IJCTA, 9(21), 2016, pp. 97-104

Telugu Speech Features Extraction by MODGDF and MFCC using Naive Bayes Classifier

Archek Praveen Kumar* Ratnadeep Roy Sanyog Rawat*** Achyut Sharma****Amit Chaurasia*******

Abstract : Speech is the most popular communication medium used in real time. Speech is a one dimensional signal which is easily corrupted by the noise. De-noising is the only technique to remove the noise which is done in preprocessing techniques. Discrete Wavelet Transform (DWT) is the efficient technique used for de-noising. Speech is recorded first and then preprocessed Spectrum is being calculated. The preprocessed speech is then jointly processed by MODGDF and MFCC technique for features extraction and later classified by Naive bias classifier technique for acquiring greater recognition accuracy. The proposed methods are implemented for 100 speakers uttering 10 words. The integrated performance of Mel-frequency cepstral coefficients (MFCC), Modified group delay function (MODGDF) feature extraction technique and Naïve Bayes classification technique (NBC) is based on recognition accuracy. Parameters like (LSP) Line spectrum, (PPF) Pitch prediction filter, (CBI) Code base indexes, Gain, Synchronization, (FEC) Forward error correction are extracted. The explored results prove the ability of these techniques for recognizing the Telugu speech.

Keywords : Speech recognition, Naïve Bayes, MODGDF, endpoint detection, Telugu.

1. INTRODUCTION

Data is a huge word which is represented in signals; every platform has its own data, which is processed for storing or transmission. Medium plays a major role in transmission or processing. There are different types of mediums used in the communication world. Medium of communication can be done in two ways natural communication and technical communication. Natural communication is done by visually, speech, images, video etc. technical communications are by radio, TV, press, cinema, books etc. communication can be done between human to human, human to machine and machine to machine [1]. All this communications need some processors to process the data for better results. Digital or analog processors play a major role in ground communication, media communication and sky communication. The communication medium includes input with touch or gesture and output with vision and hearing for humans but machine includes input with keyboard, mouse or sensors etc. and output with display and sound. The speech is the important way of communication in multiple mediums, but speech is an ocean there are different languages with different codes and slangs. Speech processing is very old area to

* Department of Electronics and Communication Engineering Amity University Rajasthan, Jaipur, Rajasthan, India archekpraveen@gmail.com,

** Department of Physics Amity University Rajasthan, Jaipur, Rajasthan, India rroy@jpr.amity.edu,

*** Department of Electronics and Communication Engineering Manipal University, Jaipur, Rajasthan, India sanyog.rawat @jaipur.manipal.edu,

**** Department of Electronics and Communication Engineering Amity University Rajasthan, Jaipur, Rajasthan, India asharma@jpr.amity.edu,

***** Department of Electronics and Communication Engineering Amity University Rajasthan, Jaipur, Rajasthan, India Email-achaurasia2583@gmail.com

research but there are new trends which are providing intense results in communication. Speech processing is part of speech recognition where this paper deals with recognition of speech by MFCC feature extraction and Naïve Bayes classification technique for south Indian Telugu language. There are many feature extraction techniques and feature classification techniques but the proper suitable technique should be used for getting higher accuracy [2].

Contemporary speech recognition technique is based on speech feature extraction and speech feature classification. Features based on functional principles like LPCC, MFCC, PLP etc. are used. Features simulation is the informative task for speech researchers. This paper proposes new features obtained from modified group delay functions and MFCC. MFCC and MODGDF are combined to produce better features. These techniques are isolated for broadcasting of news in Telugu and Tamil language. Speech generally fails to align the phonetic unit boundaries perfectly. This paper uses novel technique for better recognition efficiency which uses minimum group delay functions expressed from root spectrum. Telugu language data base is created using a base line system. The proper procedure for recognition is presented. Speech is differentiated to many factors like sex, poignant state, intonation, diction, expression, adenoidal, pitch, sound, rapidity [3].

2. GENERAL BLOCK DIAGRAM

Speech is an analog signal which is recorded first and then it is segmented to number of frames. This paper deals with two frame sizes primarily 144 bits/sec and secondly 50 bits/sec. Each frame is processed by converting continues data to discrete samples. Later each frame is preprocessed by techniques like end point detection, pre emphasis, frame blocking, frame windowing, distortion measures, time alignment, amplitude modulation, short time energy, short time zero crossing count. These preprocessing techniques enhance the speech and makes easy to extract the features. After preprocessing features are extracted using the integrated MODGF and MFCC where LSP, PPF, CBI and Gain are extracted finally the extracted features are classified by Naïve Bayes technique for recognition of speech with good accuracy. The detailed procedure is shown in figure 1.

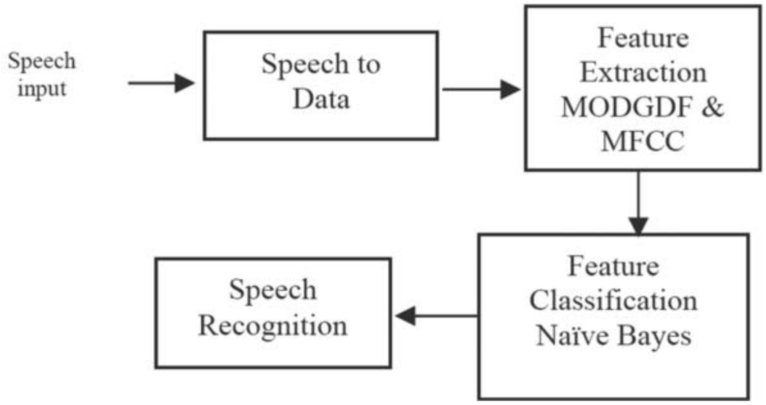

Speech input
Speech to Data
Feature Extraction MODGDF & MFCC
Feature Classification Naïve Bayes
Speech Recognition

Fig. 1. General Block diagram

2.1. Telugu Speech Input

The Nanda kings and Mauryan emperors at Pataliputra invented the Brahmi script and finally the modern day Devanagari. Dravidian script is said that the style of writing is reestablished after the death of Buddha. Telugu is the property of the Dravidian family languages. Telugu has flourish script. India was spread by a broad amount of ideological and traditional literature. Aryans invented a phonetics and convoluted rules of syntax and utterance. Telugu and Kannada are introduced through braahmee Andhra (Saatavaahana) dynasty. Thus, Telugu language is designed from the proto-Dravidian language and the script is generated from the Braahmee. Now the Telugu language is spoken in two states Andhra and Telengana. Telugu language consists of 60 symbols where 16 are vowels, 3 vowel modifiers and 41 consonants. International Phonetic Alphabet (IPA) is used for processing the Telugu language [4].

2.2. Speech to Data

Speech is a physical quantity. This physical quantity is converted to electrical signal which is analog in nature. Now these analog signals are segmented to frames for preprocessing. Each frame is sized with 144 and 80 bits per second. Firstly speech is recorded in mp3 format which is later transformed to .wave files. The wave signal is a frame but it is continuous. Now this continuous signal is discretized by sampling at 16 KHz called data sequence. The data sequence is preprocessed followed by feature extraction and classification [5].

2.3. End point Detection

End point detection is not an easy task as speech is easily affected by noise, but its major part for robust speech recognition. A feature specification like short-time energy is normally used to illustrate the speech segments from other waveforms. The speech is corrupted due to noisy environments, slang, pitch etc. frames are indicated by fixing threshold level but there is a chance of false alarm and missed detection. Suitable threshold levels are perfectly fixed foe perfect frames.

2.4. Preprocessing

Preprocessing includes, pre emphasis, frame blocking, frame windowing, distortion measures, time alignment, amplitude modulation, short time energy, short time zero crossing count which is a part in the speech recognition.

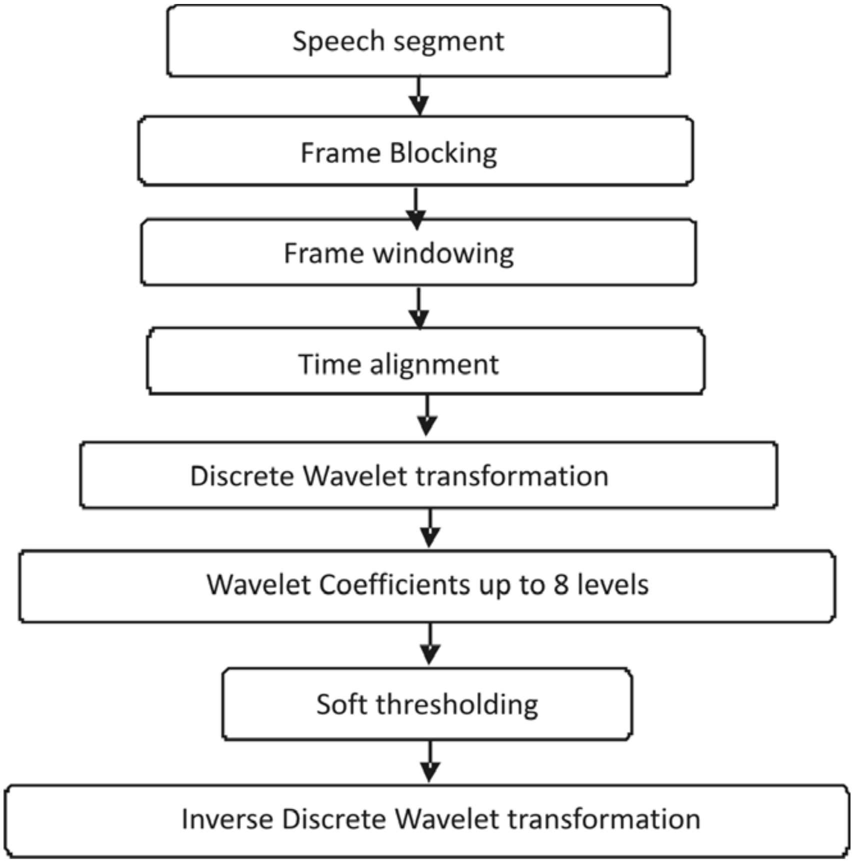

Speech segment
Frame Blocking
Frame windowing
Time alignment
Discrete Wavelet transformation
Wavelet Coefficients up to 8 levels
Soft thresholding
Inverse Discrete Wavelet transformation

Fig. 2. Proposed algorithm for preprocessing

Pre-emphasis recoup the high-frequency that will generally squash during the recording. The frame size should be equal to power of two for smooth use of any transformation .if needed zero padding is done to the closest length of power of two. Hamming window is multiplied to every frame which maintains the ongoing of the first and the last points in the frame. Amplification is done later of high-frequency formants discrete wavelet algorithm is used for de-noising the speech signal. Which use soft thresholding technique where the entire process is shown in figure 2 [6]. Universal threshold given by AWGN given in equation (1)

$$X_{soft} = \begin{cases} \text{sign}(X)\ (|X| - |\tau|) & \text{if } |x| > \tau \\ 0 & \text{if } |x| \leq \tau \end{cases} \tag{1}$$

Soft threshold where X represents wavelet coefficients, *t* is threshold value is given by equation (2)

$$\tau = \sigma(2\log(N))1/2 \tag{2}$$

Sigma is standard deviation, N is length of signal [7].

2.5. Feature extraction

Feature extraction is major important part in the speech recognition. Feature extraction is the simulated technique where feature vectors are estimated. LPCC, PLP, and RASTA-PLP are some of the techniques, but MFCC Mel frequency cepstral coefficients and MODGDF are suitable technique used. The isolation of these both techniques gives better results in extracting the features. The total process of MFCC is shown in figure 3[8].

(*i*) **MFCC :** MFCC is mel frequency cepstral coefficients which is a feature extraction technique naturally used in speech recognition

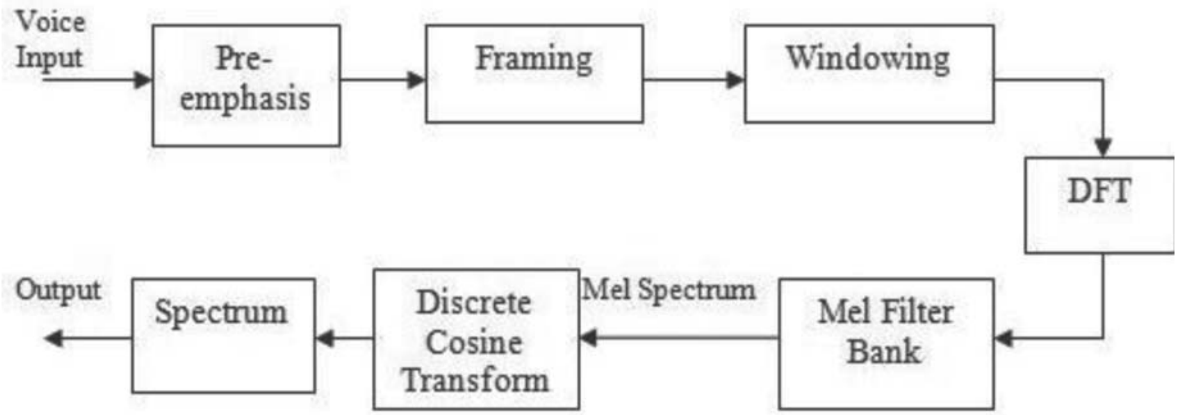


Fig. 3. MFCC Block Diagram

The delta-MFCC coefficients are determined in equation (3), where *dk* is the delta coefficient determined as $ck + \alpha$ to $ck - \alpha$. M is the configuration parameter. [9] [10].

$$d_k = \frac{\sum_{\alpha=1}^{M} \alpha\,(C_{k+\alpha} - C_{k-\alpha})}{2\sum_{\alpha=1}^{M} \alpha^2} \tag{3}$$

(*ii*) **MODGDF :** Speech has both magnitude and phase. Feature extraction by the Group delay function (GDF), defined as the negative derivative of phase, when the signal is of minimum phase defined in equation (4)

$$\tau(f) = -d(\theta(f))df \tag{4}$$

where $\theta(f)$ is the unwrapped phase function. The GDF derived in equation (5)

$$\tau z(f) = xR(f)yR(f) + yI\ (f)xI\ (f)\ |X(f)|2 \tag{5}$$

The subscripts R and I denote the real and imaginary parts of the Fourier transform. X(f) and Y (f) are the Fourier transforms of *x*(*n*) and *n x*(*n*). The pointed character of the Group delay spectrum (GDS) is controlled by substation of |X(*f*)| in the denominator of the GDF with its cepstral data, S(*f*) [11]. Features of MFCC and MODGDF are integrated which improves the efficiency of speech recognition. For frame size 144 bits 72 dimensional MODGDF stream and 72 dimensional MFCC stream is combined by feature stream combination. Similarly for 80 bits 40 dimensional for both MODGDF and MFCC are considered.

2.6. Feature Classification

The Features vectors extracted are classified by feature classification technique [12]. Classification can be done by pattern recognition, statistical, artificial neural networks, Naïve Bayes, WPD etc. classification is done by considering two phases training phase and testing phase. This paper used 70 percent for training and 30 percent for testing. Classifier gives a model from training data which estimates the target point of testing the unknown patterns. Suitable classifiers are used according to the language, pitch, and other parameters. Naïve Bayes classifier is based on Bayes theory. This is a multiclass classifier with simple estimation. Conditional features are calculated with combination of independent features with conditional features. This classifier requires only small training sets. The classifier is defined in equation (6)

$$P(A/B) = P(B/A)*P(A)/P(B) \tag{6}$$

P (A) is probability of A and P (B) is probability of B. P (A/B) is conditional probability of A given B and P (B/A) is conditional probability of B given A

3. PROPOSED ALGORITHM

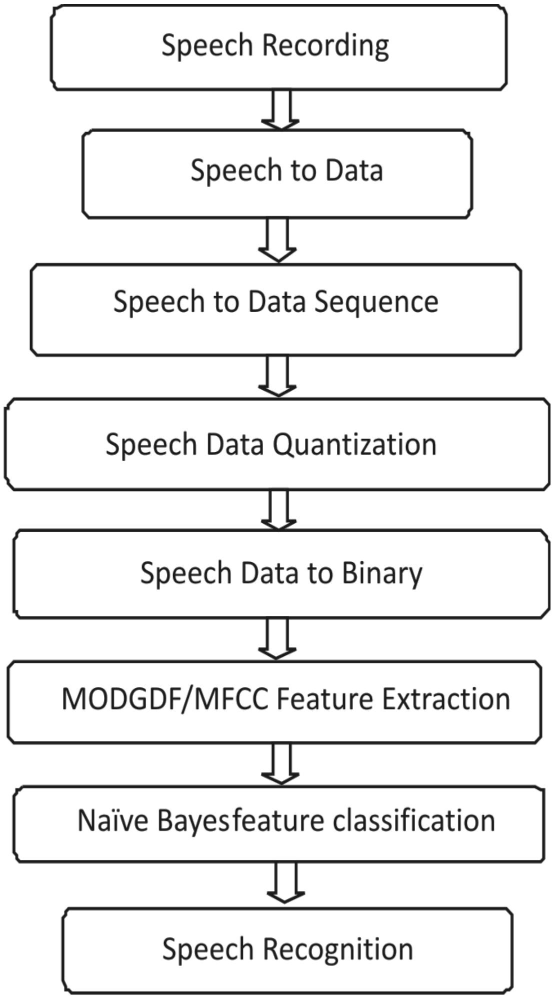
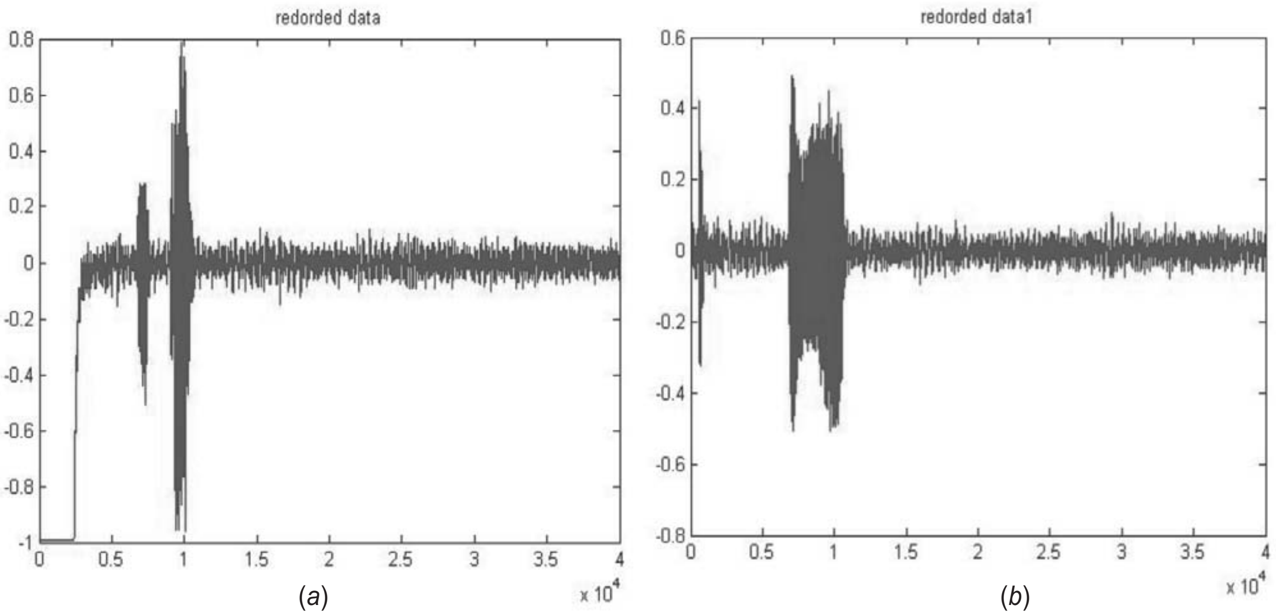


Fig. 4. Proposed algorithm

4. EXPERIMENT AND RESULT

Telugu words AKKA, ANNA, ATTHA, VACHA, RA, IVU, PUVU, RATRI, NEEKU, ILA are recognized. Telugu data base is created for 100 male voices and 100 female voices with age factor of 20 to 40 and 10 words are spoken. Results plots are shown in figure 5 for only words AKKA and ANNA



(*a*) (*b*)

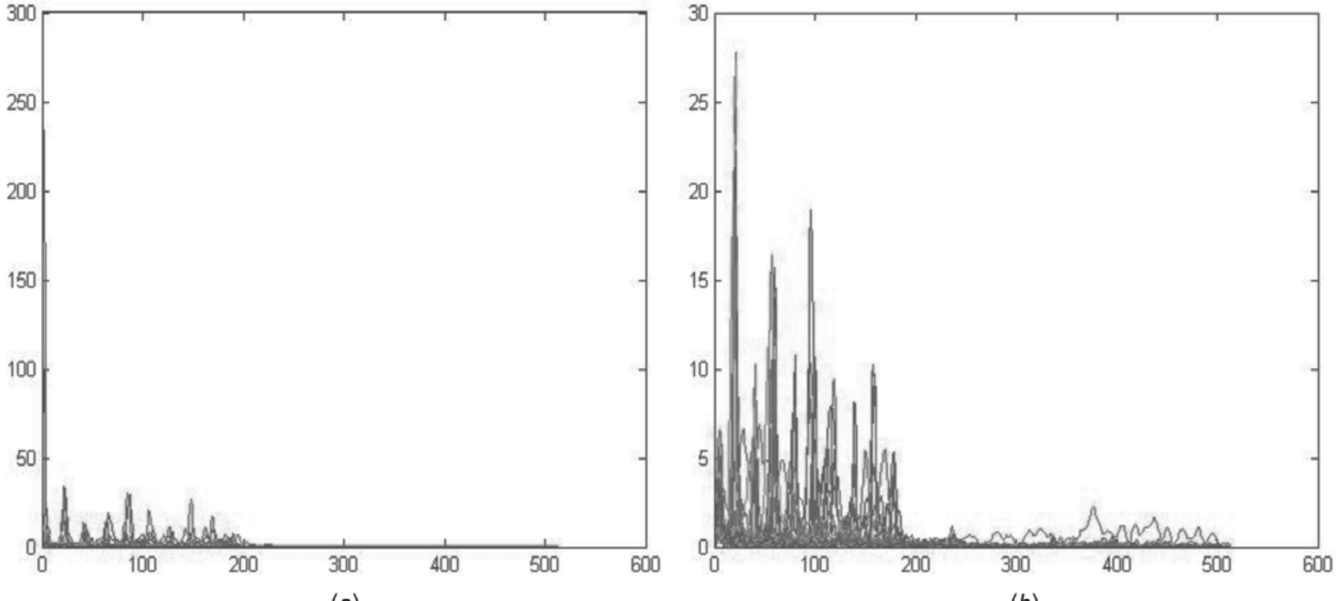
Fig. 5. (*a*) Recorded speech for word AKKA (*b*) Recorded speech for word ANNA

AKKA word is recorded and data Sequence has been calculated.

The sampling rate is 16 kHz and the frame size is 144 sample points and 80 sample points

- The frame duration is 144/16000 = 0.009 sec = 9 ms, overlap is 60 points, frame rate is 16000/(144-60) = ~190 frames per second
- The frame duration is 80/16000 = 0.005 sec = 5ms, overlap is 60 points, frame rate is 16000/(80-60) = ~800 frames per second

If the signal in a frame is denoted by *s*(*n*), *n* = 0,…N – 1, and *w*(*n*) is the Hamming, after frame windowing it is *s*(*n*)**w*(*n*) [13]. Spectrum after windowing is shown in figure 6.



(*a*) (*b*)

Fig. 6. (*a*) Recorded spectrum for word AKKA (*b*) Recorded spectrum for word ANNA

Joint Feature extraction techniques of MODGDF and MFCC were used for Recognition of speech with different frame rate are shown in figure 7 for the words AKKA and NANNA

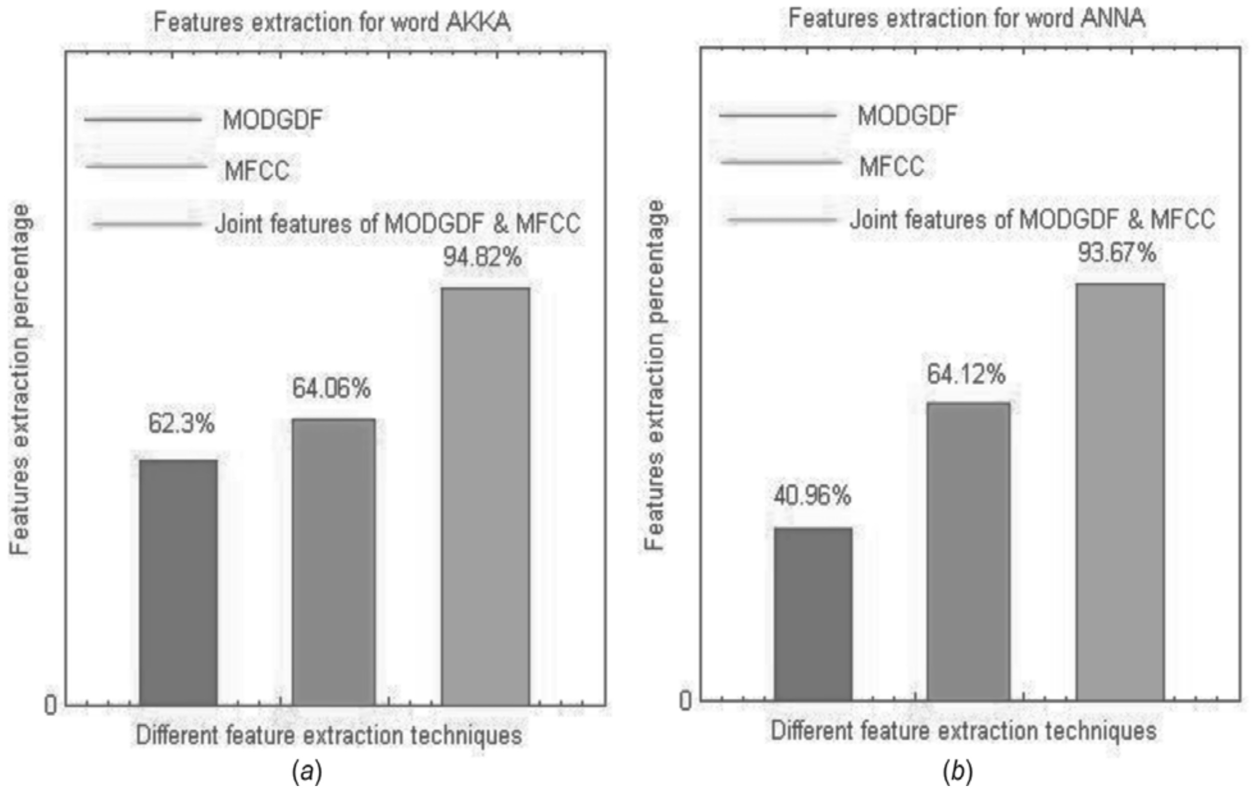


Fig. 7. (*a*) Joint feature extraction for word AKKA (*b*) Joint feature extraction for word ANNA

Feature extraction using MODGDF and MFCC are done. Various parameters or features extracted for 144 bits/ frames and 80 bits/ frame shown in TABLE I.

Table 1. Features extracted

Features	*Bits of data*	*Bits of data*
Line spectrum pairs (LSP)	30	17
Pitch prediction filter (PPF)	45	13
Code base indexes	31	33
Gain	31	17
Synchronization	2	–
FEC	5	–
Total	144	80

The recognition accuracy obtained by using MFCC and MODGDF feature extraction techniques with Naïve Bayes classification is 93.76%.

4. CONCLUSION

Speech recognition is done by feature extraction with the help of MODGDF and MFCC. Then classification is done on extracted features by Naïve Bayes classifier. The technique MFCC and MODGDF provides various parameters like LSP, Pitch prediction filter, code base indexes, gain, synchronization, FEC. Two frame sizes are taken one is 144 bits/frames and other is 80 bits/frames. Recognition results are good. Further work can be done by changing the techniques for getting greater accuracy.

5. REFERENCES

1. P. Ramesh babu–Digital Signal Processing; Fourth edition; SciTech Publications, 2003.
2. A. P. Kumar, N. Kumar, C. S. Kumar, A. K. Yadav, "Speech compression by adaptive Huffman coding using Vitter algorithm", *International Journal of Innovative Sciences, vol. 2, No. 5, May2015.*
3. A. P. Kumar, N. Kumar, C.S.kumar, A.K. Yadav, A. Sharma, "Speech Recognition Using Arithmetic Codinga MFCC for Telugu Language", *3rd International conference on computing for sustainable global development, proceeding IEEE digital library, BVICAm, 2016.*
4. N. Kalyani, K.V.N Sunitha, "Syllable analysis to build a dictation system in Telugu language", *International journal of computer science and information technology, Vol. 6, No. 3, 2009.*
5. A. K. Yadav, R. Roy "De-noising of color image using median filter", *3rd International conference on image information processing, Proceedings of IEEE Digital Library, 978-1-5090-0148-4, Dec 2015.*
6. D.L. Donoho, "Denoising by soft thresholding", *IEEE Transctaions on Information Theory, Vol. 48, PP. 927-940, 1995*.
7. A. K. Yadav, R. Roy, C.S. Kumar, "De-noising of ultrasound image using discrete wavelet transform by symlet wavelet and filters*", Proceedings of IEEE Digital Library, pp. 1204-1208, ISBN- 978-1-4799-8790-0, Aug 2015.*
8. S. Singh, E. G. Rajan, "Vector quantization approach for speaker recognition using MFCC and inverted MFCC", *International Journal of Computer Applications, Vol. 17, No. 1, March 2011.*
9. C. P. Dalmiya, V. S. Dharun, K. P. Rajesh, "An efficient method for Tamil speech recognition using MFCC and DTW for mobile applications*" IEEE Conference on Information and Communication Technologies, 2013.*
10. M. Hossan, S. Memon and M. Gregory, "A novel approach for MFCC feature extraction", *International Conference on Signal Processing and Communication Systems, pp. 1-5, 2010.*
11. M. Rajesh, A. Hema, V. R. Rao, " Continouse speech recognition using joint features derived from modified group delay function and MFCC", *semantic scholor journal, 2012.*
12. S. Sunny, D. Peter, K. Jacob, "Performance of different classifiers in speech recognition" *International Journal of Research and Engineering Technology, Vol. 2, No. 4, 2013.*
13. A. P. Kumar, & Bansal, D, "Digital Arithmetic Coding with AES Algorithm*", IJCA Special Issue on International Conference on Electronic Design and Signal Processing, Vol. 1, No. 2, pp. 15-18, Feb 2013.*