Robust Perceptual Wavelet Packet Features for Recognition of Continuous Kannada Speech

Mahadeva Swamy (mahadevaswamy@vvce.ac.in)
Vidyavardhaka College of Engineering https://orcid.org/0000-0003-4891-1236

D J Ravi
Vidyavardhaka College of Engineering

Research Article

Keywords: Wavelet Packet Decomposition, Acoustic Models, Hidden Markov Model and Deep Neural Networks.

Posted Date: June 14th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-247034/v1

Version of Record: A version of this preprint was published at Wireless Personal Communications on July 21st, 2021. See the published version at https://doi.org/10.1007/s11277-021-08736-1.

Robust Perceptual Wavelet Packet Features for Recognition of Continuous Kannada Speech

Mahadevaswamy[1], D J Ravi[2]

[1] *Research Scholar, Department of Electronics and Communication Engineering*

[2]*Professor, Department of Electronics and Communication Engineering*

[1,2] *Vidyavardhaka College of Engineering, Mysuru, Visvesvaraya Technological University, Karnataka, India*

[1] mahadevaswamy@vvce.ac.in, [2] ravidj@vvce.ac.in,

[1]Orcid ID: 0000-0003-4891-1236 [2]Orcid ID: 0000-0002-5758-5893

Abstract— ***An ASR system is built for the Continuous Kannada Speech Recognition. The acoustic and language models are created with the help of the Kaldi toolkit. The speech database is created with the native male and female Kannada speakers. The 75% of collected speech data is used for training the acoustic models and 25% of speech database is used for the system testing. The Performance of the system is presented interms of Word Error Rate (WER). Wavelet Packet Decomposition along with Mel filter bank is used to achieve feature extraction. The proposed feature extraction performs slightly better than the conventional features such as MFCC, PLP interms of WRA and WER under uncontrolled conditions. For the speech corpus collected in Kannada Language, the proposed features shows an improvement in WRA of 1.79% over baseline features.***

KEYWORDS: **Wavelet Packet Decomposition, Acoustic Models, Hidden Markov Model and Deep Neural Networks.**

1. INTRODUCTION

The frequent pauses between the speech sounds of a speech signal portrays its unique characteristic that distinguishes it from all other signals. The speech database created in uncontrolled conditions of the environment must be processed to implement a robust automatic speech recognition system. Speech is an important and efficient tool of communication. The speech research drawn the attention of infinite researchers and has emerged as one of the important

multidisciplinary research areas in the recent decades. Speaker Independent Speech recognition is the task of identifying the spoken word or sentence irrespective of the speaker. The speech recognition has been performed over the several languages. The UNESCO atlas of the world languages danger report-2009, describes that about 197 Indian languages are in critical situation of being extinct. According to Indian census report Percentage of people speaking local languages has drastically reduced[1]. Speech recognition system is implemented for Assamese Language. The vocabulary size is 10 Assamese words, The task of speech recognition is achieved using Hidden Markov Model, I-vector technique and Vector quantization technique. A 39-dimensional features are derived using Mel Frequency Cepstral Coefficients, Delta-Coefficients, Acceleration Coefficients. The Novel Fusion technique outperforms the conventional techniques such as Hidden Markov Model, I-Vector Techniques and Vector Quantization Technique by achieving speech recognition accuracy of 100%[1]. The ASR system developed and evaluated using a moderate Bengali speech. Then 39-dimensional features are extracted and used to train triphone based HMM technique. The system was able to achieve an accuracy of 87.30%[2]. The speech recognition system is developed for Bangla accent. The Mel LPC features and their delta. The HMM modeling, lead to 98.11% recognition accuracy [3]. A Hindi isolated word recognition system is realized with LPC features and HMM Modeling and an accuracy of 97.14% was achieved corresponding to the word “teen” [4]. Another isolated word recognition system was realized with MFCC features and HTK Toolkit for Hindi language. An accuracy of 94.63% and a WER of 5.37% was achieved [5]. A connected word speech recognition system for Hindi language was proposed using MFCC features and HTK Toolkit. An accuracy of 87.01% was achieved [6].

An isolated digit recognition system was designed using MFCC features and HTK Toolkit for Malayalam isolated words to achieve an

accuracy of 98.5% [7]. LPCC, MFCC, Delta-MFCC, Acceleration coefficients and vector quantization is utilized to build a speaker identification system to yield an accuracy of 96.59%. There is boost in the performance of the system by 3.5% accuracy during testing stage with a consideration to text dependent system[8]. An automatic language identification task is achieved among five Indian languages. The languages selected were are Hindi, Kannada, Telugu and Tamil. All the utterances are created from five native female speakers and five native male speakers. The cepstral features are derived from the speech signals and vector quantization technique based on the codebook concept is used to achieve the task of classification. The system achieved an recognition accuracy of 88.10% in recognizing spoken Kannada sentences[9]. A word recognition system was built for Punjabi language. The LPC feature vectors were extracted from speech signals. The vector quantization and Dynamic time warping techniques were used for implementing the speech recognition system. Experiments were carried out for different code book sizes from 8 to 256. The system was able to achieve a accuracy of 94%[10].

A speaker recognition system was developed for two speech databases. One speech database is created using microphone speech and other speech database is telephone speech. MFCC features are used with the Linear discriminant Analysis technique, Co-variance Normalization, used to train the support vector machines classifier and cosine distance scoring[11]. The speech signal is a complex signal has information of vocal tract and the excitation source. To extract the excitation source information, the Linear Prediction Residual subjected to processing. The LP residual, Phase and Magnitude components are processed at three different levels, segmental level, sub-segmental level and suprasegmental level to derive the language specific excitation source information. The Gaussian Mixture Models are used to perform the classification task[12]. The literature re-

veals that the Kannada ASR system has not been experimented with Perceptual Wavelet Packet features so far. This approach is one of the first over Kannada language by augmenting the implementation of Perceptual Wavelet Packet features over the Kaldi toolkit. The organization of the article is as follows: In section 1 and 2 provides introductory information towards automatic speech recognition and some of the important works presented in the literature. Section 3 describes about feature extraction methods.

2. RELATED WORKS

The automatic speech recognition (ASR) system is able to provide 100% accuracy under clean environment. But, its performance is degrades significantly when the spoken utterances gets contaminated by the presense of background noise or mismatch in acoustic features extracted from noisy or clean condtions [13, 14, 15] and mismatch in the labelled speech data used to train the classifier [16]. Hence, the performance of ASR system is constrained by two choices namely, correct labelling of speech data and selection of acoustic features. The well known acoustic features for speech recognition is Mel-frequency cepstral coefficients (MFCCs). MFCCs are extracted from the Mel filter banks[17]. MFCCs are obtained using short time Fourier transform (STFT). The mel cepstral coefficients are computed by allowing speech signal to pass through a bank triangular shaped filters having passbands slightly overlapping with adjacent passbands and to obtain a smooth spectrum[18,19]. Spectrum is subjected variations as the impact of background noise increases[18,20]. The popular MFCC technique consists of STFT. The STFT has a requirement that the signal to be processed must be stationary over short interval of time i.e.,semi-periodic signals[21]. Due to the trade-off between time-frequency resolution, it is not easy to detect phones that happen with a rapid burst in a slowly changing signal [22,18,20]. This problem of time-frequency resolution is alleviated by using wavelet transform(WT)

[23,24,25]. The major benefit of using wavelet transform is that, unlike using single fixed sized analysis window in STFT, it uses windows with variable duration. The high frequency portion of the speech signal is processed by the short duration window, whereas the low frequency part of the speech signal is processed by the long duration window[24,26-27]. Thus by applying wavelet transform to a speech signal, it can be inspected for the presense or absence of sudden burst (stop phonemes) in a slowly changing signal[20,22]. The conventional wavelet filter bank performed well for phoneme recognition tasks[20]. Because of the fixed resolution of frequency in time-frequency plane, the STFT was not able to find voiced stop due to their characteristic of rapid burst at higher frequencies[20,22]. Multi-resolution potential of wavelets was enormously utilized by many research professionals for feature extraction and demonstrate their benefit for several applications such as, Biomedical application like ECG[28,29], Speech enhancement[30, 31], EEG[32, 33] and Phoneme recognition[20,22,34].

3. METHODOLOGY

3.1 PREPOROCESSING

The preprocessing functions like framing, windowing and pre-emphasis are applied to all the wave files in speech database. The frame duration and frame overlap are choosen as 20msec and 10mes respectively, for performing framing and windowing.

3.2 PROPOSED FEATURES

The Multi-resolution property of the wavelet makes it appropriate tool for handling nonstatinary and semi-stationary signals. This transform can detect unvoiced sounds in the speech signal and it provides best desnoising characteritics. In the recent years, several feature extraction approaches have been invented for speech recognition in uncontrolled environment. But, majority of these feature extraction schemes use Fourier transform to compute the spectrum.

The speech signal consist of voiced (periodic) and unvoiced (aperiodic) portions throughout its existence. It's a popular fact that the STFT or windowed Fourier transform has fixed and uniform frequency resolution with respect to the time frequency plane. Therefore it is difficult for the methods relay on STFT to recognize sudden bursts in the slowly time varying speech signals. To problem is alleviated by the application of wavelet transforms in the speech recognition research [35,36,37,43-47]. The wavelet transform offers good frequency resolution.

3.2.1 Theoretical Background of Wavelet Transforms

Multi Resolution Analysis is an alternative way to STFT technique to analyze a signal. A mathematical scaling function is utilized to obtain a series of approximations to the signal. This principle has been considered by Wavelet Transforms (WT). A comparision of time- frequency resolution between STFT and WT is shown in Figure 1.

3.2.2 Continuos Wavelet Transform (CWT)

CWT of a signal x(t) is given by

$$CWT_x^{\Psi}(\tau, s) = \frac{1}{\sqrt{S}} \int_{-\infty}^{\infty} x(t) \Psi^{*(\frac{t-\tau}{s})} dt \qquad (1)$$

From equation (17), the result of transformation is function of two variables, τ *and* s that describe the translation and scaling factor respectively, and $\Psi(\mathrm{t})$ is mother wavelet.

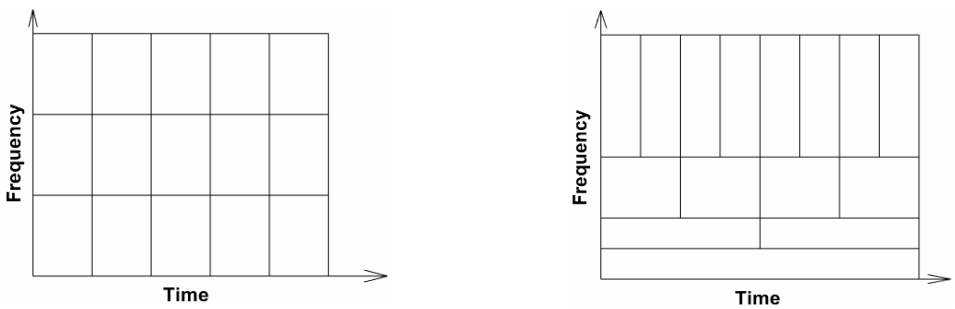

Fig. 1 Comparision of STFT with WT

The term wavelet is concatenation of two words 'wave' and 'let'. Here wave is signal and let is short. The mother wavelet acts as a model or prototype to derive other window functions. The time information is captured by the variable τ and the parameter s specifies dialation or compression operation on the wavelet.

3.2.3 Discrete Wavelet Transform (CWT)

The CWT is more complicated for signal analysis, because it involves significant computational resources. While DWT is less complicated in capture the signal information effectively[49]. The DWT of signal x(t) is defined as:

$$DWT(j,k) = \frac{1}{\sqrt{|2^j|}} \int_{-\infty}^{\infty} x(t)\psi\left(\frac{t-2^j k}{2^j}\right) \tag{2}$$

Mallat successfully demonstrated the method of wavelet decomposition by allowing a signal to pass through a series arrangement of low pass filter and high pass filter pairs. The multi resolution analysis of a signal is shown in Figure 2a and 2b shown below. Here, $h_0(n), h_1(n)$ in the decomposition tree are low pass and high filter pairs respectively. Similarly $g_0(n)$, $g_1(n)$ form the low pass and high pass filter pair in the reconstruction tree.

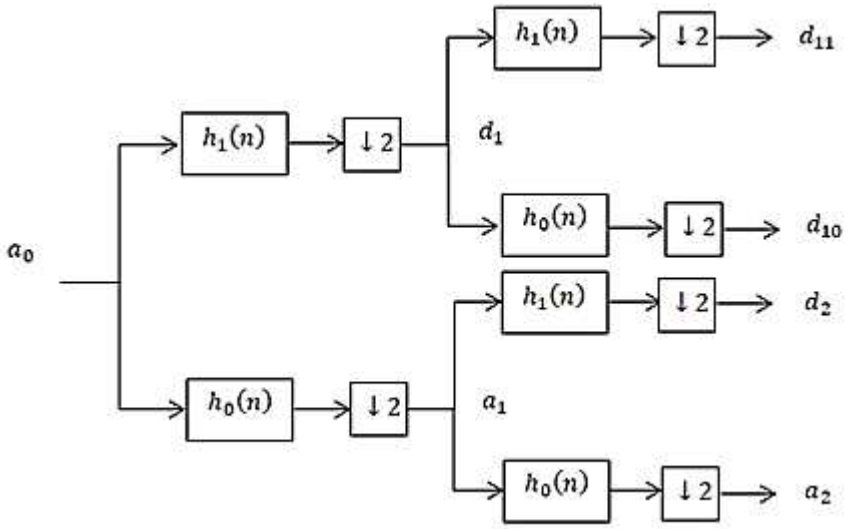

Fig. 2a The balanced 2-level analysis wavelet tree structure for a_0

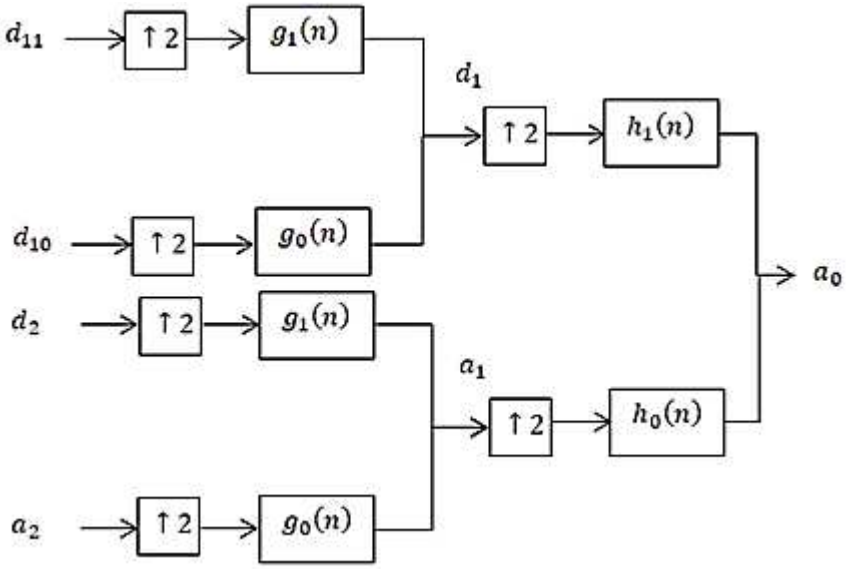

Fig. 2b The balanced 2-level synthesis wavelet tree structure for a_0

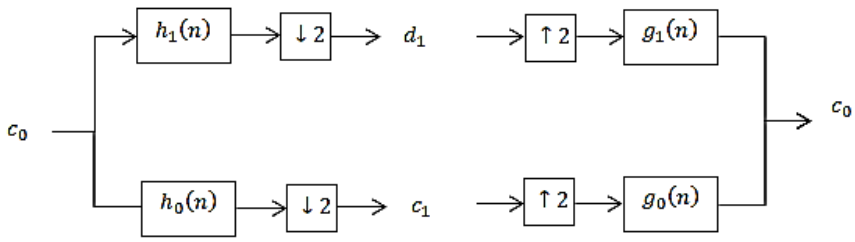

Fig. 3a The one level wavelet analysis and synthesis.

$h_0(n)$ and $h_1(n)$ are a pair filters used for analysis, whereas $g_0(n)$ and $g_1(n)$ form another pair of low, highpass filters respectively. These four filters have related as

$$h_1(n) = (-1)^n g_0(1 - n),$$
$$g_1(n) = (-1)^n h_0(1 - n) \quad (3)$$

Also, the symbols ↓2 and ↑2 presented in Figure 2a and 2b, denote the decimating and interpolating opertions carried out by a factor of 2 respectively. A pair of one level analysis and synthesis trees are shown in Figure 3. In Figure 3, $\{c_0(n)\}n \in Z$ is the input applied the one level analysis and synthesis tree respectively[23].

$$c_1(k) = \sum_n h_0(n - 2k)\, c_0(n) \quad (4)$$

$$d_1(k) = \sum_n h_1(n - 2k)\, c_0(n) \quad (5)$$

where $c_1(k)$ and $d_1(k)$ are known as the approximation space and the detail space respectively. These are created by the one level wavelet analysis of $c_0(n)$. The corresponding synthesis tree is shown in Figure 3 can be operated as

$$c_0(m) = \sum_k [g_0(2k - m)c_1(k) + g_1(2k - m)d_1(k)] \quad (6)$$

3.2.4 Wavelet based acoustic feature extraction

By repeating the iterative decomposition a desired binary wavelet packet tree is obtained. Various WP filterbank tree structures can be derived depending on application of interest. Wavelet features are extracted using Daubachies wavelet of order 4 (db4) [57]. Increaing the order of the mother wavelet may provide better results at expense of increased computational complexity.

3.2.4.1 Mel Filter like WP Decomposition

Farooq et.al.,[20] introduced 24-band Mel like Wavelet Packet Cepstral Features (WMFCC) The sound frequency f_c is mapped to the mel frequency f_{mel} according to the following equation

$$f_{mel} = 2595log_{10}\left(1 + \frac{f_c}{700}\right) \tag{7}$$

A frame size of 25msec with a frame ovelap of 15msec was used to derive the WMFCC. Intially the speech frames are subjected to pre-emphasis followed by windowing operation using Hamming window. Initially a balanced three level wavelet packet tree structure is derived. Here, the frequency axis is subdivided into eight sub-bands each of 1KHz The low frequency subband in the range 0-1KHz is again subjeted to three level balanced decomposition to get eight subbands each having a bandwidth of 125Hz. Which is almost close to 100 Hz Mel-filter. The subband with frequency range is decomposed into two level balanced WP coefficients, giving four subbands each having a bandwidth of 250Hz. The subbands in the range 1-1.25KHz and 1.25-1.5KHz are decomposed again, resulting in four subbands same bandwidth i.e., 250Hz. The subband of 3-4KHz frequency range is again processed by level decomposition, resulting in two subbands of 3-3.5KHz and 3.5-4KHz respectively. The frequency bands with ranges 4-5KHz, 5-6KHz, 6-7KHz, and &-8KHz are retained as it is. This results in 24-band Mel scale resembeled WP filter. The bandwidth of the 24 frequency bands

resulting after WP Decomposition does not exactly follow Mel scale[20] (see Table 1).

Table 1 Comparision of frequency bands of 24-band Mel scale filters and Wavelet Packet sub-band

Filters	Mel Scale	Wavelet Subband	Filters	Mel Scale	Wavelet Subband	Filters	Mel Scale	Wavelet Subband
1	100	125	9	900	1125	17	2639	2750
2	200	250	10	1000	1250	18	3031	3000
3	300	375	11	1149	1375	19	3482	3500
4	400	500	12	1320	1500	20	4000	4000
5	500	625	13	1516	1750	21	4595	5000
6	600	750	14	1741	2000	22	5278	6000
7	700	875	15	2000	2250	23	6063	7000
8	800	1000	16	2297	2500	24	6954	8000

The frequency axis is divided with the intention of matching it to frequency response of the Mel scale. The 24-band wavelet packet sub-bands resemble 24-band Mel filters is shown in Figure 5[].

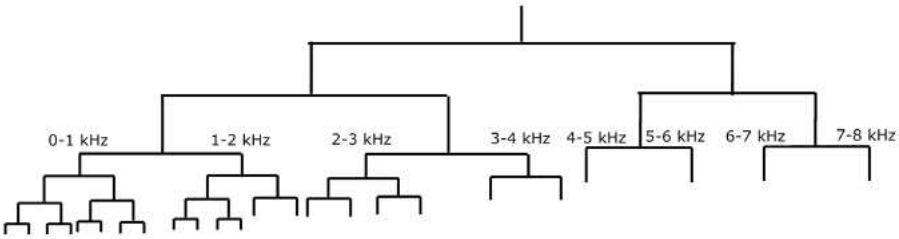

Fig. 5 24-band WP tree based on Mel scale

The energy in each subband is calculated by

$$\langle S_i \rangle_k = \sum \frac{|\omega_\Psi(x,k)_i|^2}{N_i} \tag{8}$$

where, $\omega_\Psi(x,k)_i$ is wavelet packet coefficients of the signal x, i is the subband frequency index ($1 \leq i \leq M$), k indicates the temporal frame and N_i is the number of samples in the ith suband. Similar to MFCC, the 24 energy coefficients are subjected to logarithmic compression. Finally, DCT is applied to all 24 coefficients and only

first 13 normalized DCT coefficients are considered as WMFCC features. The pictorial representation of the feature extraction process is shown in Figure 6 shown below.

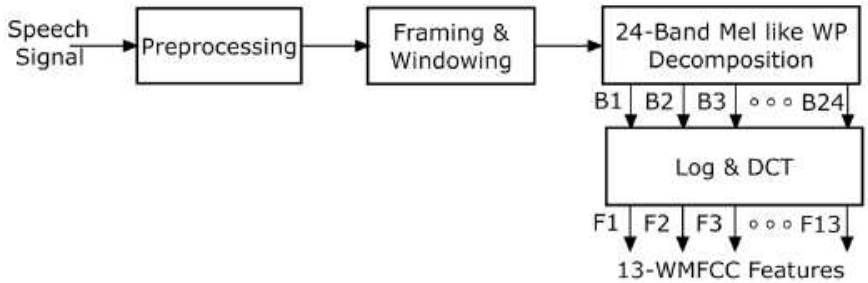

Fig. 6 Steps of acoustic WMFCC feature extraction technique

3.2.4.2 Proposed PWP tree structure for feature extraction

In this work we have proposed a 24-band wavlelet packet tree which is used to obtain the cepstral features. The feature extaction is carried out by proposing a 24-band Wavelet Packet (WP) tree structure after conducting repeated experiments. The WP tree structure shown in Figure 7. is the proposed WP tree structure for obtaining the features.

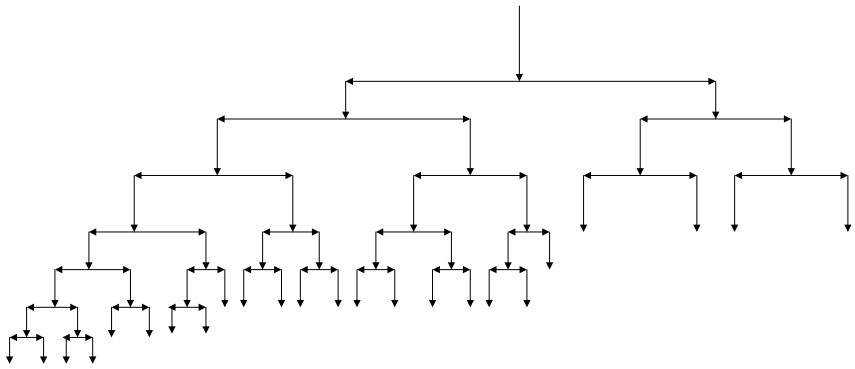

Fig. 7 Proposed 24-band WP tree based on Mel scale

The energy of the 24 band wavelt subbands are calculated. These coefficients are then logarithmically compressed and subjected to Discrete Cosine Transform. Discrete Cosine Transform (DCT)

basically achieves energy compaction. The outpus of DCT gives 24 coefficients and only first 13 coefficients are used as cepstral coefficients. The Kaldi Toolkit is used to determine the delta and delta delta coefficients to form features of 39-dimension.

3.2.5 ACOUSTIC MODELS

The acoustic models are used to map the observed feature matrix with the desired phoneme sequences of the hypothesized sentence. The creation of acoustic models is usually accomplished by using the Hidden Markov Models (HMM).

3.2.6 LANGUAGE MODELS

The ASR systems utilize n-gram language models to facilitate the detection of exact word sequence through prediction of nth word, utilizing $(n-1)$ previous words. Most popular n-gram language models are trigram $(n=3)$ and bigram $(n=2)$ language models.

3.2.7 RECOGNITION

The speech recognition task is achieved using Gaussian Mixture Model -Hidden Markov Model (GMM-HMM),Triphones 1, Triphones 2 and Triphones 3 and Deep Neural Network (DNN).

3.2.8 HIDDEN MARKOV MODEL

To determine the probability $P\left(\frac{W}{X}\right)$ a 3-state Markov chain is used here. The 3- state Markov chain is displayed in the Figure 8. In the training phase, probality of system staying in a state (π), probability of transition between states(A), and output probabilities (B) are determined applying Baum-Welch Algorithm. The HMM Acoustic Model for each word sequence is defind by using the equation

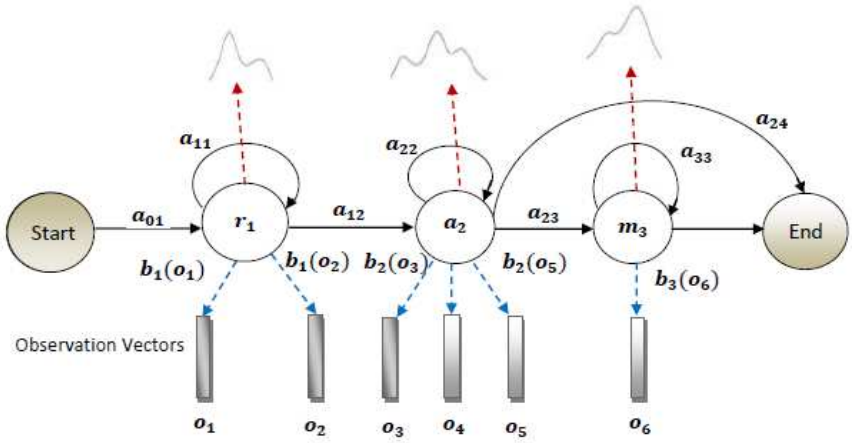

Fig. 8 A 3-state Markov Chain

$$\lambda = (A, B, \pi) \tag{9}$$

The log-likelywood of every word sequence is estimated using Viterbi Decoding technique according to the equation

$$v * = [P\ (O|\lambda v)], \quad 1 \leq v \leq V \tag{10}$$

V is word length.

3.2.9 PERFORMANCE ANALYSIS

The recognition accuracy of any ASR system is determined using the popular metric word error rate and word recognition accuracy [24] given by equations (28) and (29) respectively.

$$WER(\%) = \frac{(D + S + I)}{N} \times 100(\%) \tag{11}$$

$$WRA(\%) = 100 - WER(\%) \tag{12}$$

Where N is the total number of words present in the test set and $D, S\ and\ I$ are erroes due to deletion, substitution and Insertion respectively.

4. DATABASE

The Kannada speech Database consisting of isolated digits from 0-9, 20 isolated wods and 500 continuous speech sentences which includes combination of read speech and spontaneous speech and native language accents have been used as text script for creating the database. The database consists of 100 speakers. The continuous speech is of 15 hours database. The database is recoreded in the natural environment in the presense of room noise, vehicle noise happening nearby road at a distance of 5 meters from the recording room. The tools used for recording the database is Matlab R2019b software on Dell Laptop. The speech data has been collected at a sampling rate of 16KHz, 16-bit resolution and Mono recordings. The table indicates the 50 sample sentences from the speech database along with its goolgle transcription. The lexicon is created using Indian Language Symbol Labels version 3 (ILSLv3) prepared by Samudravijay, Indian Institute of Technology, Guwahati. The continuous speech database created here is according to the guidelines suggested by speech research experts from Indian Institute of Technolgy, Guwahati (IITG). The databsase is partitioned into training set and testing set. The 80% of the database is used for training the Acoustic Models and 20% of the database is used for testing the model.

The database consists of 3 sets for Kannada Language namely: isolated digits through (0-9), isolated words, Continuous Kannada Speech consisting of Spontaneous Spoken Kannada Sentences also. The database consists of 3 sets for English Language namely: isolated digits (TIMIT) through (0-9), isolated words (TIMIT), Librispeech of Continuous English Speech. The Kannada speech sentences along with their transcription is provided in the Table 2.

Table 2 The Kannada text sentences of the speech data collected and their corresponding transliteration.

ಕೊನೆಯಲ್ಲಿ ಮತ್ತೊಮ್ಮೆ ವಾರ್ತೆಗಳ ಮುಖ್ಯಅಂಶಗಳು
koneyalli mattomme vaartegala mukhyaamsagalu
ಸಂಸತ್ತಿನ ಉಭಯ ಸದನಗಳಲ್ಲಿ ರಾಷ್ಟ್ರಪತಿಗಳ ಭಾಷಣ ಮೇಲಿನ ವಂದನಾ ನಿರ್ಣಯದ ಚರ್ಚೆ ಆರಂಭ
samsattina ubhaya sadanagalalli raastrapatigala bhaasanada meelina vandana nirnayada carce aarambha

ಆಧಾರ್ ಕಾಯ್ದೆ ತಿದ್ದುಪಡಿ ಸೇರಿ ಲೋಕಸಭೆಯಲ್ಲಿ ಪ್ರಮುಖ ಮೂರು ಮಸೂದೆಗಳು ಮಂಡನೆ
aadhaar kaaide tiddupadi seeri lookasabheyalli pramukha mooru masuudegalu mandane
ಸೇವಾವಲಯದಲ್ಲಿ ಸುಧಾರಣೆ ತರುವ ಸರ್ಕಾರದ ಪ್ರಾಧಾನ್ಯ ಆದ್ಯತೆ ಮುಂದುವರೆಯಲಿದೆ ವಿದೀಸಾಂಗ ಸಚಿವಾ ಜಯಸಂಕರ್
seevaavalayadalli sudaarane taruva sarkaarada praadhaanya aadyate munduvareyalide videesaanga sacivaa Jayasankar
ಮತ್ತು ವಿಶ್ವಕಪ್ ಕ್ರಿಕೆಟ್ ಆಫ್ಗಾನಿಸ್ತಾನಕ್ಕೆ ಗೆಲ್ಲಲು ಇನ್ನೂರ ಅರವತ್ಮೂರು ರನ್ಗಳ ಗುರಿ ನೀಡಿದ ಬಾಂಗ್ಲಾದೇಶ
mattu visvakap kriket aafgaanistaanakke gellalu innuura aravatmooru rangala guri niidida baanglaadeesa
ಇಲ್ಲಿಗೆ ವಾರ್ತಾ ಪ್ರಸಾರ ಮುಕ್ತಾಯವಾಯಿತು
illige vaartaa prasaara muktaayavaayitu
ವಾರ್ತೆಗಳ ವಿವರ ಸಂಸತ್ತಿನ ಉಭಯ ಸದನಗಳಲ್ಲಿಂದು ರಾಷ್ಟ್ರಪತಿಗಳ ಭಾಷಣಕ್ಕೆ ವಂದನೆ ಸಲ್ಲಿಸುವ ನಿರ್ಣಯದ ಮೇಲಿನ ಚರ್ಚೆ ಆರಂಭವಾಗಿದೆ
vaartegala vivara samsattina ubhaya sadanagalallindu raastrapatigala bhaasanakke vandane sallisuva nirnayada meelina carce aarambhavaagide
ಈ ತಿಂಗಳ ಇಪ್ಪತ್ತನೇ ತಾರೀಖಿನಂದು ಲೋಕಸಭೆ ಮತ್ತು ರಾಜ್ಯಸಭೆ ಜಂಟಿ ಅಧಿವೇಶನ ಉದ್ದೇಶಿಸಿ ರಾಷ್ಟ್ರಪತಿ ರಾಮನಾಥ್ ಕೂವಿಂದ್ ಅವರು ಭಾಷಣ ಮಾಡಿದ್ದರು
ii tingala ippattane taariikhinandu lookasabhe mattu raajyasabhe janti adhiveesana uddeesisi raastrapati raamanaath koovind avaru bhaasana maadiddaru
ಲೋಕಸಭೆಯಲ್ಲಿ ಕೇಂದ್ರ ಸಚಿವ ಪ್ರತಾಪ್ ಚಂದ್ರ ಸಾರಂಗಿ ಅವರು ಚರ್ಚೆಗೆ ಚಾಲನೆ ನೀಡಿದರು
lookasabheyalli keendra saciva prataap candra saarangi avaru carcege caalane niididaru
ಪ್ರಧಾನಿ ನರೇಂದ್ರಮೋದಿ ನೇತೃತ್ವದ ಸರ್ಕಾರ ಕೈಗೊಂಡಿರುವ ಅಭಿವೃದ್ಧಿ ಕ್ರಮಗಳು ಹಾಗು ಯೋಜನೆಗಳ ಪ್ರಮುಖ ಅಂಶಗಳನ್ನು ಉಲ್ಲೇಖಿಸಿದ ಪ್ರತಾಪ್ ಚಂದ್ರ ಸಾರಂಗಿ
pradhaani nareendramoodi neetrutvada sarkaara kaigondiruva abhiruddi kramagalu haagu yoojanegala pramu kha amshagalannu ullekhisida prataap candra Sarangi
ವಿವಿಧ ಯೋಜನೆಗಳಲ್ಲಿ ನೇರ ಸೌಲಭ್ಯ ವರ್ಗಾವಣೆ ಯೋಜನೆ ಯನ್ನು ಜಾರಿಗೆ ತಂದಿರುವ ಸರ್ಕಾರ ವ್ಯವಸ್ಥೆಯಲ್ಲಿ ಮಹತ್ವದ ಬದಲಾವಣೆ ತಂದಿದೆ
vividha yojanegalalli neera saulabhya vargaavane yojaneyannu jaarige tandiruva sarkaara vyavastheyalli mahatvada badalaavane tandide
ಎಂದು ಪ್ರತಾಪ್ ಚಂದ್ರ ಸಾರಂಗಿ ಶ್ಲಾಘಿಸಿದರು
endu prataap candra saarangi shlaaghisidaru
ಉತ್ತಮ ಕಾರ್ಯಕ್ರಮ ಮತ್ತು ಯೋಜನೆಗಳ ಫಲವಾಗಿ ಎನ್ ಡಿ ಎ ಮೈತ್ರಿ ಕೂಟವನ್ನು ಮತದಾರರು ಪ್ರಚಂಡ ಬಹುಮತದಿಂದ ಎರಡನೇ ಅವಧಿಗೆ ಆಯ್ಕೆ ಮಾಡಿದ್ದಾರೆ ಎಂದು ಹೇಳಿದರು
uttama kaaryakrama mattu yojanegala phalavaagi en di e maitri kuutavannu matadaararu pracanda bahumatadinda eradane avadhige aayke maadiddare endu heelidaru
ಆಕಾಶವಾಣಿ ವಾರ್ತೆಗಳು ಓದುತ್ತಿರುವವರು ಹೇಮಂತ್
aakaashavaani vaartegalu ooduttiruvavaru heemanth
ಈ ಸಮಿತಿಯನ್ನು ಎರಡುಸಾವಿರದ ಹದಿನೇಳರ ಜೂನ್ ಇಪ್ಪತ್ತನಾಲ್ಕರಂದು ರಚಿಸಲಾಗಿದ್ದು ದೇಶಾದ್ಯಂತ ಎಪ್ಪತ್ತು ಕಡೆಗಳಲ್ಲಿ ಸಮಿತಿ ಸಭೆ ಸೇರಿ ಜನಾಭಿಪ್ರಾಯ ಸಂಗ್ರಹ ಮಾಡಿದೆ
ii samiitiyannu eradsaavrada hadineelara juun ippattnaalkarandu racisalaagiddu deesaadyanta yeppattu kadegalalli samiti sabe seeri janaabipraaya sangraha Maadide
ದೇಶ ವಿದೇಶಗಳ ಶಿಕ್ಷಣ ತಜ್ಞರಿಂದಲೂ ಸಮಿತಿ ಅಭಿಪ್ರಾಯಗಳನ್ನು ಪಡೆದಿದೆ ಎಂದು ಹೇಳಿದರು
deesa videesagala siksana tagnarindalu samiti abipraayagalannu padedide endu heelidaru
ಎಲ್ಲ ರಾಜ್ಯಗಳಲ್ಲಿ ಹಿಂದಿಯನ್ನು ಕಡ್ಡಾಯ ಭಾಷೆಯನ್ನಾಗ ಮಾಡಲು ಬಯಸುತ್ತೀರಾ ಎಂಬ ಪ್ರಶ್ನೆಗೆ ಸಚಿವರು ಪ್ರತಿಕ್ರಿಯೆ ನೀಡಿ.
ella raajyagalalli hindiyannu kaddaaya bhaaseyannaaga maadalu bayasuttiira emba prasnege sacivaru pratikriye Nidi

ವಾಸ ಶಿಕ್ಷಣ ನೀತಿಯ ಕರಡನ್ನು ಈಗಷ್ಟೆಯೇ ಸಿದ್ಧಪಡಿಸಲಾಗಿದೆ ಇದಕ್ಕಾಗಿ ಜನರಿಂದ ಸಲಹೆಗಳನ್ನು ಆಹ್ವಾನಿಸಲಾಗಿದೆ
vaasa siksana niitiya karadannu iigastee siddapadisalaagide idakkaagi janarinda salahegalannu aah-vaanisalaagide
ದಯವಿಟ್ಟು ಇದನ್ನು ಓದಿ ಮತ್ತು ಸುಧಾರಿಸಲು ಸಲಹೆಗಳನ್ನು ನೀಡಿ, ಇದು ಕೇವಲ ಕರಡು: ಮಾತ್ರ, ಸರ್ಕಾರದ ಅಂತಿಮ ನೀತಿಯಲ್ಲ ಎಂದು ಸ್ಪಷ್ಟಪಡಿಸಿದರು
dayavittu idannu odi mattu sudaarisalu salahegalannu niidi idu keevala karadu maatra sarkaarada antima niitiyalla endu spastapadisidaru
ಈ ವಾರ್ತೆಗಳನ್ನು ಆಕಾಶವಾಣಿಯಿಂದ ಕೇಳುತಿದ್ದೀರಿ
ii vaartegalannu aakaasavaaniyinda keeluttiddiiri
ಚಾಲನಾ ಪರವಾನಗಿ ಪತ್ರದಲ್ಲಿ ಬದಲಾವಣೆ ಮಾಡಲು ಕೇಂದ್ರ ಭೂಸಾರಿಗೆ ಮತ್ತು ಹೆದ್ದಾರಿ ಸಚಿವಾಲಯ ತೀರ್ಮಾನಿಸಿದ್ದು ಸ್ಮಾರ್ಟ್ ಕಾರ್ಡ್ ಮಾದರಿಯ ಪರವಾನಗಿ ಪತ್ರ ನೀಡಲು ನಿರ್ಧರಿಸಿದೆ
caalana paravaanagi patradalli badalaavane maadalu keendra bhuusaarige mattu heddari saci-vaalaya tiirmaanisiddu smaart kaard maadariya paravaanagi patra niidalu nirdariside
ರಾಜ್ಯಸಭೆಗೆ ಇಂದು ಲಿಖಿತ ಉತ್ತರದಲ್ಲಿ ವಿಷಯ ತಿಳಿಸಿರುವ ಕೇಂದ್ರ ಭೂಸಾರಿಗೆ ಹಾಗು ಹೆದ್ದಾರಿ ಸಚಿವ ನಿತಿನ್ ಗಡ್ಕರಿ
raajyasabhege indu likhita uttaradalli visaya tilisiruva keendra bhuusaarige haagu heddari saciva nitin gadkari
ಇಡೀ ದೇಶಕ್ಕೆ ಅನ್ವಯವಾಗುವಂತೆ ಏಕರೂಪ ಮಾದರಿಯ ವಿನ್ಯಾಸವನ್ನು ಸಿದ್ಧಪಡಿಸಿ ಗುಣಮಟ್ಟದ ಚಾಲನಾ ಪರವಾನಗಿ ಪತ್ರ ನೀಡಲು ತೀರ್ಮಾನಿಸಲಾಗಿದೆ ಎಂದು ತಿಳಿಸಿದ್ದಾರೆ
idii deesakke anvayavaaguvante eekaruupa maadariya vinyaasavannu siddapadisi gunamattada caalanaa paravaanagi patra niidalu tiirmaanisalaagide endu tilisiddare
ಚಾಲನಾ ಪರವಾನಗಿ ಪತ್ರದ ಅರ್ಜಿ ಸಿದ್ಧಪಡಿಸುತ್ತಿದ್ದು
caalanaa paravaanagi patrada arji siddapadisuttiddu
ಚಾಲನಾ ಪರವಾನಗಿ ಹೊಂದಿರುವ ಪ್ರತಿಯೊಬ್ಬರ ದತ್ತಾಂಶಗಳನ್ನು ಈ ವ್ಯವಸ್ಥೆಯಡಿ ಅಳವಡಿಕೆ ಮಾಡಲಾಗುವುದು
caalana paravaanagi hondiruva pratiyobbara dattaamsagalannu ii vyavasteyadi alavadike maada-laaguvudu
ಪ್ರಸ್ತುತ ಚಾಲನಾ ಪರವಾನಗಿ ಹೊಂದಿರುವ ಹದಿನೈದು ಕೋಟಿ ಜನರ ಮಾಹಿತಿ ಲಭ್ಯವಿದೆ labhya-vide
prastuta caalana paravaanagi hondiruva hadinaidu kooti janara maahiti
ಸ್ಥಳದಲ್ಲಿಯೇ ಪರವಾನಗಿದಾರರ ಸಮಗ್ರ ಮಾಹಿತಿ ಒದಗಿಸುವ ಭವಿಸ್ಯದ ಇದಾಗಿದೆ
staladalliye paravaanagidaarara samagra maahiti odagisuva bhavisyada idaagide
ದೇಶದಲ್ಲಿ ಸಣ್ಣ ಸೂಕ್ಷ್ಮ ಹಾಗೂ ಮಧ್ಯಮ ಉದ್ಯಮಗಳ ಸಂಖ್ಯೆಯನ್ನು ವಿವಿಧ ಯೋಜನೆಗಳು ಹಾಗು ಮತ್ತು ಕಾರ್ಯಕ್ರಮಗಳನ್ನು ಸಮರ್ಪಕವಾಗಿ ಬಳಸಿಕೊಳ್ಳಲು ಕೇಂದ್ರ ಸೂಕ್ಷ್ಮ, ಸಣ್ಣ ಹಾಗೂ ಮಧ್ಯಮ ಸಚಿವಾಲಯ ಮುಂದಾಗಿದೆ
deesadalli sanna suuksma haagu madhyama udyamagala sankheyannu vivida yoojanegalu haagu mattu kaaryakramagalannu samarpakavaagi balasikollalu keendra suuksma sanna haagu madyama sacivaalaya mundaagide
ರಾಜ್ಯಸಭೆಯಲ್ಲಿಂದು ಈ ಕುರಿತ ಪೂರಕ ಪ್ರಶ್ನೆಗೆ ಉತ್ತರಿಸಿದ ಕೇಂದ್ರ ಸಚಿವ ನಿತಿನ್ ಗಡ್ಕರಿ
raajyasabheyallindu ii kurita puuraka prasenege uttarisida keendra saciva nitin gadkari
ಈ ವಲಯಕ್ಕೆ ಉತ್ತಮ ಸಾಲ ಸೌಲಭ್ಯ, ಸೂಕ್ತ ಕೌಶಲ್ಯ ಮತ್ತು ಆಧುನಿಕ ತಂತ್ರಜ್ಞಾನವನ್ನು ಒದಗಿಸಿ, ಸಂಪೂರ್ಣ ಪರಿಸರ ಸ್ನೇಹಿ ವ್ಯವಸ್ಥೆ ಕಲ್ಪಿಸಲಾಗುವುದು ಎಂದು ತಿಳಿಸಿದ್ದಾರೆ
ii valayakke uttama saala saulabhya suukta kausallya mattu aadhunika tantragnaanavannu odagisi sampuurna parisara sneehi vyavaste kalpisalaaguvudu endu tilisiddare
ಪ್ರಧಾನಮಂತ್ರಿ ಉದ್ಯೋಗ ಸೃಷ್ಟಿ ಯೋಜನೆ ಸೂಕ್ಷ್ಮ ಮತ್ತು ಸಣ್ಣ ಉದ್ಯಮವಲಯದ ಅಭಿವೃದ್ಧಿಯಂತಹ ಪ್ರಮುಖ ಕಾರ್ಯಕ್ರಮಗಳನ್ನುಅನುಷ್ಠಾನಗೊಳಿಸಲಾಗುವುದು: ಪ್ರಧಾನಮಂತ್ರಿ

pradhaanamantri udyooga srusti yoojane suuksma mattu sanna udyamavalayada abhiruddiyantaha pramukha kaaryakramagalannu anustaanagolisalaaguvudu pradhaanamantri
ಉದ್ಯೋಗ ಸೃಜನ ಯೋಜನೆ ಒಳಗೊಂಡ ಸಾಲಸೌಲಭ್ಯ ಕಲ್ಪಿಸುವ ಯೋಜನೆಯಾಗಿದೆ ಎಂದು ತಿಳಿಸಿದರು
udyooga srujana yoojane olagonda saalasaulabhya kalpisuva yojaneyaagide endu tilisidaru
ಸೇವಾವಲಯದಲ್ಲಿ ಕಳೆದ ಐದು ವರ್ಷಗಳಲ್ಲಿ ತಂದ ಸುಧಾರಣೆಗಳು ಮುಂದುವರಿಯಲಿವೆ ಎಂದು ವಿದೇಶಾಂಗ ಸಚಿವಾ ಜಯಶಂಕರ್ ತಿಳಿಸಿದ್ದಾರೆ
seevaavalayadalli kaleda aidu varsagalalli tanda sudaaranegalu munduvareyalive endu videesaanga sacivaa jayasankar tilisiddaare
ನವದೆಹಲಿಯಲ್ಲಿಂದು ಏಳನೇ ದಿನದ ಅಂಗವಾಗಿ ಆಯೋಜಿಸಲಾಗಿದ್ದ ಸಮಾರಂಭದಲ್ಲಿ ಮಾತನಾಡಿದ ಅವರು ದೇಶದ ಭದ್ರತಾ ವಿಚಾರದಲ್ಲಿ ರಾಜಿ ಆಗದೆ ವಿತರಣೆ ಪ್ರಕ್ರಿಯೆ ಹಾಗು ನಿಯಮಗಳನ್ನು ಸರಳೀಕರಣಗೊಳಿಸುವುದನ್ನು ಸರ್ಕಾರ ಮುಂದುವರಿಸಲಿದೆ ಎಂದು ತಿಳಿಸಿದರು
navadehaliyallindu elane dinada angavaagi aayojisalaagidda samaarambhadalli maatanaadida avaru deesada bhadrataa vicaaradalli raaji aagade vitaraneprakriye haagu niyamagalannu saraliika-ranagolisuvudannu sarkaara munduvarisalide endu tilisidaru

4.1 KALDI TOOLKIT

Kaldi is a open source toolkit designed excusively for building Acoustic Models (AMs) and Language Models (LMs) [41]. Kaldi is built using C++ programming language. The Kaldi toolkit can be run in windows as well as Linux based operating systems. But, the support for Linux based Kaldi tasks is very good compared to that of windows. The Table 3 presents the labels for kannada phones using syllable transliteration. There are four Dravidan languages. Kannada, Telugu, Malayalam and Tamil. Kannada is the most popular Dravidan language used in Karnataka state. This language consist of 14 swara (vowels), 32 vyanhana (consonants), 2 part vowel, yogavaahaka (part consonant). The labels used for building the lexicon for phonemes of the Kannada Language is shown in Table 2. Therefore, The Kannada language ASR system is developed by modeling the 46 phonemes. The labels are used from the Indian Speech Sound Label Set (ILSL12) is shown in Table 4. The lexicon for Kannada language is written by using ILSL12 label set shown in Table 5.

Table 3 The labels using syllable transliteration tool (IT3 to UTF-8) for Kannada phones

LABEL SET USING IT3:UTF-8	KANNADA PHONEMES

a	oo	t:h	ph	ಅ	ಓ	ಠ	ಫ
aa	au	d	b	ಆ	ಔ	ಡ	ಬ
i	k	d:h	bh	ಇ	ಕ	ಢ	ಭ
ii	kh	nd	m	ಈ	ಖ	ಣ	ಮ
u	g	t	y	ಉ	ಗ	ತ	ಯ
uu	gh	th	r	ಊ	ಘ	ಥ	ರ
e	c	d	l	ಎ	ಚ	ದ	ಲ
ee	ch	dh	v	ಏ	ಛ	ಧ	ವ
ai	j	n	sh	ಐ	ಜ	ನ	ಶ
o	t:	p	s	ಒ	ಟ	ಪ	ಸ

Table 4 The labels used from the Indian Speech Sound Label Set (ILSL12) for Kannada phonemes

LABEL SET USING IT3:UTF-8				KANNADA PHONEMES			
a	oo	txh	ph	ಅ	ಓ	ಠ	ಫ
aa	au	dx	b	ಆ	ಔ	ಡ	ಬ
i	k	dxh	bh	ಇ	ಕ	ಢ	ಭ
ii	kh	nx	m	ಈ	ಖ	ಣ	ಮ
u	g	t	y	ಉ	ಗ	ತ	ಯ
uu	gh	th	r	ಊ	ಘ	ಥ	ರ
e	c	d	l	ಎ	ಚ	ದ	ಲ
ee	ch	dh	w	ಏ	ಛ	ಧ	ವ
ai	j	n	sh	ಐ	ಜ	ನ	ಶ
o	tx	p	s	ಒ	ಟ	ಪ	ಸ

Table 5 Dictionary for Kannada language is created by using ILSL12 label set

TEXT TRANSCRIPTION	LABEL SET USING ILSL12
koneyalli	k o n e y a llx i
mattomme	m a t t o mm e
vaartegala	v aa r t e g a lx
mukhyamshagalu	m u k h y aa nx s h a g a lx u
samsattina	s a nx s a tt i n a
ubhaya	u b h a y a
sadanagalalli	s a d a n a g a lx llx i
raastrapatigala	r aa s t r a p a t i g a lx

bhasanada	b h aa s a nx d a
meelina	m ee l i n a
vandana	v a n d a n a
nirnayada	n i r nx y a d a
carce	c a r c e
aarambha	aa r a nx b h a
vaartegala	v aa r t e g a lx
vivara	v i v a r a
samsattina	s a nx s a tt i n a
ubhaya	u b h a y a
sadanagalallindu	s a d a n a g a lx llx i nx d u
raastrapatigala	r aa s t r a p a t i g a lx
bhaasanakke	b h aa s a nx kk e
vandane	v a n d a n e
sallisuva	s a llx i s u v a
nirnayada	n i r nx y a d a
meelina	m ee l i n a
carce	c a r c e
aarambhavaagide	aa r a nx b h a v aa g i d e

The dictionary is created by using ILSL12. Figure 10 represents the block diagram of the proposed features.

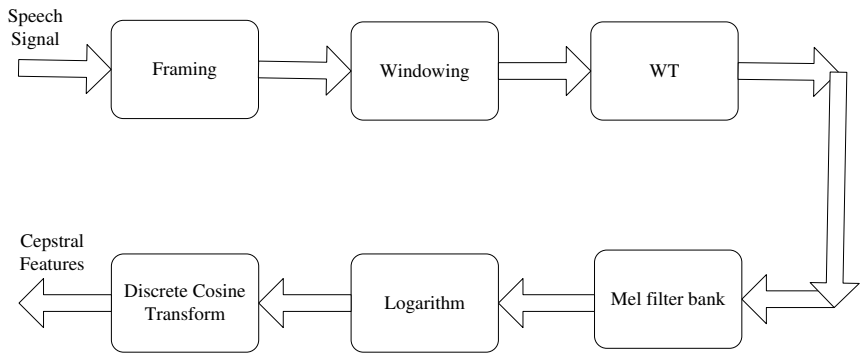

Fig. 10 Block diagram of proposed features

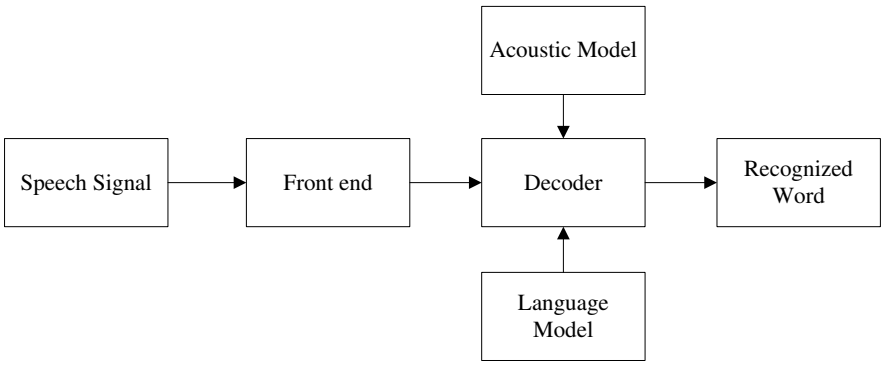

Fig. 11 ASR system architecture

The general architecture of ASR system is shown in Figure 11.

The Table 6 provides the details of parameters used for acoustic modelling. The acoustic models are generated at Monophones, Triphones1 and Triphones3 levels with number of jobs 3. The parameters used to develop Acoustic Model are as follows:

Table 6 Parameters of the Acoustic Models

Patameters Specific to Acoustic Model	Triphone 1	Triphone2	Triphone3
Number of Leaves	2500	2500	2500
Number of Gaussian	20000	20000	20000

5. RESULTS

The results of developed ASR system is presented in this section for Monophones, Triphones1, Triphones2, Triphones3 and DNN-HMM Phoneme Models. The Table 7 gives the WER details for the Kannada Isolated digit recognition task. The pictorial representation of Table 7 are presented in Figure 12.

Table 7 Kannada Isolated digit recognition

SET1 DIGITS	Features->	MFCC	PLP	Proposed
WER	Mono	4.80	7.20	20.00
	Tri 1	2.80	5.60	10.40
	Tri 2	5.60	8.00	8.80

	Tri 3	3.60	3.20	2.80
	DNN-HMM	4.40	4.00	2.80

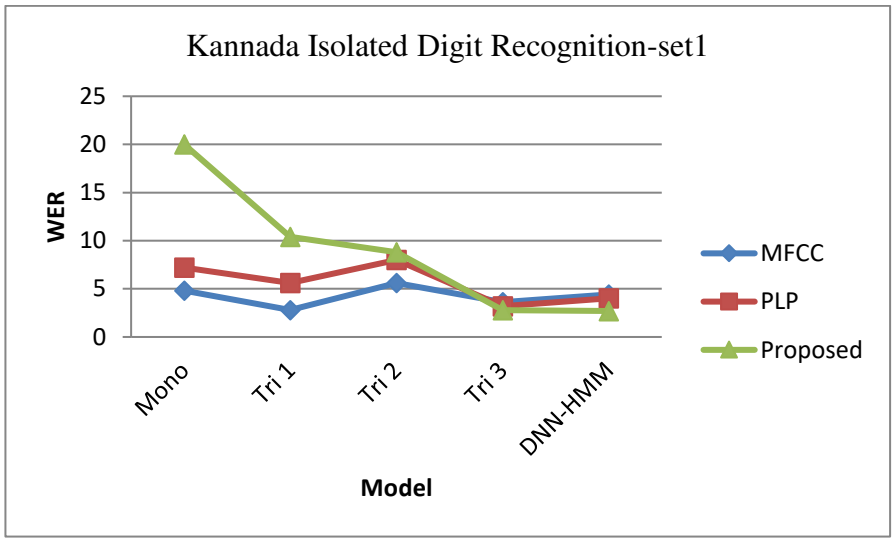

Fig. 12 Comparision of WER for Kannada Isoalted Digit Recognition over MFCC, PLP, Proposed features

The Kannada isolated word recognition results and the corresponding graph are presented in Table 8 and Figure 13 respectively.

Table 8 Kannada Isolated Word recognition

SET2_WORDS	Features	MFCC	PLP	Proposed
WER	Mono	2.25	2.80	2.77
	Tri 1	1.30	2.46	3.08
	Tri 2	1.23	1.52	1.85
	Tri 3	0.92	1.30	1.23
	DNN-HMM	0.92	1.23	0.31

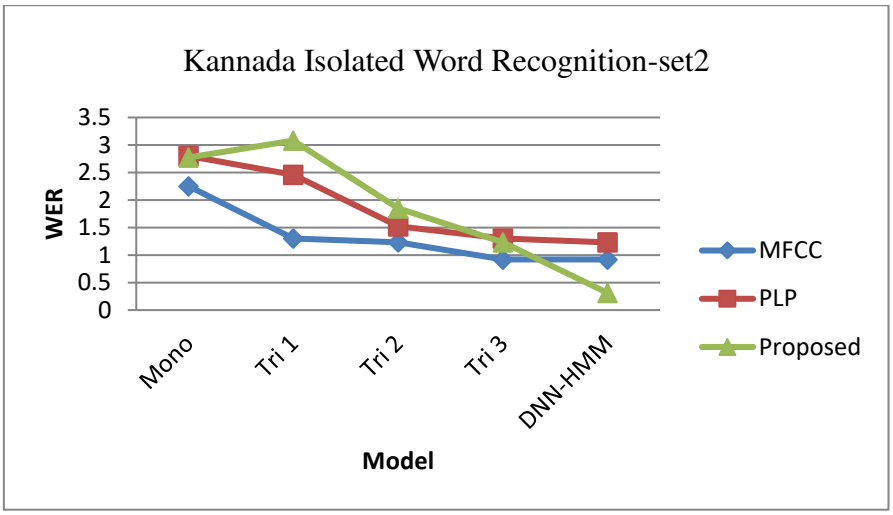

Fig. 13 Comparision of WER for Kannada Isoalted Word Recognition over MFCC, PLP, Proposed features

Table 9 Continuous Kannada speech recognition using MFCC, PLP, Proposed features

SET3_SENTENCES	Features	MFCC	PLP	Proposed
WER	Mono	07.37	06.39	16.20
	Tri 1	08.07	07.35	08.96
	Tri 2	12.23	11.49	14.09
	Tri 3	08.01	07.17	07.01
	DNN-HMM	06.06	06.33	04.27

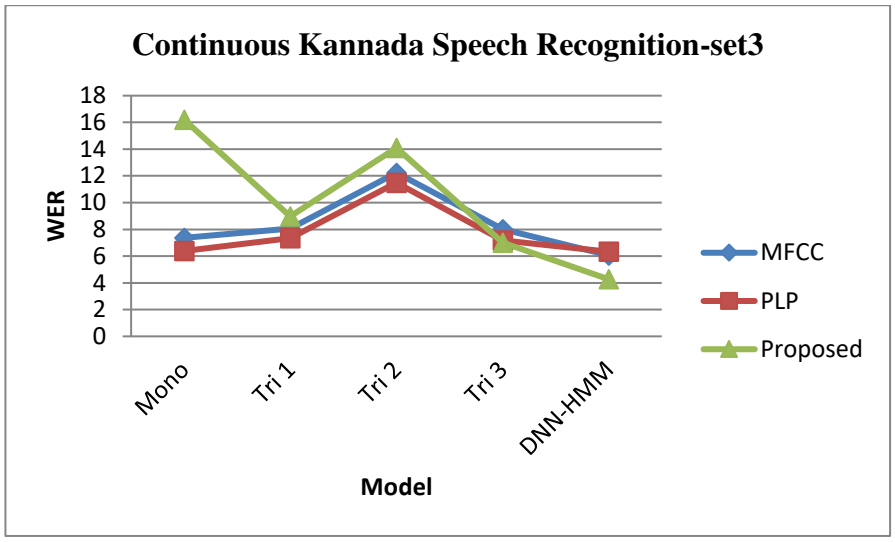

Fig. 14 Comparision of WER for Kannada Continuous Speech Recognition over MFCC, PLP, Proposed features

The WER details for Continuous Kannada speech recognition for the speech data collected in uncontrolled conditions are presented in Table 9 and Figure 14 respectively. In all the three sets of Kannada language a slight improvement in the performance can be observe with the proposed features over the MFCC and PLP features for DNN-HMM classifier.

Table 10 English Isolated digit recognition

SET4 DIGITS	Features	MFCC	PLP	Proposed
WER	Mono	4.70	7.30	20.23
	Tri 1	2.60	5.70	10.40
	Tri 2	5.64	8.00	08.82
	Tri 3	2.23	2.28	02.20
	DNN-HMM	2.00	2.00	01.23

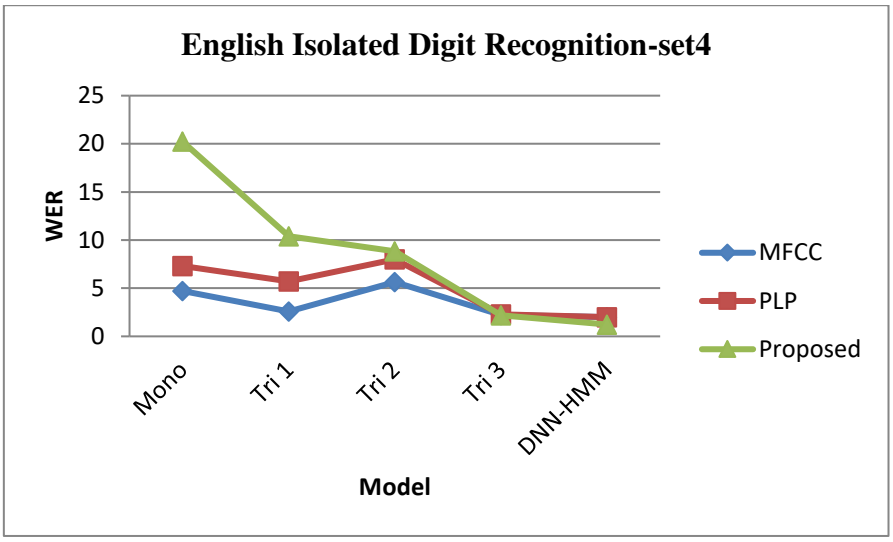

Fig. 15 Comparision of WERs for English Isolated Digit Recognition over MFCC, PLP, Proposed features

The proposed features are also experimented with the isolated digits and isolated words extracted from the TIMIT database. The Table 10 and Figure 15 describes the performance of proposed features over the MFCC and PLP features. The Table 11 and Figure 16 describe the performance of the proposed features over the MFCC and PLP features.A slight improvement in the performance can be observed in Table 11.

Table 11 English Isolated word recognition

SET5 WORDS	Features	MFCC	PLP	Proposed
WER	Mono	1.50	4.25	4.75
	Tri 1	1.39	4.00	3.25
	Tri 2	1.31	3.50	1.50
	Tri 3	1.25	2.25	1.00
	DNN-HMM	1.01	1.00	0.75

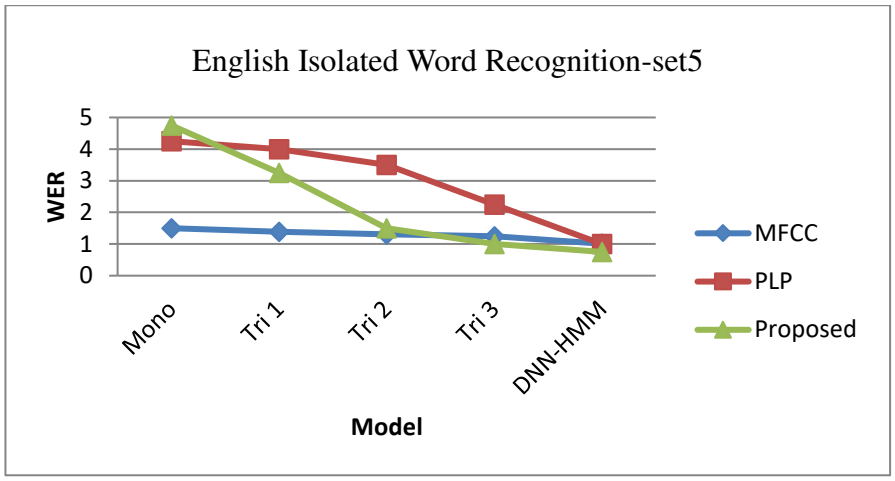

Fig. 16 Comparision of WERs for English Isolated Word Recognition over MFCC, PLP, Proposed features

The proposed features are also experimented with standard Librispeech corpus of 08 hours. A little improvement in the performance can be observed for the proposed features over the baseline features such as MFCC and PLP. The WER details are recoreded in the Table 12 and Figure 17.

Table 12 English Continuous speech recognition recognition

SET6 SENTENCES	Features	MFCC	PLP	Proposed
WER	Mono	62.04	61.17	77.21
	Tri 1	38.75	37.72	69.95
	Tri 2	35.49	35.40	40.33
	Tri 3	30.86	30.29	30.26
	DNN-HMM	30.68	35.20	30.42

The Proposed ASR system is also tested with the unseen data consisting of 512 sentences of different combinations of words. A sample of 20 sentences are shown in the Table 13 and the results of the experiment are included in the Table 14

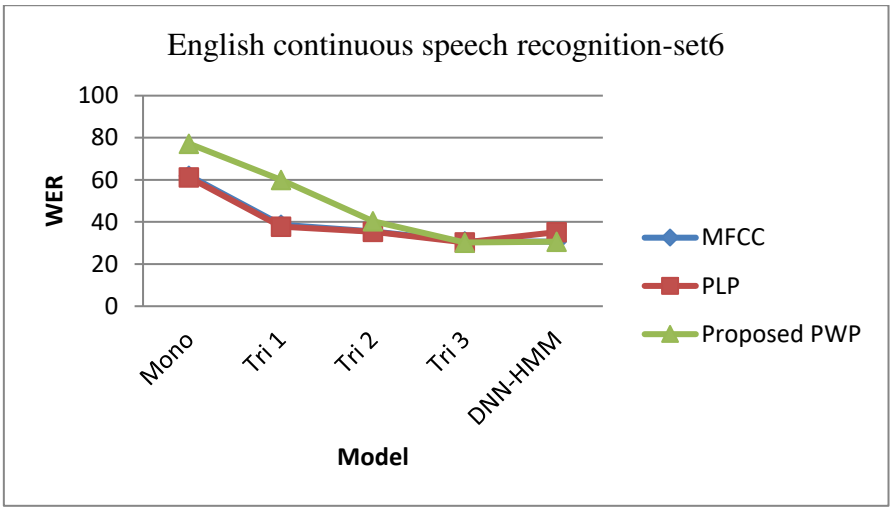

Fig. 17 Comparision of WERs for English Continuous Speech Recognition over MFCC, PLP, Proposed features

Table 14 Comparison of WER among MFCC, PLP and proposed features on Kannada database

TASKS (53-55)	Features	MFCC	PLP	Proposed
WER	Mono	07.80	07.40	10.40
	Tri 1	04.40	04.00	05.60
	Tri 2	03.20	03.80	02.80
	Tri 3	02.80	03.20	02.40
	DNN-HMM	02.60	03.00	02.20

6. CONCLUSION

The ASR work carried out in this paper are as follows.

- We have experimented the conventional as well as proposed feature extraction technique over Monophone Models, Tri-phones1, Triphones2, triphones3 and DNN-HMM.
- The database consists of 3 sets for Kannada Language namely: isolated digits through (0-9), isolated words, Continuous Kan-

nada Speech consisting of Spontaneous Spoken Kannada Sentences also.

- The database consists of 3 sets for English Language namely: isolated digits (TIMIT) through (0-9), isolated words (TIMIT), Librispeech of Continuous English Speech.
- In the experiments conducted over isolated digits and words taken from collected data of Kannada Language and from TIMIT data, the proposed features achieved significant improvement in the performance over the baseline features such as MFCC, PLP.
- For the experiments on collected Kannada Continuous Speech and Librispeech the proposed features are shown to perform better than the conventional features such as MFCC and PLP features.
- The Proposed ASR system is tested with the unseen data of 512 sentences and the performance on this test data set reveals that the proposed system performs better than the conventional features such as MFCC and PLP features.

DECLARATION

This work is not funded by any of the funding agencies. The authors comply with the ethical standards of the journal. The authors also declare that they have no conflict of interest**.** Author's would like to express their gratitude to Management, Principal, HOD-ECE, Vidyavardhaka College of Engineering, Mysuru for their constant support, motivation and encouragement in completing this work.

Table 13 Continuous Kannada speech sentences (unseen data) used for only for testing Kannada ASR system.

Sl. No.	Kannada Sentence	Kannada Transcription	English Translation
1.	ನಮ್ಮ ಜೀವನದ ಪ್ರಯಾಣ ತುಂಬಾ ಕಷ	Nam'ma jīvanada prayāṇa tumbā kaṣṭa	The journey of our life is very difficult
2.	ಆರೋಗ್ಯ ಮುಖ್ಯ ಎಂದು	Ārōgya mukhya endu kathe bari	Tell the story that health

	ಕಥೆ ಬರಿ		is important
3.	ನಿನ್ನ ಮುಂದಿನ ಪ್ರಯಾಣ ಏನೇ ಇರಲಿ	Ninna mundina prayāṇa ēnē irali	whatever your next journey
4.	ಆದರೆ ನಿನಗೆ ಇಲ್ಲಿ ಎಲ್ಲಕ್ಕಿಂತ ಕಷ್ಟ	Ādare ninage illi ellakkinta kaṣṭa	but it's hard for you here
5.	ಆದರೆ ನಂಬಿಕೆ ನಿನಗೆ ತುಂಬಾ ಮುಖ್ಯ	Ādare nambike ninage tumbā mukhya	But faith is very important to you
6.	ಮುಂದಿನ ಪ್ರಯಾಣ ಯಾವುದು ಎಂದು ಬರಿ	Mundina prayāṇa yāvudu endu bari	Write about the next journey
7.	ನಿನ್ನ ಮುಂದಿನ ