International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8, Issue-2S11, September 2019

# MFCC Using Speech Recognition in Computer Applications for Deaf

S. Elavarasi, G. Suseendran

***Abstract---* *Speech Recognition has lately evolved as a beneficial computer technology involving various interactive speech based applications in this domain. For communicating with one another, speech forms the basic foundation. Utilizing this way of communication in the technology domain, the technique of speech recognition is formulated. This technique considers the input speech in order to extract significant information thereby drawing accurate decision related to the concerned text. Here the research proposes a technique by effectively incorporating speech recognition approach for handling web portal via voice (be it from any place or time) for accessing data. Advent and growth in technology has led to formation of STT (speech–to–text) conversion model which generates a text format of the speech that's beneficial for the Deaf individuals and in other realms too. Data mining has achieved great success in examining the acoustic features related to sound and speech. The existing research employs the MFCC (Mel Frequency Cepstrum Coefficients) for extracting acoustic features of a voice that is identifying the gender of a given voice. Following are the stages in the system proposed: 1. speech pre-processing, 2. feature extraction, 3. converting to text4. Classification5. Displaying the web portal and 6. Storing the web portal. The research depicts speech recognition and analysis as a speech process. The STT (speech to text) model includes two levels, first being the document level and second is the sentence level. The proposed system is beneficial for the deaf individuals for accessing the data be it from any place or time and in other realms too. The system yields in great accuracy in the given time span.***

***Keywords---* *Speech Recognition, speech–to–text (STT), Mel Frequency Cepstrum Coefficients (MFCC), Feature Extraction, Preprocessing, Classification.***

## I. INTRODUCTION

Of lately, data mining techniques are being excessively adopted to extract and explore hidden patterns from voluminous dataset thereby resolving various concerns faced by science and technology. AI (Artificial Intelligence) has also gained prime importance. The conventional machine learning approach also employs the similar vital algorithms for identifying patterns from the dataset. For communicating with one another, speech forms the basic foundation. With the presence of virtual book technique the students are enabled to access books/study material from any place and at any point of time. Because of the remarkable growth in the realm of telecommunication, Internet, network computers, use of virtual books and e-Learning has gained a paramount level for undergoing training and learning. The conventional trend of higher education is ruling in various developing countries. The speech recognition model tends to be beneficial for deaf and dumb students in the realm of education. The model imbibes methodologies like speech coding, speech processing, speech synthesis, speech recognition and speaker recognition techniques. From all of the above, speech recognition represents the essential one. The speech recognition system works by transforming the acoustic signal received viaphone/microphone for producing the desired text. Computer technology must be incorporated for extracting and identifying the linguistic information expressed by the speech/sound wave. This technique is employed for various applications for instance household appliances, security device, ATM, mobile phones and computers. Stages in the proposed system includes: pre-processing, feature extraction via MFCC (Mel-frequency Cepstral Coefficients), classification. The method of pre-processing involves identifying and discarding the inconsistent, noisy and incomplete data which being inappropriate to be used directly for data mining and even if used yields undesired output. Feature extraction performs effective analysis viaMFCC (Mel-frequency Cepstral Coefficients). For selecting the most suitable characteristics of the speech two levels are considered, first being the document-level and second is the sentence- level. MFCC is a short-term power spectrum related to sound illustration using speech signals coefficients for feature extraction. The MFCC techniques reviews assist in selecting the filter shape, no: of filters to be employed and many more. Assessment of MFCC implementation is being compared and examined for further comprehension of signals processing techniques. The MFCC coefficients examines the speech tract without relying upon the speech folds that may be hampered because of voice signal. The proposed system targets to convert Speech in form of Text, thereby storing the same in a cloud storage space. Utilizing the above system deaf and dumb can recollect the required information, accessing the text from anywhere/anytime. The web portals are adopted for storing such large volume of data which can be retrieved from numerous web databases with the help of web portal login and user key.

Following is the journal classification. Section 2 illustrates work of previous author. Section 3 discusses the proposed speech recognition along with the views of different levels. Section 4, represents experimental output. And lastly, Section 5 presents the conclusion and recommends research work for future.

---

**Manuscript received September 16, 2019.**

**S. Elavarasi,** Ph.D, Research Scholar, Department of Computer Science, School of Computing Sciences, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, T.N, India. (e-mail: elavarasi_msc@yahoo.co.in)

**Dr.G. Suseendran,** Assistant Professor, Department of Information Technology, School of Computing Sciences, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, T.N, India. (e-mail: suseendar_1234@yahoo.co.in)

## II. RELATED WORKS

Mojtaba Talafi Daryani et.al, presents that the dataset contains 3,168 recorded speech samples obtained from both male and female speakers. Using acoustic analysis, extraction of 20 features with the desired labels was carried out and was made ready for the process of data mining. Eventually, the Python programming language tools along with 6 type of techniques were employed for building a suitable problem solving system. The 6 techniques include: SVM (support vector machines), RF (random forest), logistic regression, regression and classification trees, KNN (K-nearest neighbor) and adaptive boosting [1].

Megha Agrawal et.al proposes the different techniques for building of ASR system, elucidating its benefits and shortcomings. The research compares ASR system's performance depending on feature extraction and speech recognition technique related to a specific language. Lately, demand for speech recognition research on the basis of vast vocabulary speaker-independent-continuous speech has risen. According to the review result, combination of HMM technique and MFCC characteristics proves to be ideal for the above needs, providing effective recognition output [2].

Jayashree Padmanabhan et.al put forth various concerns related to ASR such as recognition in noisy atmosphere, multi-modal and multilingual recognition which really needs attention. A thorough review of standard ML (machine learning)approaches is presented which includes ANN (artificial neural networks), SVM (support vector machines) Gaussian mixture models and hidden Markov models. Also a complete review is illustrated related to the growth and expansion in deep learning which has resulted in improvised ASR performance, as well as future application of ASR [3].

EmreOner Tartan et.al, presents that the image and voice based applications have become mandatory in an individual's mundane life and are utilized for multiple tasks like security, recognition, tracking etc... The prime methods adopted in this domain relies upon ML (machine learning). Simultaneously to this, new processor are built that allows implementing learning algorithm on huge data and matrix oriented operations. At the end of parallel processing deep learning is employed in image and voice applications. Talking about the hardware employed, GPU-processors aids in learning and testing deep learning techniques [4].

Jonathan Chang et.al, presents the technique of representation learning that obtains discriminative illustration of emotional speech in an automated way. Two machine learning approaches are examined for enhancing classifier performance in employing unlabeled data via multi-task learning and DCGAN (deep convolutional generative adversarial network). The experiment controls both a multi-task annotated emotional corpus and a large unlabeled meeting corpus (nearly up to 100 hours). The experiment related to speaker independent classification specifically reveals that the usage of unlabeled data during the analyses raises classifiers performance, also both completely monitored baseline techniques are overtaken remarkably [5].

Tayseer M. F. Taha et.al emphasizes on various approaches present in the study for enhancing the speech signal. Different methods that are adopted include: basic spectral subtraction method, subspace method, wiener filter, statistical methods and spectral subtraction. The methods are illustrated specifying the merits and demerits. There is a review of studies carried out by the researchers concerning ML (machine learning) techniques viz, Deep Neural Network, NN (Neural network), CNN (Convolution Neural Networks) and optimization methodologies utilized for speech enhancement [6].

DeLiang Wang et.al, presents a thorough overview related to the supervised speech separation (deep learning based) approach over the past years. At first, background of speech separation and development of supervised separation is presented.

The review basically emphasizes on separation algorithms which involves reviewing monaural methods as well as speaker separation (multi-talker separation), speech enhancement (speech-non-speech separation) and speech de-reverberation and multi-microphone approaches. There is a discussion on the significant problem of generalization that's distinct to supervised learning [7].

Mohamad A. A. Al- Rababah et.al, performed a study on the behavior of LSTM (Long Short-Term Memory) based NN-neural networks concerning a particular part of automatic speech processing - speech detection. LSTM model is being compared withMLP (Multi-Layer Perceptron) and Elman's RNN (Recurrent Neural Network) neural models.

Experiments are carried on five speech detection process which reveals that the LSTM model is efficient. Whereas RNN model is also of interest to perform automatic speech processing because of its various features. [8].

WEI YU et.al, performs a comprehensive analysis of deep learning concerning its methods and applications. Particularly for the categorical compilation of state-of-the-art related to deep-learning research, a wide reference is offered to the ones looking out for an introduction on deep learning, platforms, algorithms, its uses and implementations in numerous smart-world models [9].

Namrata Dave presents feature extraction methods with their merits and demerits.

LPC(Linear Predictive Codes) is unacceptable due to its linear computation nature.

The observation reveals that MFCC, LPC and PLP being the popularly utilized features extraction approaches concerning the applications of speech recognition and speaker verification.

Whereas Neural Network and HMM techniques being the dominating ones in recognizing patterns concerning speech recognition [10].

Deepak Baby et.al, presents modulation spectrogram features for the exemplar-based tasks making use of this approach.

The system is analyzed with different input exemplars on AURORA-2 and AURORA-4 databases, resulting in better speech enhancement performances.

For directly calculating the full-resolution frequency estimates of speech and noise an effective way is proposed by employing coupled dictionaries which includes an input dictionary possessing atoms from required exemplar space for gaining decomposition and a coupled output dictionary having exemplars from the full-resolution frequency region.

Using this approach, modulation spectrogram features are being presented for the exemplar-based tasks [11].

Ji Ming et.al, put forth a new method that focuses to minimize or eliminate this remove this requirement in an effective manner. For gaining precise estimate from noise without having particular knowledge about the noise, the ZNCC (Zero-mean Normalized Correlation Coefficient) is being employed as the comparison measure and by widening the actual length of speech segment matching to sentence-long speech utterances, accurate speech can be gained. The above approach helps in attending the issues of unpredictable noise or noise with inappropriate training data [12].

Geoffrey S. Meltzneret.al, illustrates the capability of employing silent speech recognition system (sEMG-based) which acts as a substitute communication device for the individuals surviving with laryngectomy. Employing our full 8-sensor set, the proposed system yields a mean WER as 10.3% considering a 2000-word vocabulary, on the other hand the best 4-sensor subset yields a WER of 13.6%. For enhancing sub-vocal word recognition, it’s detected that a moderate rise in the quantity of training data for two subjects was a success, lowering the WERs less than 9% [13].

Bo Wu et.al, targets to build a high dimensional non-linear regression technique concerning the classical speech de-reverberation issue.

The research illustrates via thorough set of trials that ASR performance is enhanced by adopting the proposed joint training strategy. Joint training involving additional discriminative ASR characteristics and better DNN based language models enables to achieve a WER below 4.46% with one system [14].

Michael Wand et.al, proposes EMG (electromyographic) silent speech recognizer which identifies speech by acquiring the electric potentials of the human articulatory muscles, helping the user to quietly communicate. On forming a baseline EMG based continuous speech recognizer, speaking mode variations are analyzed in the research, which resembles inconsistencies amidst audible and silent speech that degrades the recognition accuracy. multimode systems are launched which permits continuous switching amidst audible and silent speech, examine various measures which measure the speaking mode variations, and lays out the spectral mapping algorithm, improvising the WER (word error rate) on silent speech by around 14.3% relative [15].

Mr.Yoghesh Dawande et.al, presents multiple techniques and methods for the purpose of feature extraction in speaker recognition.

The research dealing in speaker recognition ranges from short time features representing spectral attributes of speech (low-level or physical traits) to the high level attributes or behavioral features like phonetic information, prosody, conversational patterns and many more. Techniques of feature extraction - MFCC,PLP and LPC are compared. [16].

Gurpreet Kaur et.al, presents the techniques of feature extraction for speaker-dependent speech recognition related to the isolated words.

Review of various feature extraction techniques is presented such as MFCC (Mel-Frequency Cepstral Coefficients), LPCC (Linear Predictive Coding Coefficients), PLP (Perceptual Linear Prediction), RASTA-PLP (Relative Spectra Perceptual linear Predictive) and their assessment is performed. [17].

Shabnam Ghaffarzadegan et.al, presents a review of reasonable techniques that can help minimize the differences amidst neutral-trained-acoustic models of a speech recognizer and the input whispered speech.

These methods rely upon the algorithm of VTS (Vector Taylor Series)and de-noising auto encoders and assess feature space transformations from neutral to whispered speech for the voiced and unvoiced phone classes. In Vector Taylor Series, the transformations are re-assessed taking in account each input utterance whereas the de-noising auto encoders (DAE) strives for transformations that are global class-specific [18].

## III. PROPOSED WORK

### *3.1 Overview*

Speech Recognition has lately evolved as a beneficial computer technology involving various interactive speech based applications in this domain.

There is no doubt that in presence of ideal atmosphere (i.e. absence of disturbing sound signal) the SR applications perform effectively, but to implement this technique in more actual environment the genuine issue termed as ‘sound source separation’ needs to be attended.

The research presents a technique for achieving better performance of speech recognition technology for handling the web portalvia voice to access data at any point of time and from any place.

The existing research employs the MFCC (Mel Frequency Cepstrum Coefficients) for extracting acoustic features of a voice that is identifying the gender of a given voice. Following are the stages in the system proposed: 1. speech pre-processing, 2.

Feature extraction via MFCC, 3. converting to text4. Classification 5.

Displaying the web portal and 6. Storing the web portal. Our approach reveals that with no evaluation of the humming noise, it remarkably surpass the method which employs optimized noise removal, related to different objective measures considering the speech recognition too.

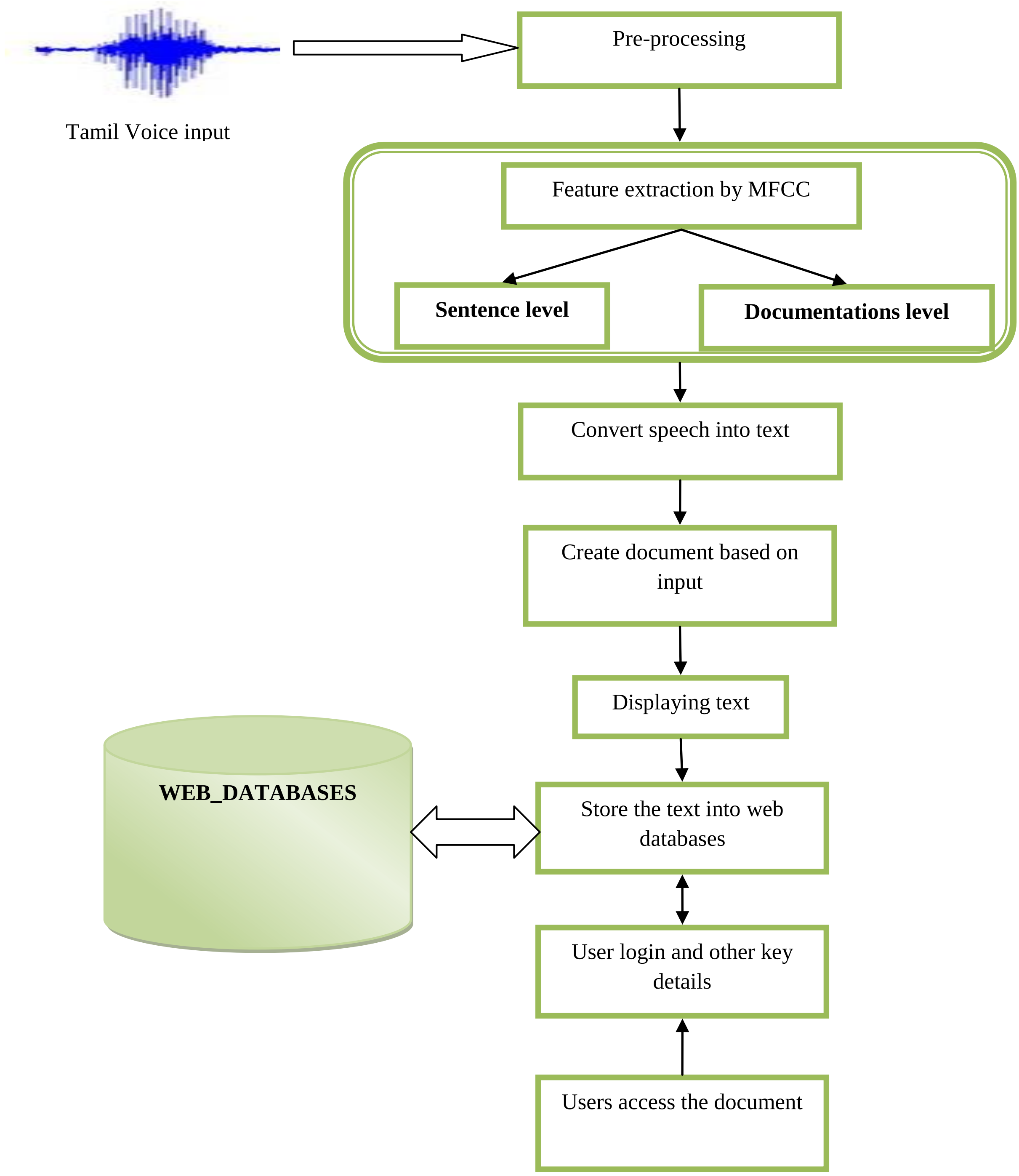


**Fig. 1: Design of Architecture**

*3.2 Signal processing*

Speech recognition tends to be complicated in case of processing the analogue signal (speech signals) which changes randomly. The speech recognition system works by transforming the acoustic signal received via phone/microphone for producing the desired text. STT (speech-to-text) and ASR (automatic speech recognition) are highly evolving techniques in the science and technology domain and assures to be the most upcoming and demanding revolution in human-machine interaction. The SR (Speech recognition)model has the potential to identify thousands of words. There exist innumerous applications of automatic speech recognition in our everyday lives for instance telephone/mobile communication, applications for the physically challenged individuals and uneducated people and much more in the realm of computers.

*3.3 Collection of Input voice data*

Speech technologies assist by offering normal interfaces with which the digital data is made available to the public and simplifies information exchange between various individuals, say speaking Tamil language. The STT (Speech to Text) approach accept live input (audio) from a microphone which is thereafter transformed into text format and displayed over the monitor/desktop.

In the work, Tamil voice is considered as an input data which is being converted in a text format. The system works by accepting user voice/speech as input which then undergoes pre-processing to eliminate any silent segments in the voice stream. The data trains and educates systems for interacting with humans.

### *3.4 Voice data pre- processing*

Normally the data pre-processing is ignored which being an essential process for implementing the techniques of data mining. Before extracting the signal's features, it must undergo preprocessing. Applications which doesn't favor any noise disturbance or silence, the stage of pre-processing holds very significant. The Speech signal confronts disturbances of electromagnetic interference from nearby environment, interference of power frequency and interference of some humming sound within the individual and voice sounds. Such noise interferences hampers the analysis of speech recognition data, making it quiet intolerable.

### *3.5 Feature Extraction of speech signal*

Feature Extraction is a significant process of speech recognition which involves separating one speech from the other since there are various personalized features associated with every speech utterance. These features are retrieved using varied proposed feature extraction techniques thereby exploiting them effectively for the process of speech recognition. The feature extraction is carried out in 3 levels: the first being the speech analysis (acoustic front end). It conducts a spectra temporal analysis of the signal, producing raw characteristics features depicting the envelope of the power spectrum related to short speech intervals. At the second level an extended feature vector is being compiled that includes static and dynamic features. The third and the final stage (which is optional) converts the above extended feature vectors into highly compact and robust vectors which are forwarded as input to the recognizer.

### *3.6 Mel Frequency Cepstrum Coefficients (MFCC)*

Recognizing the signs created by deaf people for generating similar textual data used for communication by the normal individuals is indeed necessary and a tough issue for speech recognition. According to the Psychophysical survey it's elucidated that an individual's observationconcerning the sound frequency contents for speech signals is non-linear. Therefore for every tone having actual frequency-f, computed in Hz, subjective pitch is calculated on a scale referred to as the 'Mel' scale. MFCC coefficients involves collection of DCT de-correlated factors, calculated via transformation of the logarithmically condensed filter output energies, obtained via perceptually spaced triangular-filter-bank which carries out the processing of DFT (Discrete Fourier Transformed) speech signal. The process minimizes the no: of feature vectors by equivalent no: of filters in contrast to the traditional MFCC.The speech sample is considered as input for extracting the coefficients. By employing MFCC the voice/speech is converted to a text format.

#### *3.6.1 MFCC Structure*

Fetching the most apt parametric depiction of acoustic signals is essential for generating effective recognition performance. Efficiency of present stage is vital for the subsequent stage as it can impact its behavior.

**Step 1: Pre–emphasis–**here traversing of signal via filter is processed which underscores high frequencies. Here the energy of signal is increased withhigh frequency.

$$Y[n] = X[n]-0.95\,X[n-1]$$

**Step 2: Framing–**It involves partitioning the speech samples retrieved from ADC (analog to digital conversion) into small frame having length between 20 to 40 msec. The retrieved voice signal is segmented into frames with N samples. Frames that lie adjacent are divided by M (where M<N). Standard values utilized are M = 100 and N= 256.

**Step 3: Hamming windowing -** Hamming window is employed taking into account the subsequent block in the processing chain of feature extraction which combines all the nearest frequency lines. Below is the Hamming window equation:

Consider the window stated as W (n), $0 \leq n \leq N-1$ where

N = number of samples in each frame

Y[n] = Output signal

X (n) = input signal

W (n) = Hamming window, output of the windowing signal is represented as:

$$Y(n)=X(n)\times W(n)$$

**Step 4: Fast Fourier Transform–**Every single frame of N samples from time domain is converted into frequency domain. The Fourier Transform represents converting the convolution of the glottal pulse-U[n] and the vocal tract impulse response-H[n] into time domain. The below equation supports this statement:

$$Y(w) = FFT[h(t)*X(t)] = H(w)*X(w)$$

**Step 5: Mel Filter Bank Processing–**There exist large frequencies range in FFT spectrum also the voice signal is non-linear. The filter's magnitude frequency response has a triangular shape and equivalent to unity at the center frequency and reduce linearly to 0 at center frequency of the two concerned adjacent filters. Every single filter result denotes the sum of its filtered spectral elements. Thereafter following equation is utilized for measuring the Mel for existing frequency (f )in terms of HZ:

$$F(Mel) = [2595 * \log 10\,(1+f)]700]$$

**PSEUDO CODE FOR PROPOSED MFCC FEATURE EXTRACTION**

***Input:*** *$i^{th}$ frame, $x_{p,i}(n)$*

***Output:*** *MFCC cepstral co-efficients for $i^{th}$ frame*

*for i=1, 2… Numbers of frames do*

*$X_{p,i}(f) = DFT\{x_{p,i}(n)\}$*

*for j=1, 2…., $N_F$ do*

*$N_F$- number of sub band filters used in the Mel filter bank*

*$Y_{i,j}(f)=A_j(f)X_{p,i}(f)$ {output of $j^{th}$ filter of Mel filter bank A}*

$Y_{i,j}(n)=IDFT\{Y_{i,j}(f)\}$ *{sentence level output}*
*Where*
$Y_{i,j}(n)=$ *Mel sub band filtered signal*
$Z_{i,j}(n)=\{y_{i,j}(n),DI\}$
*Where*
$Z_{i,j}(n)$- *Mel filter bank*
*DI-Dependency index*
$E_{i,j}=Mean\{Z_{i,j}(n)\}$
$S_{i,j}=log\{E_{i,j}\}$ *{document level output}*
*end for*
*end for*

### *3.6.1 Sentence level*

Sentence level classifies an individual's speech with minimum or least keywords. Sentence-level static modeling forms the baseline system. The overall sentence's features are equivalent to the ones utilized for segments. The data is divided in a manner so that every single fold has equivalent no: of speech and sentences categories. It's required to adopt the segment speech classifier's predictions for training some other decision model, another cross validation is required for the purpose of decision model training.

### *3.6.2 Documentation level*

Documentation level classifies an individual's speech with maximum keywords. Keyword basically express the overall document view. The entire text is considered as a fundamental data unit. Routing of speech takes place via speech-recognition machine and the accepted draft document is routed alongside the original voice file wherein the draft is modified and the report being finalized.

### *3.7 Convert speech into text*

This technique assists the systems to give a precise and reliable response to people's live voices, thus offering beneficial and useful services. Since interaction via computer system through live voice is prompt in contrast to the microphone, individuals opt for such a quick system. Considering the Tamil language ruling among the individuals communication, it's obvious that the public demands voice interfaces with computer. To rectify this concern, the speech recognition system has been built involving STT (speech-to-text) that permits the convert voice/audio request and dictation in a text format. Speech recognition in relation to STT involves the act of converting an acoustic signal (obtained via microphone)to a text format. The data retrieved can be further utilized for the purpose of documentation.

### *3.8 Displaying Text and Store to web portal*

The speech signal is fetched using the MFCC display thereafter stored on the web portal. The web-portal-speech application assists the researchers for transforming speech to text by incorporating speech recognition systems with the easily utilized application. The speech uploaded in the request is identified, combining it with speech storage on web portal in Cloud Storage. The proposed system of speech recognition system is speaker-independent and is founded on speech recognition along with MFCC. Generally, usage of voice/speech lies in routine tasks, one among them is performing search over WWW using voice. Cloud-computing is adopted by Google for performing speech recognition tasks.

## IV. RESULT AND DISCUSION

The research attempts to explore and analyze the speech recognition and making it execute by incorporating the feature extraction technique of MFCC (Mel Frequency Cepstrum Coefficients).By making use of such techniques speech signal set is examined and most appropriate factors are identified for the purpose of speech recognition. Using MFCC(Mel Frequency Cepstrum Coefficients) irrelevant attributes are being extracted thus generating an equation for the speech recognition. Resultant increased performance and accuracy is yielded by employing MFCC for the process of speech recognition. The benefit of the proposed technique lies in making accurate decisions, producing desired results and attending complicated issue with help of JAVA.

Table 1 depicts the accuracy of the proposed technique of feature extraction in contrast to the prevailing techniques such as LPCC (Linear Predictive Coding Co-efficient), PLP(Perceptual Linear Prediction) and LPC(Linear Predictive Coding). It's elucidated from the results that the proposed technique of MFCC(Mel Frequency Cepstrum Coefficients) yields in high accuracy and improvised performance when compared to rest of the techniques.

**Table 1: Comparison of Feature extraction techniques**

| S.No | No. of Techniques | Accuracy (%) |
|---|---|---|
| 1 | Linear Predictive Coding Co-efficient (LPCC) | 78.5 |
| 2 | Perceptual Linear Prediction (PLP) | 89.7 |
| 3 | Linear Predictive Coding (LPC) | 65.8 |
| 4 | Mel Frequency Cepstrum Coefficients (MFCC) | 93.2 |

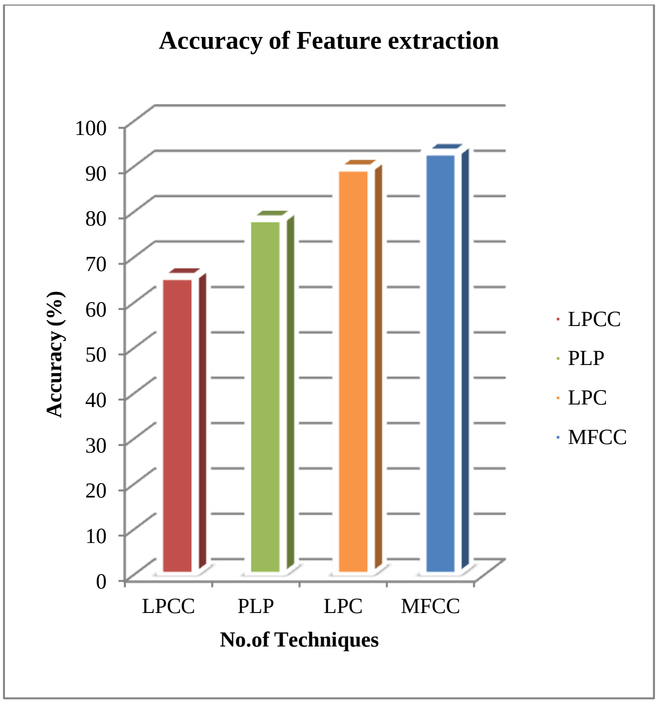


**Fig. 2: Accuracy of Feature extraction techniques**

Fig.2mentioned above compares the proposed technique of feature extraction (MFCC)in contrast to the prevailing techniques such as LPCC (Linear Predictive Coding Co-efficient), PLP (Perceptual Linear Prediction) and LPC (Linear Predictive Coding). Technique of MFCC(Mel Frequency Cepstrum Coefficients) yields in high accuracy and improvised performance when compared to rest of the techniques.

Table 2 illustrates the proposed MFCC (Mel Frequency Cepstrum Coefficients)feature extraction technique on the basis of 2 levels that is, sentence level accuracy and other being documentation level accuracy.

**Table 2: MFCC level**

| S . N o | Mel Frequency Cepstrum Coefficients (MFCC) | Accuracy (%) |
|---|---|---|
| 1 | Sentence level | 90.5 |
| 2 | Documentation level | 87.3 |

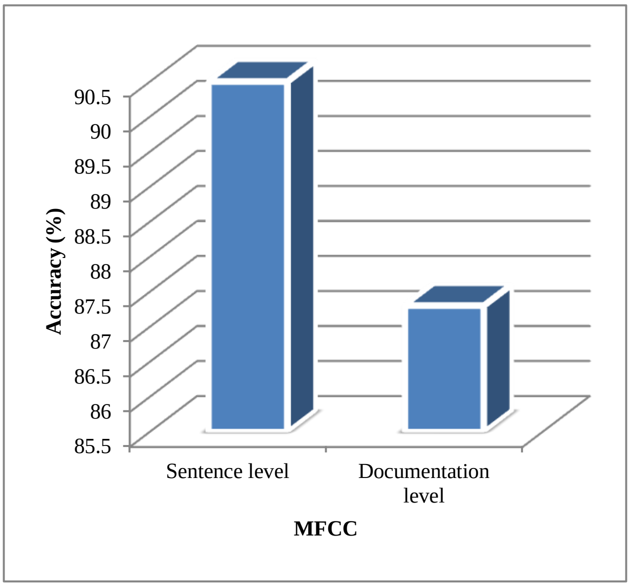


**Fig. 3: MFCC level**

Similarly, Figure 3 illustrates the proposed MFCC (Mel Frequency Cepstrum Coefficients)feature extraction technique on the basis of 2 levels that is, sentence level accuracy and the other being documentation level accuracy

## V. CONCLUSION

The research investigates and examines issues concerning speech recognition by incorporating MFCC (Mel Frequency Cepstrum Coefficients). The work and analysis founded that the approach of MFCC (Mel Frequency Cepstrum Coefficients) yields high accuracy and efficiency in contrast to other prevailing techniques. Tamil language database is required to identify emotion pertaining to a particular language itself. It's not simple to identify speech by the machines unless intense intelligence and training is involved. Building such a controllable intelligence system can simplify the human-machine communication than previously. It has been observed that few speeches can be precisely identified whereas there are speech with undetermined condition which makes reasoning and classification difficult. The area of Speech Recognition is quiet challenging and complicated and the more research it undergoes the more success and growth it will bring forth. With new innovations and discoveries in this domain will produce innumerable benefits thus enhancing theoverall living style and quality of individuals and society at large.

## REFERENCES

1. Mojtaba Talafi Daryani, Hossein Khabiri and Zahra Yamini "Application of Data Mining Techniques in the Analysis of Acoustic Sound Characteristics", Journal of Information Technology & Software Engineering, Vol. 8( 3), pp.1- 6.
2. Megha Agrawal, Tina Raikwar "Speech Recognition Using Signal Processing Techniques", International Journal of Engineering and Innovative Technology, Vol. 5( 8), 2016, pp. 65-68.
3. Jayashree Padmanabhan, Melvin Jose Johnson Premkumar "Machine Learning in Automatic Speech Recognition: A Survey", Vol. 32(4), 2015, pp. 240-251.
4. Emre Oner Tartan, Ali Berkol, Ali Yücelen "A SURVEY OF SPEECH SEPERATION BY DEEP LEARNING", International Journal of Scientific Research in Information Systems and Engineering, Vol. 4, 2018, pp.92 - 98.
5. Jonathan Chang, Stefan Scherer "Arning Representations Of Emotional Speech With Deep Convolutional Generative Adversarial Networks", © IEEE, 2017, pp. 2746 - 2750.
6. Tayseer M. F. Taha, Ahsan Adeel, Amir Hussain "A Survey on Techniques for Enhancing Speech", International Journal of Computer Applications, Vol. 179,(17), 2018, pp.1-14.
7. De Liang Wang, Jitong Chen "Supervised Speech Separation Based on Deep Learning: An Overview", © IEEE, ACM TRANSACTIONS ON AUDIO, 2018, pp.1-28.
8. Mohamad A. A. Al- Rababah, Abdusamad Al-Marghilani, Akram Aref Hamarshi "Automatic Detection Technique for Speech Recognition based on Neural Networks Inter-Disciplinary", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9(3) 2018, pp.179-184.
9. WILLIAM GRANT HATCHER and WEI YU "A Survey of Deep Learning: Platforms, Applications and Emerging Research Trends", © IEEE ACCESS, SPECIAL SECTION ON HUMAN-CENTERED SMART SYSTEMS AND TECHNOLOGIES, Vol. 6, 2018, pp. 24411-24432.
10. Namrata Dave "Feature Extraction Methods LPC, PLP and MFCC in Speech Recognition", INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY, Vol. 1(6), 2013, pp. 1-5.
11. Deepak Baby, Tuomas Virtanen, Jort F. Gemmeke, Hugo Van hamme "Coupled Dictionaries for Exemplar-Based Speech Enhancement and Automatic Speech Recognition", IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, Vol. 23( 11), 2015, pp. 1788 – 1799.
12. Ji Ming, Danny Crookes "Speech Enhancement Based on Full-Sentence Correlation and Clean Speech Recognition", © IEEE/ACM Transactions on Audio,

Speech and Language Processing, 2016, pp. 1- 13.

13. Geoffrey S. Meltzner, James T. Heaton, Yunbin Deng, Gianluca De Luca, Serge H. Roy, and Joshua C. Kline "Silent Speech Recognition as an Alternative Communication Device for Persons With Laryngectomy", IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, Vol. 25,(12), 2017, pp. 2386 - 2398.
14. Bo Wu, Kehuang Li, Fengpei Ge, Zhen Huang, Minglei Yang, Sabato Marco Siniscalchi, and Chin-Hui Lee "An End-to-End Deep Learning Approach to Simultaneous Speech Dereverberation and Acoustic Modeling for Robust Speech Recognition", IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, 2017, pp.1-11.
15. Michael Wand, Matthias Janke, and Tanja Schultz "Tackling Speaking Mode Varieties in EMG-Based Speech Recognition", © IEEE, TRANSACTIONS ON BIOMEDICAL ENGINEERING, Vol. 61(10), 2014, pp. 2515 - 2526.
16. Mr.Yoghesh Dawande, Dr. Mukta Dhopeshwarkar " ANALYSIS OF DIFFERENT FEATURE EXTRACTION TECHNIQUES FOR SPEAKER RECOGNITION SYSTEM: A REVIEW", International Journal of Advanced Technology & Engineering Research, Vol. 5(1), 2015, pp. 5-7.
17. Gurpreet Kaur, Mohit Srivastava, Amod Kumar "Analysis of Feature Extraction Methods for Speaker Dependent Speech Recognition", International Journal of Engineering and Technology Innovation,Vol. 7(2), 2017, pp. 78 – 88.
18. Shabnam Ghaffarzadegan, Hynek Boˇril, John H. L. Hansen "Generative Modeling of Pseudo-Whisper for Robust Whispered Speech Recognition", © IEEE/ACM Transactions on Audio, Speech, and Language Processing, pp. 1-17.
19. Elavarasi .S,Dr.G.Suseendran, "Implementing Speech Recognitionon Handling Device: A Review", Journal of Advanced Research in Dynamical and control systems, Vol.9, (6),2017, pp.97-103.