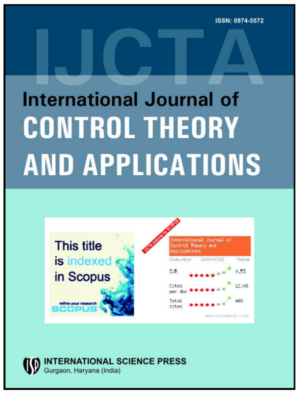

International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 10 • Number 18 • 2017

Developing an Isolated Word Recognition System for Animal voice recognition identification System

V.D. Ambeth Kumar[1], A.G. Balamurugan[2], G. Ajay[1] and S. Lokesh Kumar[1]

[1] Department of Computer Science and Engineering, Panimalar Engineering College, Chennai, India- 600123, Emails: ambeth_20in@yahoo.co.in, ggcrazyajay@gmail.com, slokeshkumar12@gmail.com
[2] Department of Computer Science and Engineering, T.J.S. Engineering College, Chennai, India- 601206, Email: balamurugan.cse.ap@gmail.com

Abstract: Speech recognition comes under the field of computational linguistics. It includes research and implementation techniques that empower the identification, recognition and translation of speech detected into text by computers. It is used in mobile phones and voice activated systems. Speech recognition is classified as isolated, continuous, dependent and independent. Isolated word recognition has a brief pause between each word spoken, whereas continuous speech recognition does not have any pause. A speaker dependent system only recognizes speech from one particular speaker, whereas a speaker independent system can recognize speech from anyone. The main objective of this paper is to use the technique of speech recognition to detect, translate and identify animal voices. This system consists of two stages training and testing. Training involves teaching the system by building a dictionary, an acoustic model for each word that the system needs to recognize (Offline analysis). Testing stage involves usage of acoustic models to recognize isolated words using a classification algorithm (Online analysis). This system can be used in animal survey processes, voice storage audio book applications to identify different animal voices in the future with more accuracy.

Keywords: Mel-Frequency cepstral coefficients (MFCC), Power spectrum, Vector quantization, Acoustic vectors and models, Dynamic Time Warping.

1. INTRODUCTION

The Basic concept of voice recognition refers to identifying the speaker, rather than what they are saying. Recognizing the speech can reduce the task of translating speech in systems that have been trained on a specific person's voice or it can be used to verify or validate the identity of a speaker as part of a security process. There are a list of voice recognition systems that are available on the market. The best system can recognize easily a minimum of thousands words. However, they generally need an extra training session during which the system becomes adapted to a particular vocal and accent. Voice recognition programs are used by many physically challenged persons to use their computers more efficiently and effectively. This type of system are called to be speaker dependent and there are other systems which requires the speaker to speak calmly and sharply and part each word with a short gap. These type of systems are called to be discrete speech systems. This needs knowledge on signal processing and statistical modeling. In this type of recognition there are two types which are acoustic modeling and language modeling, Both are important parts of today's statistically-based speech recognition methods. Hidden Markov models are largely used in many areas. Language modeling is used widely in natural language processing applications such as document classification or statistical machine translation.

The main objective followed for this work is to use the technique of speech recognition to detect, translate and identify animal voices and develop algorithm for isolated digit recognition as portrayed by the previous authors [14] and the drawback that lies here would be that comparisons would be a complex task to follow. In speech recognition the system consists of two stages training and testing. Training involves teaching the system by building a dictionary, an acoustic model for each word that the system needs to recognize (Offline analysis). Testing stage involves usage of acoustic models to recognize isolated words using a classification algorithm (Online analysis). Thus there was a system proposed in [12] Can detect each voice using its unique voice print accurately, this process recorded voice signal is processed to get its power spectrum estimation. The feature vector is derived from the power spectrum and its adjacent plots and Töeplitz matrices. This vector has proved to furnish a unique unrepeated print. but the drawback raised here was Cannot differentiate between male and female voices.

