*Special Issue of Engineering and Scientific International Journal (ESIJ)*
*Technical Seminar & Report Writing - Master of Computer Applications - S. A. Engineering College (TSRW-MCA-SAEC) – May 2015*
ISSN 2394-187(Online)
ISSN 2394-7179 (Print)

# Perfect and Secure Speech Recognizer

V.Niranjan[#1], P.Mohankumar[*2]
*Master of Computer Application, S.A Engineering College, Chennai-77, India*
vensanniranjan93@gmail.com, mohankumarkumar499@gmail.com

***Abstract***—Speech recognition is the new emerging technology in the field of computer and artificial intelligence. It has changed the way we communicate with computer and other intelligent devices of same caliber like smart phones. It is a major area of interest for research in this field which is related to artificial intelligence. Through this paper I am going to reduce the noisy disturbances received by speech recognizer by following some approaches and some specialized sensors. It gives a solution to reduce the disturbances which means noise and how securely the speech is recognized by not only by audio it recognizes also the mood of the person who gives a speech to it. Recognizer gives appropriate action according to command and mood of the person.

***Keywords***— *FaceRecognition, Phoneti cApproach, Cepstrum, Spectrogram, Butterworth Filter.*

## 1. Introduction

Basically, the microphone converts the voice to an analog signal. This is processed by the sound card in the computer, which takes the signal to the digital stage. Input from user is also known as utterance (Spoken input from the user of a speech application. An utterance may be a single word, an entire phrase, a sentence, or even several sentences.)This is the binary form of —1s and —0s that make up computer programming languages. Computers don't hear sounds in any other way.

I place some sensors for recognizing the mood of the person the person have to express his command not only by voice he has to express it by expression also and the command should be in same frequency not so high and not so low .The speech recognition is defined as the process of considering the spoken word as an input speech and matches it with the previously recorded speeches on basis of various parameters. This can be done by various methods. It is a process of automatically recognizing who is speaking on the basis of features of speaker of the speech signal.

You can provide the input to an application with your voice just like clicking the mouse& typing your keyboard. The acoustic signal captured by microphone is converted to a set of words and recognized words are the first results of the application like command and control, data entry, document preparation this is you can give the commands like switch to calculator, open a notepad by voice instead of using mouse or keyboard.

That system does not require the sample of the speech .Generally speech recognition is difficult when vocabulary is large or the system has many similar sounding words. When speech is generated as a sequence of words artificial grammar or language models are used to restrict the combination of words. The vocal tract system including coupling of nasal tract accurately described by the position of articulator like tongue, jaws etc.

## 2. Acoustic Phonetic Approach

Acoustic-phonetic approach assumes that the phonetic units are broadly characterized by a set of features such as format frequency, voiced/unvoiced and pitch. These features are extracted from the speech signal and are used to segment and level the speech Hidden markov modeling. Speech is split into the smallest audible entities.

The earliest approaches to speech recognition were based on finding speech sounds and providing appropriate labels to these sounds. This is the basis of the acoustic phonetic approach (Hemdal and Hughes1967), which postulates that there exist finite, distinctive phonetic units (phonemes) in spoken language and that three these units are broadly characterized by a set of acoustics properties that are manifested in the speech signal over time.

Even though, the acoustic properties of phonetic units are highly variable, both with speakers and with neighboring sounds (theso-called co articulation effect), it is assumed in the acoustic-phonetic approach that the rules governing the variability are straightforward and can be readily learned by a machine.

The first step in the acoustic phonetic approach is a spectral analysis of the speech combined with a feature detection that converts the spectral measurements to a set of features that describe the broad acoustic properties of the different phonetic units. The next step is a segmentation and labeling phase in which the speech signal is segmented into stable acoustic regions, followed by attaching one or more phonetic labels teach segmented region, resulting in a phoneme lattice characterization of the speech.

The last step in this approach attempts to determine a valid word (or string of words) from the phonetic label sequences produced by the segmentation to labeling. In the validation process, linguistic constraints on the task (i.e., the vocabulary, the syntax, and other semantic rules) are invoked in order to access the lexicon for word decoding based on the phoneme lattice. The acoustic phonetic approach has not been widely used in most commercial applications.

The Fig.1 shows how the characters are pronounced under various circumstances it gives a perfect pronunciation of the word or a character thus the phonetic chart is used to know the pronunciation of the word in an perfect manner and which helps in the verification of the voice are a speech which we speaks on to the system of acoustic phonetics approach.

ISSN 2394-187(Online)
ISSN 2394-7179 (Print)

Fig.1: Phonetics chart

## 3. Butterworth filter and spectrogram

Butterworth filter is used to remove the noise from the system. Speech signals degrade due to the presence of background noise and noise reduction is an important field of speech processing. Butterworth stop-band filter is used to minimize the disturbance from the speech signals.

The background disturbance in the signals degrades the performance and noise reduction is an important field of speech processing. The filtered audio signal waveform shows that background noise was successfully removed from a signal by using Butterworth stop-band filter.

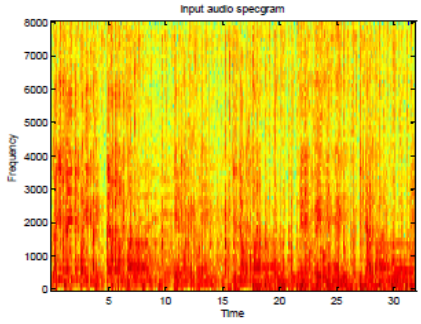


Fig.2: Input audio spectrogram

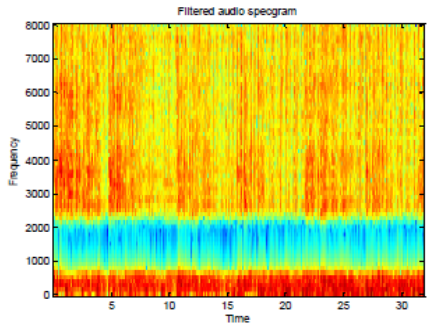


Fig.3: Filtered audio spectrogram

There is a better representation domain, namely the spectrogram. This representation domain shows the change in amplitude spectra over time.

It has three dimensions:

X-axis: Time (ms)

Y-axis: Frequency

Z -axis: Color intensity represents magnitude

The complete sample is split into different time-frames (with a 50% overlap). For every time- frame, the short-term frequency spectrum is calculated. Although the spectrogram provides a good visual representation of speech it still varies significantly between samples. Samples never start at exactly the same moment, words may be pronounced faster or slower and they might have different intensities at different times.

They are calculated from two different samples. As you can see, they both show somewhat the same pattern, but the second sample is shifted in time compared to the first sample. As these patterns vary so much, makes them useless as input for the neural network unless some more signal preprocessing is performed.

Spectrogram is Time-dependent frequency analysis. In the spectrogram the time axis is the horizontal axis and frequency is the vertical axis. The input audio specgram shows that background disturbance is present in the speech signals and the accuracy rate slightly decreases.

## 4. Cepstrum

We know that human ears, for frequencies lower than 1 kHz, hear tones with a linear scale instead of logarithmic scale for the frequencies higher that 1 kHz. The Mel frequency scale is linear frequency spacing below 1000Hz and a logarithmic spacing above 1000 Hz. The voice signals have most of their energy in the low frequencies. It is also very natural to use a Mel-spaced filter bank showing the above characteristics. If a spectrum contains several sets of sidebands or harmonic series, they can be confusing because of overlap. But in the cepstrum, they will be separated in a way similar to the way the spectrum separates repetitive time patterns in the waveform. The cepstrum of the words 'left' and 'one' is shown. Both charts show a different shape characteristic for that specific word. We discussed that the spectrogram have time dependent problems and the cepstrum is an ideal method for coping with these problems Cepstrum of the words ‘left’ and 'one' represents the cepstrum of two different samples of the word 'left'. It is clear that they almost have the same shape. A cepstral analysis is a popular method for feature extraction in speech recognition applications, and can be accomplished using the Mel Frequency Cepstrum Coefficient analysis (MFCC).

