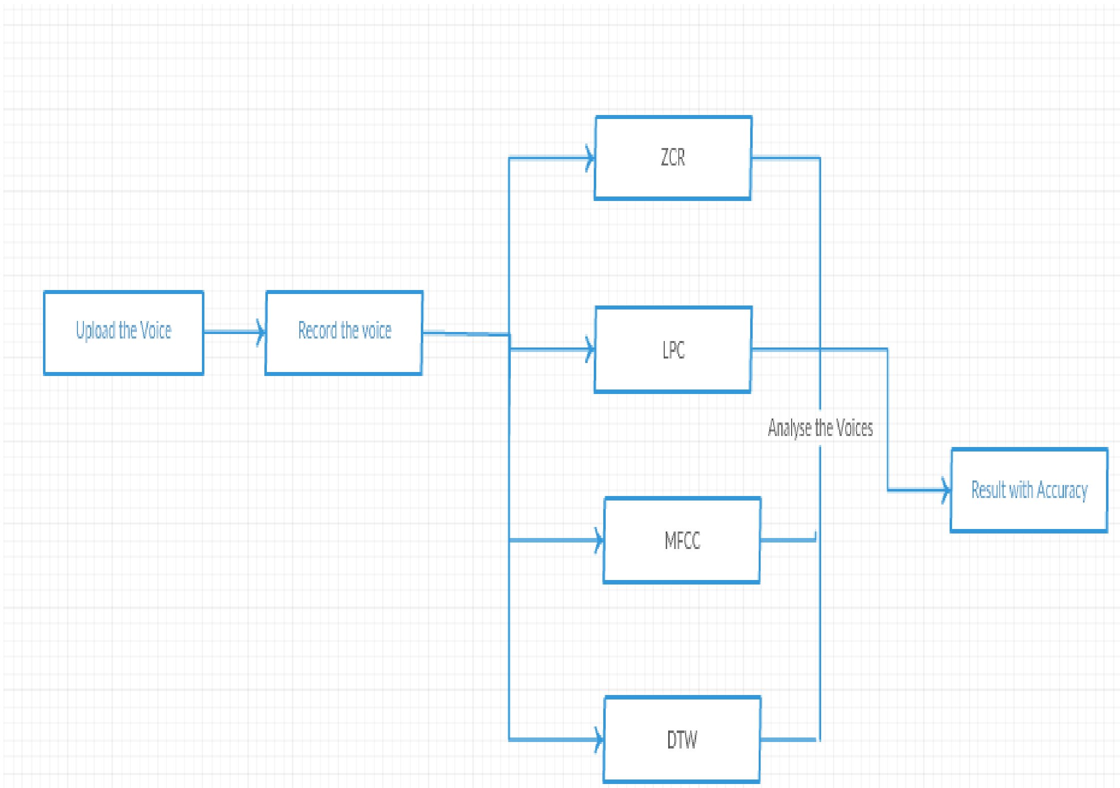

Figure 3.10: User Interface With ZCR,MFCC, DTW and LPC Algorithms

4. EXPERIMENT RESULTS

An Animal Sound Recognition System has been developed using ZCR, MFCC, DTW and LPC algorithms along with SVM classification method. Over 200 sounds of different dog species have been trained and tested and different accuracy levels were obtained. Experiments were performed in Matlab and the results obtained for recognizing a dog's emotion is given in [1].

A 50 set data for each emotion of angry, crying, happy and howling were passed and the accuracy rate was analyzed separately for ZCR, MFCC, DTW and LPC algorithms. Out of these four algorithms, DTW and LPC possess a higher accuracy level ranging between 65% and 90% than that compared to ZCR and MFCC with accuracy level ranging between 55% and 80% in detecting a dog's emotion.

4.1. ACCURACY RATE FOR RECOGNIZING DOG'S EMOTION

The results given below show the accuracy rate of each algorithm and the algorithm's efficiency in recognizing a dog's emotion.

Table 4.1 shows the accuracy rate of different dog's emotion using Dynamic Time Wrapping algorithm and its efficiency to detect a dog's emotion.

ANGRY	CRYING	HAPPY	HOWL
93	96	99	91
92	95	95	86
91	92	85	84
89	90	83	77
87	84	82	76
85	79	80	75
83	77	77	74

Table 4.1: Accuracy Rate For DTW

The above table shows that DTW has an efficiency range of 80%-90% in recognizing a dog's emotion: Angry- 88.57%, Crying- 87.57%, Happy- 85.85% and Howling- 81.28%.

The graphical representation of accuracy rate of DTW is shown as:

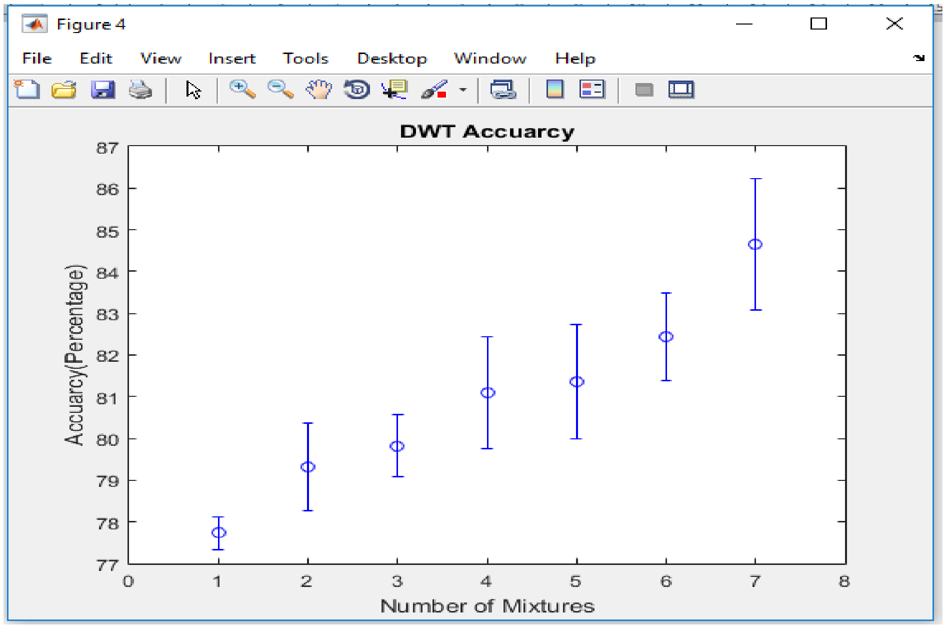

Figure 4.11: Accuracy of Dog's Emotion Based on DTW

Table 4.2 shows the accuracy rate of different dog's emotion using Linear Predictive Coding (LPC) algorithm and its efficiency to detect a dog's emotion.

ANGRY	CRYING	HAPPY	HOWL
97	92	94	96
93	90	91	75
86	67	69	72
84	64	68	70
83	63	67	66
78	55	65	55
76	48	57	38

Table 4.2Accuracy Rate For LPC

The above table shows that LPC has an efficiency range of 65%-90% in recognizing a dog's emotion: Angry- 85.28%, Crying- 68.42%, Happy- 73% and Howling- 67.42%.

The graphical representation of accuracy rate of LPC is shown as:

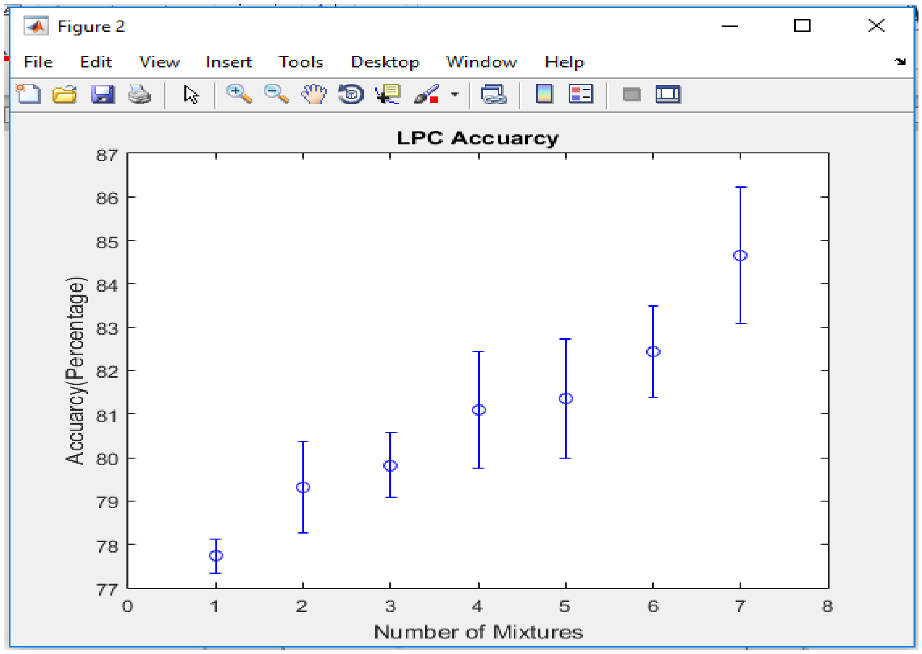

Figure 4.12: Accuracy of Dog's Emotion Based On LPC

Table 4.3 shows the accuracy rate of different dog's emotion using Zero Cross Rate (ZCR) algorithm and its efficiency to detect a dog's emotion.

ANGRY	HAPPY	CRYING	HOWL
91	98	98	99
85	97	91	81
82	80	86	77
79	77	74	66
78	73	72	53
74	80	66	60
70	49	65	61

Table 4.3: Accuracy Rate For ZCR

The above table shows that ZCR has an efficiency range of 70%-80% in recognizing a dog's emotion: Angry- 79.85%, Crying- 79.14%, Happy- 78.85% and Howling- 71%.

The graphical representation of accuracy rate of LPC is shown as:

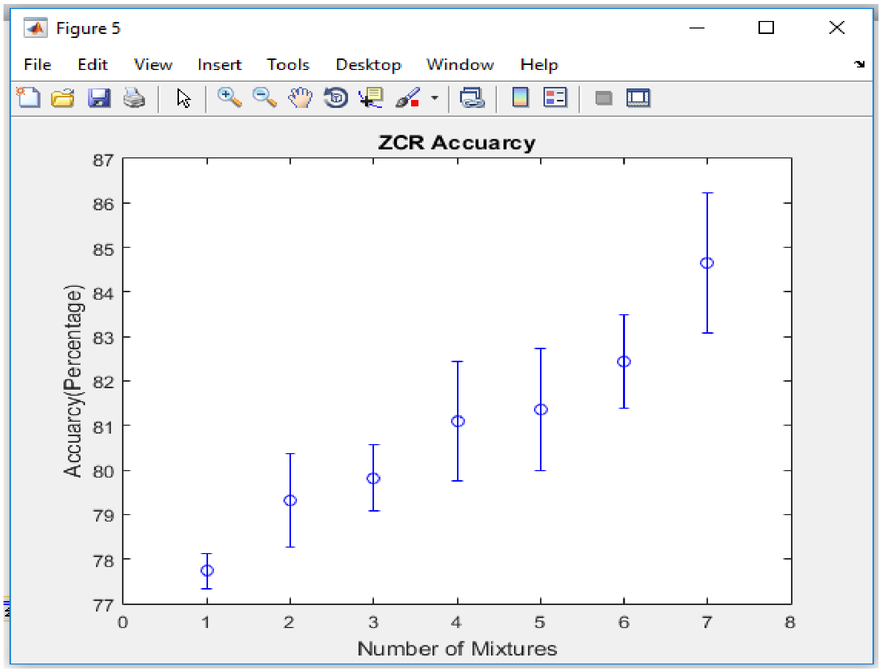

Figure 413: Accuracy of Dog's Emotion based On ZCR

Table 4.4 shows the accuracy rate of different dog's emotion using Mel Frequency Cepstrum Coefficient (MFCC) algorithm and its efficiency to detect a dog's emotion.

ANGRY	CRYING	HAPPY	HOWL
86	91	86	70
80	69	84	62
60	65	65	68
52	62	46	57
51	58	48	56
50	54	42	52
46	50	40	50

Table 4.4: Accuracy Rate for MFCC

The above table shows that MFCC has an efficiency range of 55%-65% in recognizing a dog's emotion: Angry- 60.71%, Crying- 64.14%, Happy- 58.71% and Howling- 59.28%.

The graphical representation of accuracy rate of MFCC is shown as:

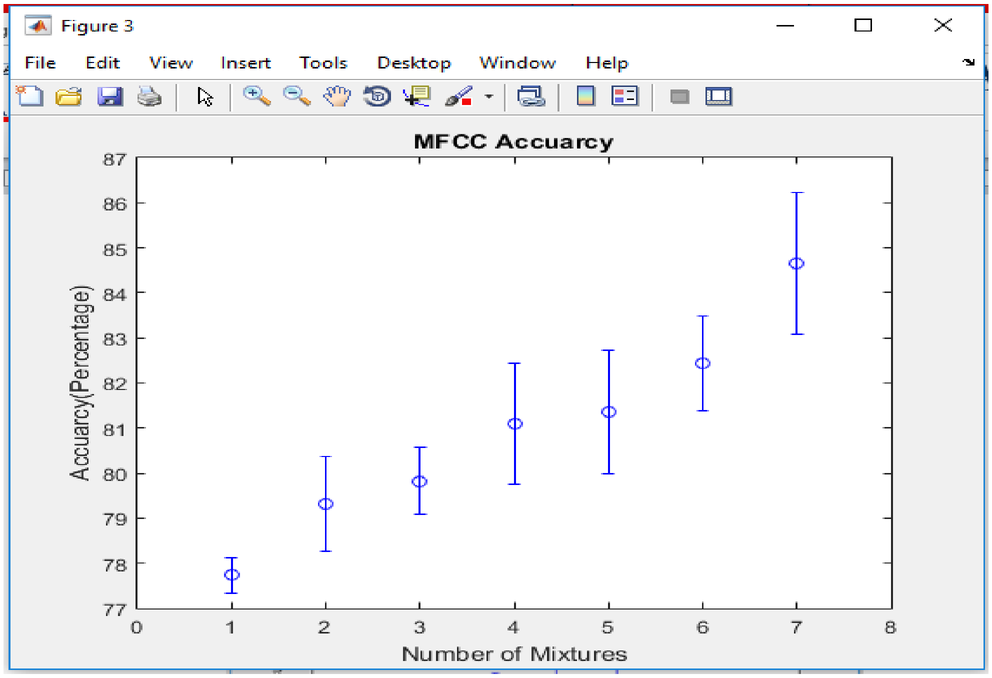

Figure 414Accuracy of Dog's Emotion Based on MFCC

Table 4.5 gives us the overall signal accuracy rate of different dog's emotion. The accuracy rate varies from 80% to 90%inspite of environmental sounds and disturbances by associating all four algorithms in animal sound recognition system.

ANGRY	CRYING	HAPPY	HAWL
83.8127	87.3288	84.1281	83.3248
82.4759	82.5764	83.6625	80.7626
80.5178	82.8322	82.8021	79.5218
80.0096	82.8687	82.251	79.5607
80.5178	80.5687	80.4942	78.8405
79.3806	79.7435	80.4844	77.4738
79.3805	77.2398	78.0256	77.6816

Table 4.5: Overall Signal Accuracy Rate

The below graph shows the overall signal accuracy rate of: ZCR, MFCC, DTW and LPC algorithms.

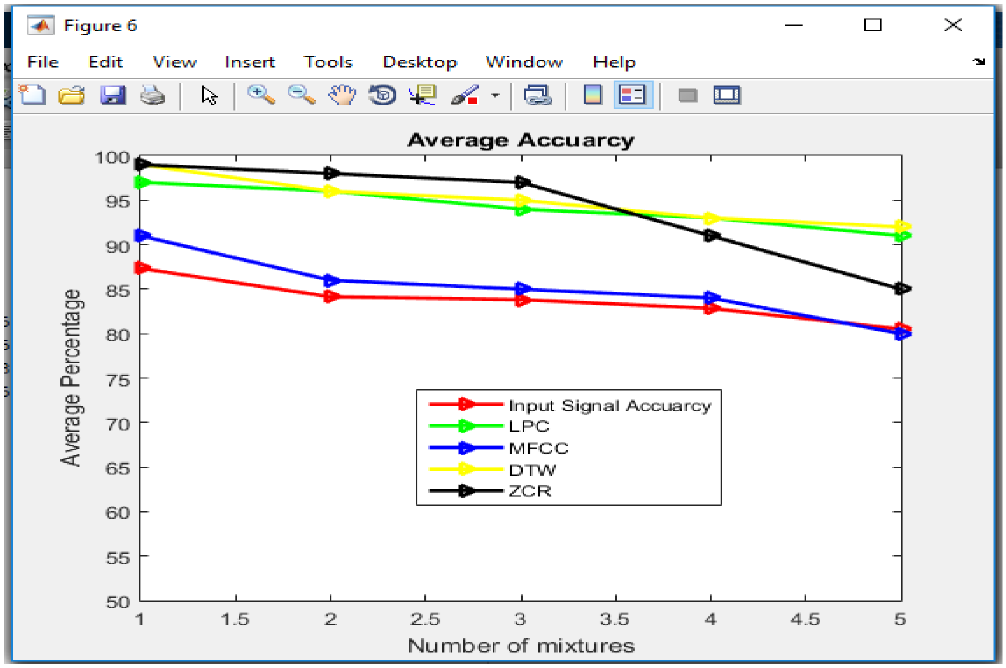

Figure 415: Overall Accuracy Of Dog's Voice

4.2. GRAPHICAL USER INTERFACE

The Graphical User Interface (GUI) created, enables the user to upload a directory with the help of upload button; upload a file using the upload file button; record a sound using the record option and finally get the output results and can also compare the accuracy rate of each algorithm used.

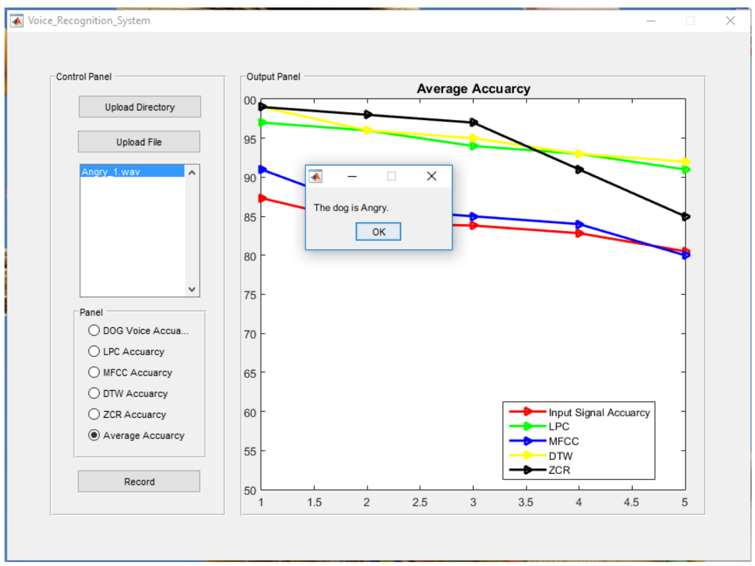

Figure 416: User Interface

4.3. LIMITATION

The only limitation with the ZCR, MFCC, DTW and LPC algorithms is that they respond to very high sound, which means, animals with high sound and frequency can be recognized and analyzed; whereas human sound and the sound of smaller animals and birds with low frequency levels is difficult to analyze.

5. CONCLUSION

In this paper, Animal Sound Recognition Pattern has been developed successfully with the help of four algorithms: ZCR, MFCC, DTW, LPC and SVM classification method. A data set of 200 sounds of different dog species were given as input to the system and the experimental results were obtained successfully. Algorithm analysis and acquired experimental results allowed to select the most time-efficient, simple and precise algorithms.

1. The analysis of the database allowed the composition of the requirements for a new Animal Sound Recognition system. The system is based on the precise definition of the intuitive modelling principles and time- efficient algorithms.
2. The comparison of feature vectors: mean value, standard deviation, variance and elapse time has been made and a desired result has been achieved.
3. Based on the experimental results, the used ZCR algorithm can accurately detect the silent voice of the input sound before it is removed. A more compact and a less redundant part of the input voice was successfully extracted from the main the part of the input voice by using the MFCC algorithm. The DTW algorithm determined the minimum wrapping path of the input sound in a more efficient manner. LPC algorithm predicts the amount of extra noise added to the original sound so that the signal accuracy of the original sound is extracted successfully.
4. The analysis of experimental results show that DTW and LPC algorithms have a higher accuracy rate ranging from 65% to 90% than that compared to the accuracy rate of ZCR and MFCC ranging from 55% to 80%; these results have been provided and compared with the methodologies associated with the system developed and results show that DTW and LPC algorithms would be better suitable for the Sound Recognition System.

FUTURE SCOPE

The Animal Voice recognition system is not only used for veterinary purposes, but can also be used in the wild animals' voice recognizing system too; in a dense forest or place where wild animals can be frequently found. This system can be used to detect the mood of animal, which might be important to the people living near the forest; as well as for people going for a wild forest trekking or conducting research

REFERENCES

1. About.com, "Animal Groups - The 6 Basic Animal Groups," http://animals.about.com/od/zoologybasics/tp/sixbasicanimalgroups.htm, Access on October 30, 2010
2. R. B. Chen and S. J. Zhang, "Video-based face recognition technology for automotive security," Mechanic Automation and Control Engineering (MACE), International Conference, pp. 2947 – 2950, 2010
3. Liying Lang and Hong Yue, "The Application of Face Recognition in Network Security," Computational Intelligence and Security. CIS '08. International Conference, vol. 2, pp. 395 – 398, 2008
4. Ming Zhao and Tat-Seng Chua, "Markovian mixture face recognition with discriminative face alignment," Automatic Face & Gesture Recognition. FG '08. 8th IEEE International Conference, pp. 1 – 6, 2008
5. R.A. Rashid, N.H. Mahalin, M.A. Sarijari, and A.A. Abdul Aziz, "Security system using biometric technology: Design and implementation of Voice Recognition System (VRS)," Computer and Communication Engineering. ICCCE 2008. International Conference, pp. 898 – 902, 2008
6. Chia-Te Chou, Sheng-Wen Shih, Wen-Shiung Chen, V.W. Cheng, and Duan-Yu Chen, "Non-Orthogonal View Iris Recognition System," Circuits and Systems for Video Technology, IEEE Transactions, vol. 20, pp. 417 – 430, 2010
7. Kaisheng Zhang, Jiao She, Mingxing Gao, and Wenbo Ma, "Study on the Embedded Fingerprint Image Recognition System," Information Science and Management Engineering (ISME), International Conference, vol. 2, pp. 169 – 172, 2010
8. Shou-Jue Wang, and Xu Chen, "Biomimetic (topological) pattern recognition - a new model of pattern recognition theory and its application," Neural Networks. Proceedings of the International Joint Conference, vol. 3, pp. 2258 – 2262, 2003
9. Liming Zhang, and Jianfeng Mei, "Theoretical confirmation of simple cell's receptive field of animal's visual systems and efficient navigation applications," Neural Networks. Proceedings of the International Joint Conference, vol. 4, pp. 3179 – 3184, 2003

10. L. X. Kong, F. H. She, S. Nahavandi and A. Z. Kouzani, “Fuzzy pattern recognition and classification of animal fibers,” IFSA World Congress and 20th NAFIPS International Conference, vol. 2, pp. 1050 – 1055, 2001

    A. Houpt Katherine, “Domestic Animal Behavior for Veterinarians and Animal Scientists,” John Wiley and Sons, 5th edition, 2010

11. Mitrovic D., Zeppelzauer M., and Breiteneder C., “Discrimination and retrieval of animal sounds,” Multi-Media Modelling Conference Proceedings, 12th International, 2006
12. Moscow Zoo, “Gallery of animals’ sounds,” http://www.moscowzoo.ru/get.asp?id=C130, Access on October 30, 2010
13. Seman. Noraini, Bakar. Zainab Abu, and Bakar. Nordin Abu, “An Evaluation of Endpoint Detection Measures for Malay Speech Recognition of an Isolated Words,” Information Technology (ITSim), International Symposium, vol. 3, pp. 1628 - 1635, 2010
14. Soo Yee Cheang, and A.M. Ahmad, “Malay language text-independent speaker verification using NN-MLP classifier with MFCC,” Electronic Design, ICED 2008, International Conference, pp. 1 – 5, 2008
15. Mark Gales 1 and Steve Young 2 “The Application of Hidden Morkov Models in speech Recognition”
16. Jan Cernocky “Speech Recognition- Intro and DTW”
17. http://andrew.gibiansky.com/blog/machine-learning/speech-recognition-neural-networks/
18. Scott Shaw, University of Sheffield - “An Evaluation of Birdsong Recognition Techniques”
19. http://staffwww.dcs.shef.ac.uk/people/P.Green/birdsong/Shaw2011.pdf
20. A.D.Mane, Rashmi R.A. and S.I. Tade “ Identification and Detection System for Animals from their Vocalization”- an “International Journal Of Advanced Computer Research (ISSN(print): 2249-7277 ISSN(online): 2277-7970), Volume- 3 Number- 3 Issue-11 September-2013”
21. Dalibor Mitrovic, Matthias Zeppelzauer and Christian Breiteneder- “Discrimination and Retrieval of Animal Sounds” Vienna University of Technology
22. Khalid T. Al- Sarayreh, Rafa E. Al- Qutaish, Basil M. Al- kasasbeh-“ Using the Sound Recognition Techniques to Reduce the Electricity Consumption in highways”

ANNEXES

1. COMPARISON OF ACCURACY RATES OF ALGORITHMS' USED

The below tables show the list of comparative results obtained by experimenting the input data of different dog's emotion using the ZCR, MFCC, DTW and LPC algorithms.

Table 4.6 shows the accuracy of an angry dog's sound using ZCR, MFCC, DTW and LPC algorithms and the signal accuracy of each input sound.

ZCR	MFCC	DTW	LPC	SIGNAL ACCUARY
16	52.41	82	77	77.4367
17	14.43	63	26.61	51.7806
19	20.32	81	41	77.6721
35	10	75	40	73.0161
16	50	87	33.17	78.5919
85	37	85	84	78.3497
82	60	57	76	67.6737
74	50	79	10	67.1654
52	12	89	21	78.4207
60	17	84	31	74.212
29	51	74	97	78.291
70	21	75	16	71.0874
29	51	74	97	78.291
16	50	87	33	78.5919
85	37	85	84	78.3497
79	18	91	14	78.7619
19	10	92	86	74.8503
82	28	75	17	79.3806
71	30	82	86	61.6415
19	10	92	21	74.8503
52	12	89	25	78.4207
12	44	82	46	78.374
22	80	74	37	76.0976
21	46	74	17	80.5178
67	37	91	17	75.111
52	28	75	93	79.3806
91	22	85	32	77.794
25	37	76	78	76.3948
68	52	72	59	79.5695

71	30	82	16	61.6415
58	46	83	21	77.3363
52	12	89	21	78.4207
22	31	89	25	83.8127
21	46	74	37	80.5178
16	45	93	16	72.498
78	20	76	15	82.4759
91	22	85	93	77.794
25	37	76	32	76.3948
13	10	77	46	79.4558
19	42	82	12	80.0096
58	13	71	38	78.2048
52	28	75	17	79.3806
71	30	82	59	61.6415
19	14.7	92	86	74.8503
52	12	89	21	78.4207
12	44	82	25	78.374
22	86	74	46	76.0976
21	46	74	37	80.5178
67	37	91	17	76.111
52	28	75	17	79.3806

Table 4.6: Accuracy Rate Of a Angry Dog Voice

Table 4.7 shows the accuracy of a crying dog's sound using ZCR, MFCC, DTW and LPC algorithms and the signal accuracy of each input sound.

ZCR	MFCC	DTW	LPC	SIGNAL ACCURACY
18	16	74	67	76.4779
18	16	72	67	76.4779
12	87	65	10	74.4572
35	34	76	92	79.7435
12	57	65.14	10	74.4572
11	10	64	11	74.7116
65	65	74	11	75.5274
15	19	63	31	82.5764
86	47	57	40	77.3915
51	69	79	19	76.2224
65	65	72	11	75.5274
39	41	95	27	74.2499
12	25	72	44	73.759
23	91	75	17	76.2256
38	84	73	16	71.2585
11	10	64	11	74.7116
81	28	92	36	66.8277
13	24	74	30	75.7525
12	25	72	44	73.759
40	16	77	15	63.7949
44	62	65	10	73.2385
42	26	77	64	74.5958
74	49	76	11	73.4016
34	39	74	11	87.3288
32	39	74	55	87.3288
98	25	72	32	74.3931
14	41	84	34	80.5687
11	45	64	22	77.9763
91	58	84	48	74.91
66	42	90	90	77.2536
74	49	76	15	73.4016
45	34	75	55	45.3384
98	25	72	55	74.3931
72	23	74	15	82.8322
14	50	72	63	76.325
18	34	74	67	77.2398

91	58	84	22	74.91
66	42	90	48	77.2536
14	41	84	32	80.5687
45	34	75	15	45.3384
72	23	90	15	82.8322
14	50	96	63	76.325
91	58	84	22	74.91
66	42	90	48	77.2536
14	41	84	32	80.5687
45	34	75	15	45.3384
72	23	90	15	82.8322
14	50	96	63	76.325
65	65	74	11	75.5274
24	62	65	10	73.2385

Table 4.7: Accuracy Rate of Crying a Dog's Sound

Table 4.8 shows the accuracy of a happy dog's sound using ZCR, MFCC, DTW and LPC algorithms and the signal accuracy of each input sound.

ZCR	MFCC	DTW	LPC	SIGNAL ACCURACY
17	18	74	44	74.1335
54	51	75	68	80.4944
13	31	82	10	78.0256
13	31	82	10	78.0286
13	31	82	10	78.0286
13	31	82	10	78.0286
27	16	74	16	76.1572
16	24	57	69	84.1281
13	31	82	10	78.0286
19	20	74	65	73.9308
77	15	74	31	82.251
77	15	74	31	82.251
27	16	74	16	76.1572
98	34	72	40	83.6625
98	34	74	40	83.6625
98	12	95	57	78.3542
18	46	83	17	74.8644
49	15	57	67	78.3787
23	84	99	35	76.5922
46	85	72	11	71.3202
23	84	99	35	76.5922
97	37	65	16	74.8029
51	39	65	91	80.4942
98	42	85	68	76.5922
73	37	74	47	74.8029
50	45	77	24	80.4949
80	49	83	13	76.212
98	12	95	57	75.7413
18	46	83	17	76.1028
49	15	57	67	70.5501
97	37	65	16	74.8029
23	84	99	35	76.5922
51	39	74	91	80.4949
27	16	74	16	76.1512

32	65	80	31	74.6754
98	42	85	68	76.212
47	22	62	94	82.8021
47	22	62	94	82.8021
98	42	85	68	76.212
51	39	74	91	80.4949
97	37	65	16	74.8029
23	84	99	35	76.5922
46	85	72	11	71.3202
23	84	99	35	76.5922
49	15	57	67	78.3787
18	46	83	17	74.8644
98	12	95	57	78.3542
27	16	74	16	76.1572
98	34	74	40	83.6625
27	16	74	16	76.1572

Table 4.8: Accuracy Rate of a Happy Dog's Sound

Table 4.9 shows the accuracy of a howling dog's sound using ZCR, MFCC, DTW and LPC algorithms and the signal accuracy of each input sound.

ZCR	MFCC	DTW	LPC	SIGNAL ACCURACY
26	26	75	34	59.0979
26	26	75	34	59.0979
26	26	75	34	59.0972
36	38	76	26	63.5152
99	24	64	22	68.2143
26	26	75	34	59.0979
15	17	82	55	80.7626
99	24	64	22	68.2843
13	52	74	17	78.2472
13	52	74	17	78.2472
53	70	74	75	78.9056
77	11	72	23	68.1001
13	52	74	17	78.2472
15	17	82	55	80.7626
12	18	77	37	77.6516
53	70	74	75	78.9056
36	38	76	26	63.5152
77	11	74	23	68.1001
26	20	84	38	79.5218
46	14	74	86	75.6391

51	51	74	11	83.3248
13	57	65	72	78.8405
51	51	74	11	83.3248
13	57	65	72	78.8405
81	15	72	31	73.3147
17	38	63	10	77.4738
17	38	63	10	77.4238
24	62	91	18	75.292
50	44	63	96	66.7515
66	50	86	55	65.0021
29	57	72	27	69.8796
26	20	84	38	79.5218
51	51	74	11	83.3248
17	38	63	10	77.4738
13	57	65	72	78.8405
10	56	74	19	71.9533
50	44	63	96	66.7515
50	44	63	96	66.7515
66	50	86	55	65.0021
66	50	86	55	65.0021
29	57	74	27	69.8796
51	51	74	11	83.3248
17	38	63	10	77.4738
13	57	65	72	78.8405

10	56	74	19	71.9533
29	57	74	27	69.8796
66	50	86	55	65.0021
66	50	86	55	65.0021
50	44	63	96	66.7515
23	88	74	70	79.5607

Table 4.9: Accuracy Rate Of a Howling Dog's Sound

2. COMPARISON OF FEATURE VECTORS OF EACH EMOTION

RESULT 1: Results for angry emotion of a dog; out of 50 sounds, first 20 sounds with mean value, standard deviation, Variants and elapse time are given below in the table.

MEAN VALUE	STANDARD DEVIATION	VARIANTS	ELAPSE TIME
3.2657e−05	0.0063	4.0303e−05	0.227946
0.0049	0.04683	0.0023	0.233513
−5.5946e−04	0.0016	2.6335e−06	0.251555
0.0023	0.3053	0.0932	0.233418
−6.7041e−05	0.002	3.8033e−06	0.22602
−5.7385e−06	0.178	0.0317	0.232587
−0.0014	0.0039	1.5306e−05	0.227483
6.1859e−06	0.0553	0.0031	0.228505
−0.0122	0.2607	0.068	0.227655
0.0033	0.1222	0.0149	0.239998
−2.5580e−05	0.0135	1.8313e−04	0.290908
2.3331e−06	0.0328	0.0011	0.239369
−2.5680e−05	0.0136	1.8313e−04	0.251955
−6.7041e−05	0.002	3.8033e−06	0.228554
−5.7385e−05	0.178	0.0317	0.241513
9.0851e−06	0.0018	1.0936e−06	0.232058
6.2149e−04	0.1089	0.0119	0.25904
−0.015	0.1241	0.0154	0.228077
−0.0037	0.0619	0.0038	0.227239

6.2149e−04	0.1089	0.0119	0.226296

Table 4.10: Comparison Of Feature Vectors Of Angry Emotion

RESULT 2: Results for crying emotion of a dog; out of 50 sounds, first 20 sounds with mean value, standard deviation, Variants and elapse time are given below in the table.

MEAN VALUE	STANDARD DEVIATION	VARIANTS	ELAPSE TIME
−4.6741e−04	0.1576	0.0248	0.230445
2.8275e−05	0.0079	6.2895e−05	0.227615
−2.2195e−05	0.0056	3.0517e−05	0.230442
−2.2195e−05	0.0055	3.0517e−05	0.229931
−2.2195e−05	0.0055	3.0517e−05	0.269575
−2.2195e−05	0.0655	3.0517e−05	0.225383
−1.6596e−04	0.004	1.6132e−05	0.224353
−1.7157e−04	0.005	2.4708e−05	0.274477
−2.2195e−05	0.0055	3.0517e−05	0.262855
−2.6303e−05	0.1237	0.0153	0.225286
−2.3018e−04	0.0927	0.0082	0.227753
−2.3018e−04	0.0907	0.0082	0.229985
−1.6596e−04	0.004	1.6132e−05	0.226107
−4.0039e−05	0.0113	1.2708e−04	0.276127
−4.0039e−05	0.0113	1.2708e−04	0.231944
−2.0121e−04	0.0051	2.5914e−05	0.265144

−0.0075	0.2363	0.0015	0.225095
−4.0494e−05	0.0028	7.8596e−06	0.231245
0.015	0.6781	0.4598	0.225637
−1.7651e−04	0.0168	2.8152e−04	0.227174

Table 4.11: Comparison Of Feature Vectors Of Crying Emotion

RESULT 3: Results for happy emotion of a dog; out of 50 sounds, first 20 sounds with mean value, standard deviation, Variants and elapse time are given below in the table.

MEAN VALUE	STANDARD DEVIATION	VARIANTS	ELAPSE TIME
−2.5365e−04	0.0978	0.0096	0.22977
−2.5365e−04	0.0978	0.0096	0.375484
−2.5365e−04	0.0978	0.0096	0.281414
8.7445e−05	0.1275	0.0163	0.268432
−7.0536e−04	0.0355	0.0013	0.229455
−2.5365e−04	0.0978	0.0096	0.235673
−5.5997e−04	0.0339	0.0012	0.227256
−7.0536e−04	0.0355	0.0013	0.227235
5.7846e−05	0.0157	2.4548e−04	0.241516
5.7846e−05	0.0187	2.4548e−04	0.230597
−5.1896e−04	0.074	0.0055	0.229341
−0.0038	0.2818	0.0794	0.23018

5.7846e−05	0.0157	2.4548e−04	0.229941
−5.5997e−04	0.0339	0.0012	0.235042
−1.5502e−04	0.0169	2.8672e−04	0.23469
−5.1896e−04	0.074	0.0055	0.226526
8.7445e−05	0.1275	0.0163	0.227991
−0.0038	0.2818	0.0794	0.225615
−3.9490e-05	0.0144	2.0678e−04	0.243939

Table 4.12: Comparison Of Feature Vectors Of Happy Emotion

RESULT 4: Results for howling emotion of a dog; out of 50 sounds, first 20 sounds with mean value, standard deviation, Variants and elapse time are given below in the table.

MEAN VALUE	STANDARD DEVIATION	VARIANTS	ELAPSE TIME
−1.7459e−05	0.0205	4.1897e−04	0.27616
−1.7459e−05	0.0205	4.1897e−04	0.232288
−4.9473e−05	0.02	4.1897e−04	0.24848
−3.4882e−05	0.0398	0.0016	0.228151
4.9473e−05	0.02	4.1897e−04	0.2295
1.1633e−05	0.0152	2.3038e−04	0.233648
1.2265e−04	0.0161	2.6035e−04	0.238723
−6.2326e−08	0.0071	5.0822e−05	0.23785
−0.0064	0.03	9.0179e−04	0.239111
−1.9373e−04	0.0052	2.7479e−05	0.244689
1.2265e−04	0.0161	2.6035e−04	0.229376

5.3897e−05	0.0183	3.3516e−04	0.23548
0.0056	0.0687	0.0047	0.231855
−0.0119	0.2588	0.0654	0.229808
−1.5881e−05	0.03	8.9729e−04	0.226467
1.1633e−05	0.0152	2.3038e−04	0.228813
−3.7039e−04	0.0162	2.6292e−04	0.236008
3.9008e−05	0.0251	0.3154e−04	0.230634
0.0056	0.0687	0.0047	0.240038
4.4647e−04	0.066	0.0044	0.233756

Table 4.13: Comparison Of Feature Vectors Of Howling Emotion