The development workflow is follows three methods which are speech acquisition, speech analysis and user (UID) development. Where speech acquisition process is done during the development of spoken words by a person. Where grammatical and syntactic learning is shown as a part of language acquisition, speech acquisition mainly focuses on the developing the speech production and speech approach over the initial stage. The next process in the workflow is Speech analytics which can be used to monitor and teach agents as well as helping in managing database. which is a specialized speech recognition software to find out the information through the recorded spoken data. This type has a software which finds out the words and tests audio ornament to detect stress in a speaker's voice. Using ART approach [8] the author identifies a method to find bird songs which was introduced to find a single existence alone but it can only find out bird songs. this method was planned using Angular Radial Transform(ART) method to find out the grey level variations within an image boundary in both angular and radial directions which further continues to shape the methods to form a spectrogram picture thus to overcome the demerit in [7] developed a signal processing technique for automatic recognition of bird spices this process was helpful because using this system we can detect species using more than one method but only over certain percentage can be found and for other variety of the bird songs of most species requires further analysis. The MPEG-7 angular radial transform used in [12] was mentioned in [4] as tools which is used for sound recognition this used to provide a header for contents of audio or indexing of audio that is further classified into two significant types text based description by category labels and quantitative descriptor using probabilistic models. This process provides a consistent framework for indication and querying sounds. Complying for a wide range of applications that have audio components. Then cacophony was a one of the problem, where the quality of the sound was not reachable thus in [5] the extracting the sound from cacophony situated on audio spectrum basis (ASB) and audio spectrum projection (ASP) of MPEG-7 audio descriptors was the method followed. On following this method there was a positive model that achieved a rate of 96% accuracy [4,5]. Feature extraction, Indexing and retrieval in animal sound databases were found in a system proposed in [2] where by using a method to extract algorithm that identifies animal sounds by analyzing the curve like time frequency trajectories in their signal. this method was very useful for searching animal sounds for various purposes but the drawback here was it was less useful or helpful in searching noise like animal sounds and also that speed of query processing was slow. Retrieval method is used in audio which was a method followed in [1,2] where implementing an audio analysis, search, classification engine which classifies sound based on loudness, pitch, brightness and bandwidth that improves the process but the main drawback in [1] is if two audio files have similar acoustic features they are difficult to comprehend. this issue was solved in [2] where sound signal can be reconstructed even at a low error rate thus retrieval audio is more effective and efficient than [1] and simpler methods can be expected in future [1,2,3].

Routinely audio processing and feature extraction was a issue that was on a wide range and various types of audio signals that are present. In [10] audio feature extraction and a multi group classification scheme was used to identify biased time- frequency subspaces using the local discriminant bases (LDB) technique. The extracted features classified as artificial and natural sounds which were sub classified as human, instrumental, animal sounds. This process has a high level of classification accuracy and with this accuracy we should detect multiple animal voices which was proposed in [11] to identify and detect animals from their vocalizations that has been developed using ZCR to detect the voice from the noise. MFCCs are used due to their ability to represent the speech spectrum in a compact form. DWT is used for voice pattern classification [10,11]. A Real time speaker recognition system is proposed which can detect a single person with very high accuracy [13] this is done by converting the acoustic audio signal into computer readable form using feature extraction and feature matching.

The approach carried out here is to recognize the sounds made by different animals and find out using a algorithm that filters the noise and finds the isolated sound of an animal and display the name of the animal. Acoustic model is also used which contains statistical representation of each word called phoneme Our system can be used in animal survey processes, voice storage audio book applications to identify different animal voices in the future with more accuracy.

2. REALTED WORK

Here we give a general overview about the related work concerning about animal voice recognition based on isolated word recognition system.

In [1] the author is explaining, Speed and capacity growth in computers and networks allows inclusion of audio as a data type in today's computer applications. Classifying and searching for audio based on its perceptual features instead of an opaque collection of bytes such as primitive fields i.e. name, size, file format. The main approach would be implementing an Audio analysis, search, classification engine which classifies sound based on loudness, pitch, brightness, and bandwidth. By doing this A particular audio can be easily identified and retrieved. But if two audio files have similar acoustic features they are difficult to differentiate. Here it accesses the audio files that are present that are only in the database and not that is represented in the database. Each sound has a background voice which has to be separated from the main voice, for this Gestalt psychology or non perceptual signal processing is used. Hence this improves the process of classification, search and retrieval of audio than the previously used methods. In [2] Rolf Bardeli proposed system for similarity search in animal sound databases using existing algorithms By using an feature extraction algorithm which identifies animal sounds by analyzing the curve-like time-frequency trajectories in their signal. By following this, animal databases sounds will be very helpful to find. But less helpful in Searching noise-like animal sounds, speed of query processing is slow. In Older systems Hidden Markov Model (HMM) was used for recognition. In this, the author uses feature extraction algorithm or getting a compact finger like prescription of the sound to be searched or identified. The audio signals produced by the animals create a curve like spectrum which can be used for identifying the particular animal by using image processing techniques. The main aim is to uniquely recognize and find similar sounds in database. For implementing this method indexing has to be done which is a time taking process so processing on large databases takes time. Thus Feature extraction, indexing and retrieval in animal sound databases can be done using this method. Guodong Guo and Stan Z. Li in [3] initially tell about Effective classification and retrieval of audio. The main approach here is perceptual Features, Mel-cepstral features and their combinations. Cepstral coefficients capture the shape of the frequency spectrum of the audio. The author uses support vector machines (SVM) with a binary tree recognition used to tackle the audio classification problem which is a newly proposed algorithm for pattern recognition. Rapid increase of audio data demands for a modern method which allows efficient and computerized content-based classification and recovering of audio from a database. SVM is usually is used for classifying audio which belongs to a single class, But here the functionality of SVM is extended to solving classifying multi class problems which during experimental review has a very less error rate compared to the traditional method An astonishing recent work is done by Wold et al. which is called to be Muscle Fish is used that distinguishes it from the previous audio retrieval work in its capability. In the Muscle Fish system, various features, such as loudness, brightness, pitch, timbre are used to represent a sound. A normalized Euclidean also known as Mahalanobis which is distance and its nearest neighbor (NN) rule used to classify the query sound into the sound classes of its the database. In [4] Michael Casey researches that efficient indexing of audio files in databases Use of a robust framework which is accepted universally. The main approach in this paper is The MPEG-7 sound-recognition tools provide a header for indexing of audio. Description is divided into two types: text-based description by category labels and quantitative description using probabilistic models. MPEG-7 sound recognition tool provides a unified interface for automatic indexing of audio using skilled sound classes in a prototype recognition framework. By doing this it provides a consistent framework for indexing and querying sounds Adaptable for a wide range of applications that have audio components. Thus Sound Recognition Classifier functions as a header that provides a description about the audio. This can be used to find the audio. For this methodology to work we have to store basic functions and HMM parameters, models of reference in a database. Using these features the header is created for efficient labeling and extraction of audio using MPEG-7. In [5] method Hyoung-Gook Kim, Nicolas Moreau, and Thomas Sikora introduced a technique to evaluate the efficiency of an audio indexing and retrieval system. Extraction of sound from noise. This technique uses low level and high level description schemes. For low level audio spectrum based signal is generated and used. For high level reduced dimensions feature modeling is used. The main approach is to estimate the effectiveness of an audio indexing and retrieval method based on audio spectrum basis (ASB) and audio spectrum projection (ASP) of MPEG-7 audio descriptors. The major merits are Audio sounds that classify into selected sound classes in real time with an accuracy of 96%. But Accuracy decreases with sound similarity. In this manuscript, they apply the ASB and ASP MPEG-7 audio descriptors to two recognition methods, a speaker recognizer and a sound categorization and retrieval system. The speaker recognizer, tested with female and male speakers, yield a high identification rate. For evaluation, Usual MFCC with delta and double-delta features were extracted. The trial results showed that the recognition rate using dimensional ASP features was minor than dimensional MFCC characteristic vectors. the ASP type using ICA basis functions verified better speaker and gender recognition act than the NASE features and the PCA basis projection sort. Thus we can identify extraction of sound under various environments and general enough to describe various sound classes. Michael Clausen and Frank Kurth proposed a system in [6] on implementing an approach for fast index based music recognition which is widely used in music information retrieval (MIR). The main idea behind this is to search and identify the music file in a database with high accuracy. The approach that tried by this author is polyphonic score data and the identification of pulse-code modulation audio material from a given acoustic waveform. The main merit here is The sizes of our search indexes are considerably small. Equally, the indexing and the retrieval performance is very high. But it is complex. The author uses both notify and audentify features for this purpose. Hence Searching and retrieval is based on score-base music and the identification of PCM audio is done here. In [7] Panu Somervuo et al. developed a Development of signal processing techniques for automatic recognition of bird species. The main approach by the author is sinusoidal modeling for high toned birds Mel-cepstrum parameters recognize bird species directly from syllables (elements) which is the building block of bird song. Where More than one method is present to detect the bird song of different species but the percentage of correct recognition is over 70%, but the average is only around 50% depending on the method followed. Thus the author concludes that certain bird species were predictable using presently available methods, while the variety of the bird songs of most species requires further analysis. Deepika M, Nagalinga Rajan found a way in [8] to create a system to identify bird songs. The main approach by the author is MPEG-7 Angular Radial Transform(ART) descriptor is used to image is used to explain the grey level variations in an image region in both angular and radial orders, to extract the shape features from the spectrogram image. The merit here is it Can be used to identify different bird species but the main drawback is only birds songs are identified. The author finally concludes that A new descriptor is proposed to identify the bird species in the recorded bird song. Yoshio Ikeda , Yohei Ishii in the year 2008 proposed a model [9] to find the changes in the voice of a cow under two conditions 1. When being hungry 2. When being separated from calf. Where the main approach is to find the variance was computed and the section of the highest variance was spectrally found by the linear predictive coding (LPC) model. The merit found in this paper is information regarding different resonance frequencies present in the voice was found but this is not accurate which lies as a major drawback in this paper. Finally the author concludes that It was found that the psychosomatic unrest of the cow below division from her calf resulted in voices with lower resonant frequencies than those of voices under hunger condition. Karthikeyan Umapathy et al found [10] a solution for Audio processing and feature extraction on a wider range and categories of audio signals. The way the authors approach this problem was by extraction audio feature and a multi group classification scheme was used to identify discriminatory time-frequency subspaces using the local discriminant bases (LDB) technique. The extracted features classified as artificial and natural sounds which were sub classified as human instrumental, animal sounds. The merit in this paper is High level of classification accuracy is found. The author finally concludes that LDB-based audio classification scheme covering a wide range of audio signals was presented. High classification accuracies were achieved using the proposed methodology.

In [11] the author explains how they identifying different voices from the various vocalization frequencies. Where ZCR is used for detecting the voice from the noise. MFCCs are used due to their ability to represent the speech spectrum in a compact form. DWT is used for voice pattern classification. The merit found here is that it can detect multiple animal voices and biggest drawback here is quality of the signal is minimal and Needs high quality signal for recognition and finally the author concludes that Identification and Detection of animals from their vocalizations can be found through this method. The problem addressed in [12] is to identify each unique voice by its unique voice print. And the approach followed by the author was to record voice signal is processed to get its power spectrum estimation. The characteristic vector is imitative from the power spectrum and its nearby plots and Töeplitz matrices. This vector has proved to furnish a unique unrepeated print. The merit here is it can detect each voice using its unique voice print accurately but cannot distinguish between male and female voices and can be easily bypassed. The brief review given in this paper manifests a new approach to the voiceprint as a means for simple data description.Roma Bharti proposed a system [13] which is a real time speaker that recognition system that has been trained for a particular speaker and verifies the speaker. The approach followed here is by converting the acoustic audio signal into computer readable form.Using feature extraction and feature matching. The merit found in this paper is it can easily detect single person with very high accuracy and the drawback here is requires good voice signal Can detect only one person’s voice at a time. The author finally concludes that This system recognizes real-time speaker with the help of stored database and can be extended with more number of users.In [14] Yu-Hsiang Bosco Chiu et al researched and raised a problem that Better recognition accuracy in the presence of backdrop noise than recognized MFCC feature extraction and the approach followed here is Maximum mutual information (MMI) criteria is used to determine the signal from the noise. The merit found in this paper is it can recognize voice even during the presence of noise and the drawback found is Uses complex databases for comparisons and finally Has higher accuracy than traditional speech recognition methods.

3. PROPOSED SYSTEM

The system architecture for developing an isolated word recognition system for animal voice recognition system can be divided into the following modules namely: voice signal procurement, signal processing, feature extraction using MFCC, HMM decoder, similarity detector, classifier, decision. Voice database is used to store voices of different animals for comparison purposes. Training models involves teaching the system by building a dictionary, an acoustic model for each word that the system needs to recognize.

First the voice of the animal is recorded using a microphone i.e. the analog voice is converted to digital voice and given to the system for voice processing and recognition.

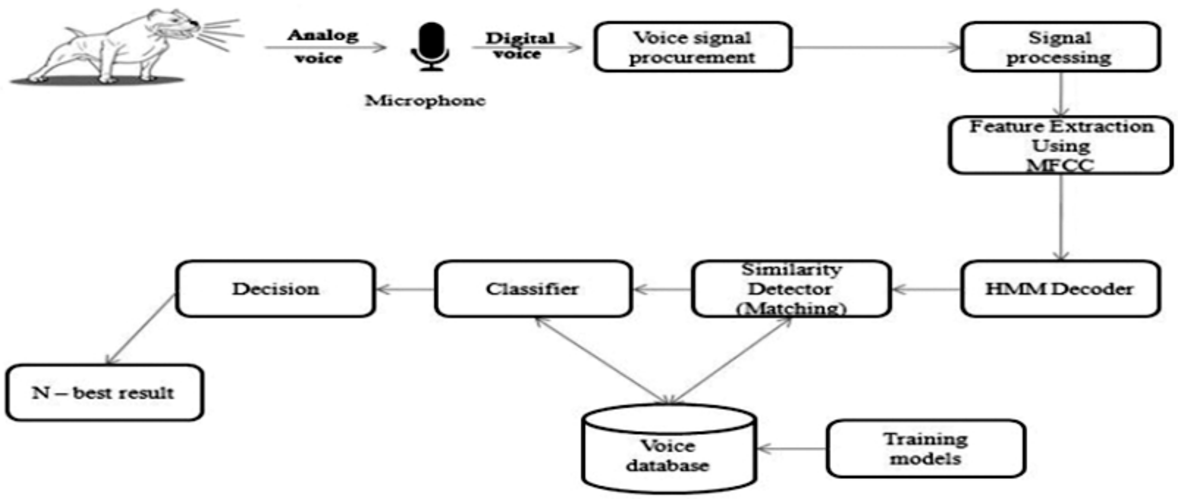

Figure 1: Architecture of the proposed system for isolated word recognition system for animal voice identification system.

3.1. Signal Procurement

The signal is procured using a distortion free microphone as mentioned above in Fig. 1, then from resulting signal undesirable information is removed to preserve the unique voice properties of the animal sound which has to be identified during signal processing.

3.2. Signal Processing

The recorded signal is processed at 44100Hz frequency, saved in 8 bit format and normalized so that the amplitude of the signal is in the range of [-1 to 1].now noise filters ,spectral noise grating and signal analysis are employed to remove the noise during signal processing and get the anti noising parameters to get a good quality signal form the noisy background present in the recorded audio signal.

3.2.1. Power continuum or Power spectrum

Power continuum or spectrum of a time sequence describes the allocation of frequency components forming that signal. According to Fourier analysis any substantial signal can be decomposed into a number of distinct

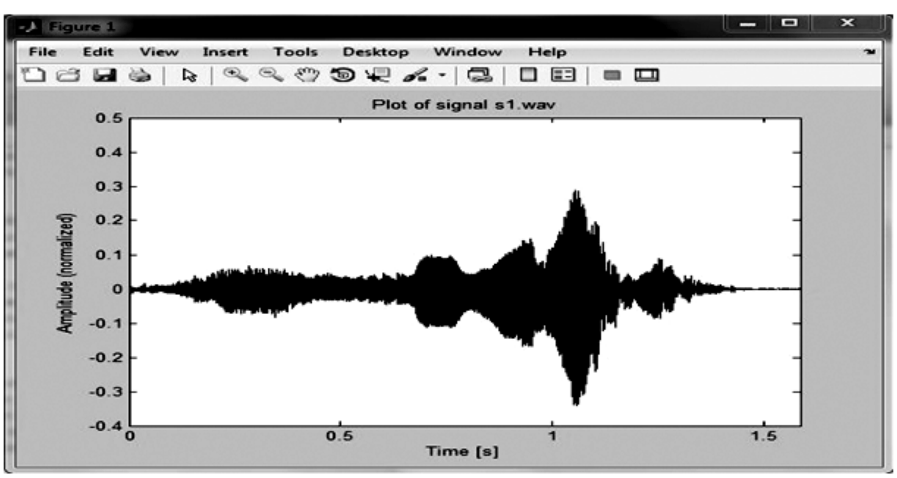

Figure 2: Power spectrum of the given input audio signal in waveform format

frequencies or variety of frequencies over a continuous range. The arithmetical average of a certain signal including noise as analyzed in stipulations of its frequency content is called its spectrum. When the energy of the signal is intensified around a limited time interval, in particular if its total energy is limited, we can compute the energy spectral density. The recorded signal is processed to get the following spectrum as show in Fig. 2

3.2.2. Spectogram

A spectrogram is a illustration of the continuum of frequencies in a sound as they differ with instance or some other variable. Spectrograms are sometimes called spectral voiceprints.

The spectral centroid of a signal is the mean of its spectral density function, i.e. the occurrence that divides the allocation into two equal parts. The spectral border frequency of a signal is an addition of the previous concept to any quantity in its place of two equal parts. The spectral density is a function of frequency, not a role of time. However, the spectral density of undersized windows of a longer signal may be calculated, and plotted against time linked with the window. Such graph is called a *spectrogram*. This is the source of a number of spectral study techniques such as the short-time Fourier transform and wavelets. The spectrogram is calculated and plotted as shown in Fig 3.

Spectrograms can be used to examine the outcome of passing a analysis signal through a signal checker such as a filter in order to ensure its performance.

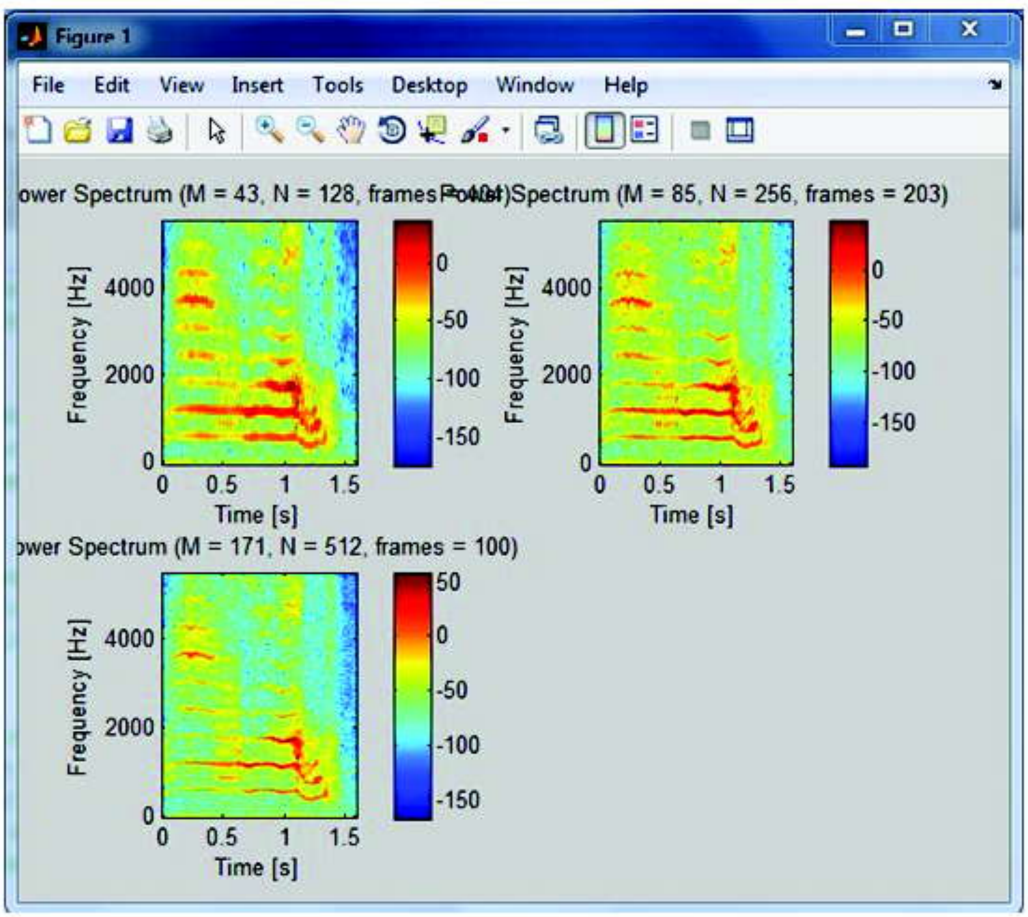

Figure 3: Spectrogram of the given input signal

Spectrograms are typically formed in two ways which are approximated as a sift bank that results as of a sequence of band pass or premeditated from the instance signal using the FFT.

3.3. Feature Extraction using Mel-Frequency cepstral coefficient

The most common technique used in automatic speech recognition system to extract features is Mel-Frequency cepstral coefficient it is used to detect the components of a audio signal that are useful for accurately detecting the linguistic content and removing environment noises, distortion, etc.

Sounds produced by human beings, animals and other species are filtered by the vocal tract which gives a particular shape to the sound. This shape determines which is produced. If we can determine this shape precisely, we can get an exact image of the phoneme being created. The shape of the vocal tract itself forms the packet of the short time power spectrum, and the objective of MFCCs is to exactly characterize this packet. An audio signal is continuously altering, so to make things easy we presume that on small time interval the audio signal does not change much. This is we divide the signal into frames (30-50ms frames per signal). If the frame is a lot smaller then we cannot obtain enough samples to get a consistent spectral estimation, else if longer the signal changes too much within the frame.

As shown in fig 4.The subsequently step is to estimate the power spectrum of every frame. For this we use Mel filter bank. The first filter is very thin and gives an hint of how to a great extent energy exists close to zero Hertz. As the frequencies get elevated filters get thicker as we grow to be less disturbed about variations. We are merely concerned in approximately how much energy occurs at every spot. The Mel level tells us precisely how to space the filter banks and how wide to place them.

The last step is to calculate the DCT of the log filter bank energies. There are two reasons why is done. First, filter banks are all overlapping each other; the filter bank energies are quite interrelated to each other. The DCT de links these energies whose diagonal covariance matrices can be used to represent the features in e.g. a

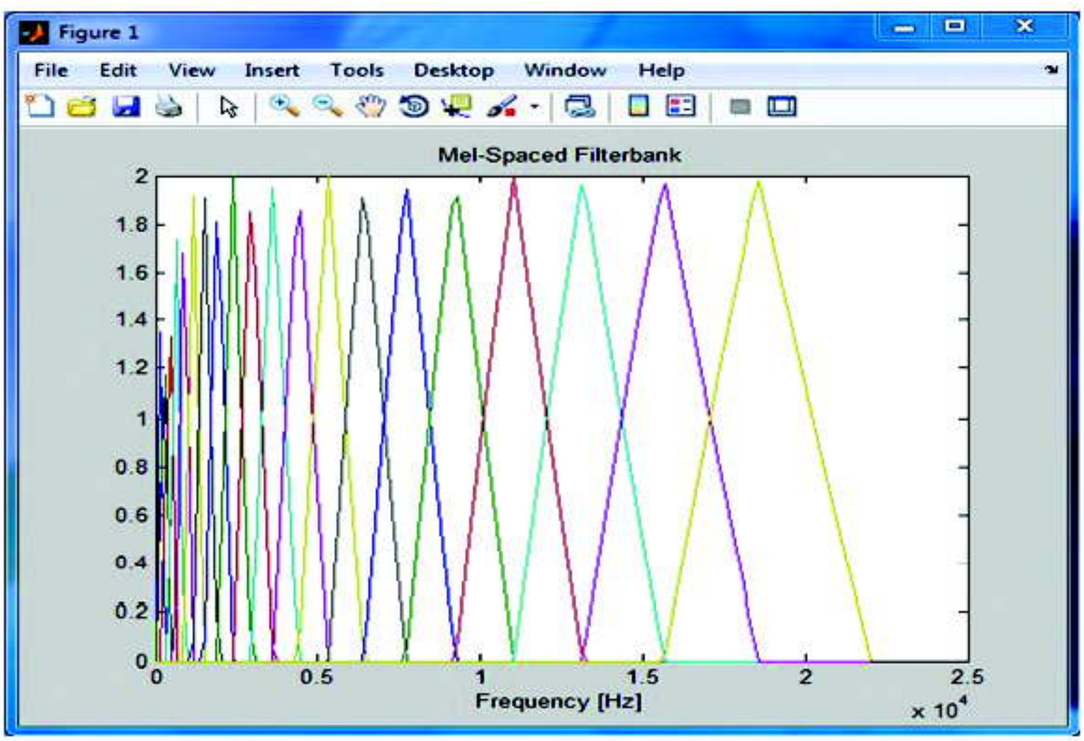

Figure 4: Mel spaced filter bank representation of the given signal

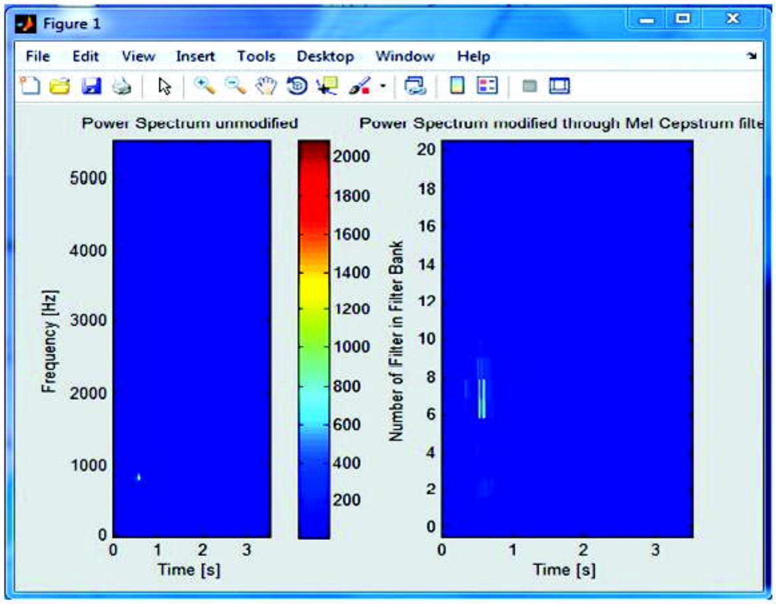

Figure 5: Representation of modified and unmodified spectrum before and after processing through Mel filter bank

HMM classifier. The elevated DCT coefficients symbolize quick changes in the filter bank energies and it turns these changes degrade ASR performance. Fig 5 shows the difference between modified and unmodified spectrum after Mel filtration

3.4. HMM decoder

HMM (Hidden Markov Model) is implemented in speech recognition because a speech signal can be represented as a piecewise stationary signal. A short instance speech can be viewed as a stationary process. Verbal communication can be consideration as a Markov model for a lot of stochastic purposes.

Another basis why HMM is well-liked is because it can be trained automatically, uncomplicated and computationally realistic to use. In speech recognition, the hidden Markov model would yield a series of n-dimensional real valued vectors. The vectors contain cepstral coefficients, which are obtained by taking Fourier transform of a small time window of speech and de linking the spectrum using a cosine transform, then considering the first coefficients. The hidden Markov model will be inclined to have in every state a statistical distribution that is a combination of diagonal covariance Gaussians, which will give possibility for each experimental vector. Each utterance, or phoneme, will have a different output allocation .A hidden Markov model for a series of utterances or phonemes is prepared by concatenating the individually taught hidden Markov models for the separate utterances and phonemes.

3.5. Classifier

Here classifier is used to classify different animal voices based on their acoustic characteristics and cepstral values. Here vector quantization is used to plot the values and classify the values based on the position of the vectors on the graph.

3.5.1. Vector quantization – plotting of 2D acoustic vectors

Vector quantization is a traditional quantization method in signal processing used for modeling of probability density functions by the allocation of model vectors as shown in fig 6. The density matching ability of vector

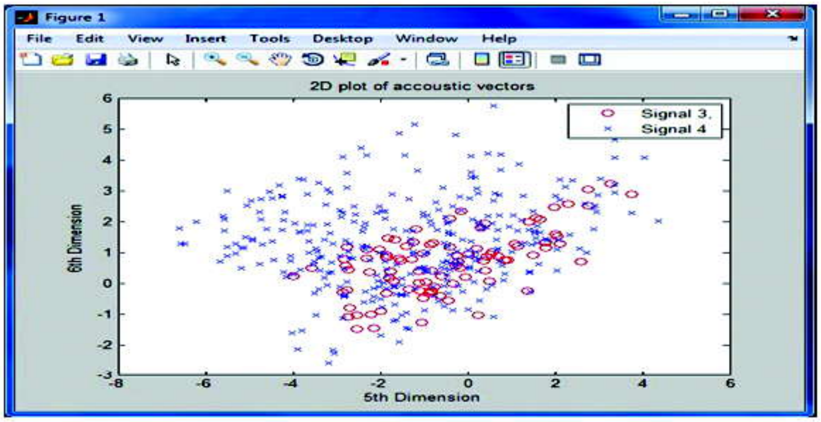

Figure 6: Plotting of 2D acoustic vectors

quantization is great, especially for identifying the density of big and high-dimensioned information. Since information points are represented by the index of their nearby centroid, frequently occurring information have low error as well as rare data high inaccuracy.

Here for recognizing a pattern, a codebook is created for each using acoustic vectors of the user. In the testing segment the quantization distortion of a testing signal is compared with the complete set of codebooks obtained in the training segment as shown in fig 7. The codebook which provides the least vector quantization distortion gives the identified user.

The main benefit of Vector quantization is in pattern recognition is it's near to the ground computational load when compared with other methods such as DTW and HMM. The major disadvantage while compared to DTW, HMM is that it does not take into explanation the sequential growth of the signals because all the vectors are diverse.

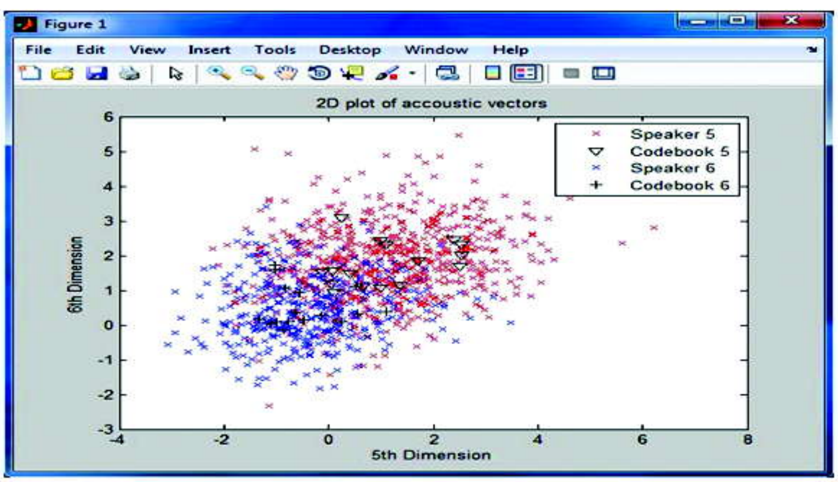

Figure 7: Plotting of vector quantization code words

3.6. Decision

All the above mention methods such as MFCC, DTW, Power spectrum , HMM, Vector quantization are used to compare and identify the voice of a animal in real time. The system mentioned above is a combinations of multiple methodologies for getting higher accuracy in recognition of voices in real life situations.

4. CONCLUSION

Isolated Word Recognition System for Animal voice recognition identification System is proposed in this work, with the objective of successfully recognizing animal voices in the outside surroundings. With this proposed system, we can identify the animal by its voice by comparing it with the samples which are already present in the database. This system can be used for animal surveying, veterinary sciences, and zoological research. In future, we intend to use the basis of this system to detect the changes in voices of animals in their different psychological state for understanding the psychological state of the animal and for use in the training of classifiers, and so we anticipate to be able to attain a higher accuracy rate in recognizing the voice of different animal present in the world.

REFERENCES

[1] Erling Wold,Thom Blum, and Douglas Keislar, and James Wheaton, "Content-Based Classification, Search and Retrieval of Audio" IEEE Multimedia, vol. 3, no. 3, pp.27 – 36, 1996.

[2] Rolf Bardeli, "Similarity Search in Animal Sound Databases", IEEE Multimedia, vol. 11, no.1, pp. 68-76, Jan 2009.

[3] Guodong Guo and Stan Z. Li, "Content-Based Audio Classification and Retrieval by Support Vector Machines", IEEE Neural Networks, vol. 14, no. 1, Jan 2003.

[4] Michael Casey, "MPEG-7 Sound-Recognition Tools", IEEE Trans. Circuits and systems for video technology, vol.11, no.6, June 2001.

[5] Hyoung-Gook Kim, Nicolas Moreau, and Thomas Sikora, "Audio Classification Based on MPEG-7 Spectral Basis Representations", IEEE Tans. Circuits and systems for video technology, vol. 14, no. 5, May 2004.

[6] Michael Clausen and Frank Kurth, "A Unified Approach to Content-Based and Fault-Tolerant Music Recognition", IEEE Trans. Multimedia, vol. 6, no. 5, Oct 2004.

[7] Panu Somervuo, Aki Härmä, and Seppo Fagerlund, "Parametric Representations of Bird Sound for Automatic Species Recognition", IEEE Trans. Audio, Speech and Language processing, vol. 14, no. 6, Nov 2006.

[8] Deepika M and Nagalinga Rajan, "Automatic Identification of Bird Species from the Recorded Bird Song Using ART Approach", Int. Conf. on innovations in engg., vol. 3, no. 3, Mar 2014.

[9] Yoshio Ikeda and Yohei Ishii, "Recognition of two psychological conditions of a single cow by her voice", Journal in Computers and Electronics in Agriculture , vol. 62. no. 1, pp. 67-72 • June 2008

[10] Karthikeyan Umapathy, Sridhar Krishnan and Raveendra K. Rao, "Audio Signal Feature Extraction and Classification Using Local Discriminant Bases", IEEE Trans. Audio, Speech and Language processing, vol. 15, no. 4, May 2007.

[11] A. D. Mane, Rashmi R. A, and S. L. Tade, "Identification & Detection System for Animals from their Vocalization", Int. Jour. Advanced computer research, vol. 3, no. 3, Sept 2013.

[12] Khalid Saeed, "Sound and Voice Verification and Identification A Brief Review of Töeplitz Approach",7th Conference Znalosti, pp. 22-27, 2008.

[13] Roma Bharti and Priyanka Bansal, "Real Time Speaker Recognition System using MFCC and Vector Quantization Technqiue", Int. Jour. Computer Applications, vol. 117, no. 1, May 2015.

[14] Yu-Hsiang Bosco Chiu, Bhiksha Raj and Richard M. Stern, "Learning-Based Auditory Encoding for Robust Speech Recognition", IEEE Trans. Audio, Speech and Language processing, vol. 0, no. 0, 2011.