ISSN 2394-187(Online)
ISSN 2394-7179 (Print)

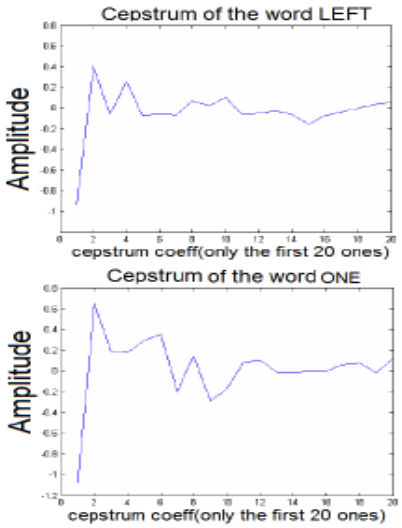



Fig.4: Cepstrum representation of word left and one

## 5. Face recognition

Intelligent systems are being increasingly developed aiming to simulate our perception of various inputs (patterns) such as images, sounds…etc. Biometrics is an example of popular applications for artificial intelligent systems. Face recognition by machines can be invaluable and has various important applications in real life. The development of an intelligent face recognition system requires providing sufficient information and meaningful data during machine learning of a face.

Here, multiple face images of a person with different facial expressions are used, where only eyes, nose and mouth patterns are considered. These essential features from different facial expressions are averaged and then used to train a supervised neural network. Face recognition represents an intuitive and non-intrusive method of recognizing people and this is why it became one of three identification methods used in e-passports and a biometric of choice for many other security applications.

There are many technological issues to be solved as well, some of which have been addressed in recent ANSI and ISO standards. The different types of facial expressions are recorded as shown in Fig.5 and stored in the base and its test for a match occurs exactly on 80% of should be match if it matches it should executes a following action to be performed.

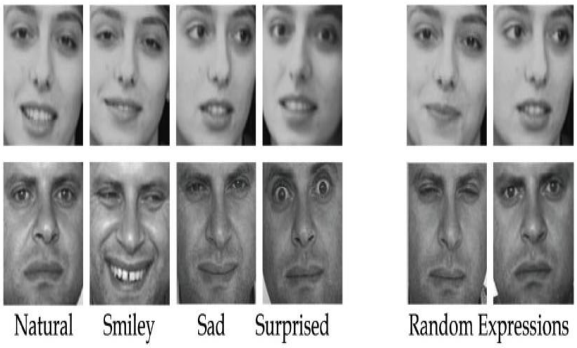



Fig.5: Different types of facial expression

The particular places of the face are noted and it test for match as shown in the Fig.6 this results that the match should be perfect thus the face recognition technique is implemented in voice recognition phase.

Fig.6: Particular nodes of face is recorded

The speech which are given by the user is received by the microphone and then the speech is moved to the database before going to move on database the speech is split into various chunks and checks for an approach which are all fixed on the system of recognizing to next only a new mode of recognizing I fixed here is going to check the face of the person is to be recognized. The particular parts of the face can be recorded as with the help of the sensor lights as shown in the fig.7.

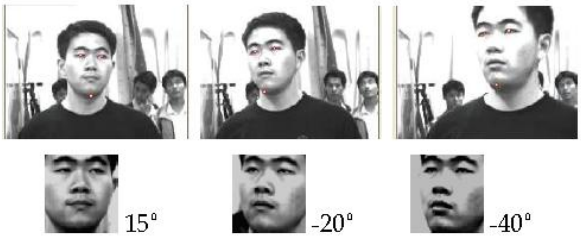
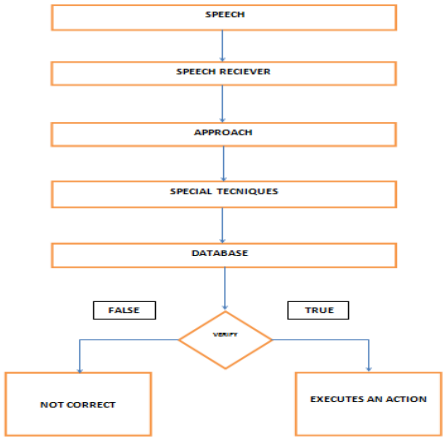



Fig.7: Sensor light sensing parts of a face

The facial expression of the person while expressing a command is stored already in the database in order of the corresponding commands expressing view this should be checked as it is same as recorded one or not. Then only it executes an action to be performed if it is true or if it is false the command gives an action as a message stating wrong command this all actions of modes were explained by the following flow chart.




Flow chart- Working structure of speech recognizer

Fig.8 Flow chart

ISSN 2394-187(Online)
ISSN 2394-7179 (Print)

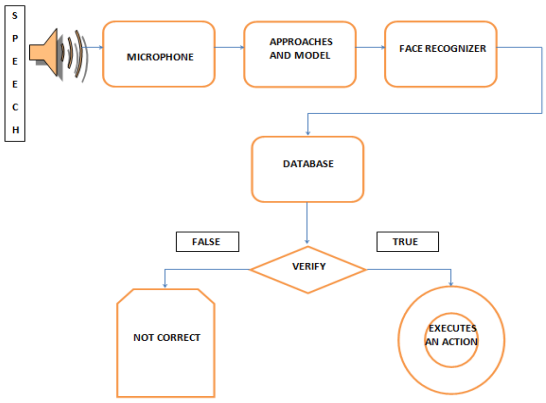



Fig.9Architecture of speech recognizer

The system of speech recognition is acts like the following architecture in which the mode of speech is recognized by audio as well as facial expression also recognized to execute an action the architecture is shown below.

## 6. Conclusion and future work

Disturbance can be reduced by placing Butterworth filter. Spectrogram shows the frequency of the signal it help to measure the exact frequency measure which is already record. The facial expression of the person is watched in order to give the command is acceptable. This paper gives a solution for miss-recognizing voice command. It is not possible to exact match of command so the range I fit here is 80% of should be correct means it provides the result so how it should be satisfied by cent percentage.

## References

[1] Ankush Sharma 1, Srinivas Perala, 2 Priya Darshni 3 Objects Control through Speech Recognition Using LabVIEW Available Online at www.ijecse. International Journal of Electronics and Computer Science Engineering 102Available online at www.ijecse.org ISSN- 2277-1956.

[2] WouterGevaert, GeorgiTsenov, ValeriMladenov, Senior Member, IEEE Neural Networks used for Speech Recognition Journal of Automatic Control, University Of Belgrade, Vol. 20:1-7, 2010©.

[3] SukhdeepKaur, Er. GurwinderKaur Enhancement of Speech Recognition Algorithm Using DCT and Inverse Wave Transformation
Sukhdeep Kaur et al Int. Journal of Engineering Research and Applications www.ijera.com ISSN: 2248-9622, Vol. 3, Issue 6, Nov-Dec 2013, pp.749-754.

[4] Speech Recognition as Emerging Revolutionary Technology Parwinder pal Singh Er. Bhupindersingh.

[5] AnupamChoudhary, Ravi Kshirsagar Process Speech Recognition System using Artificial Intelligence Technique International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-5, November 2012.

[6] Kresimirdelac, Mislavgrgic And Marian Stewart Bartlet.
Recent Advances In Face Recognition, Published By In- The First Published November 2008 Printed In Croatia.

**V.Niranjan** is studying MCA in S.A Engineering College, Chennai. He holds B.sc [computer science] degree from the University of Madras.

**P.Mohankumar** is studying MCA in S.A Engineering College, Chennai. He holds B.sc [computer science] degree from the University of Madras.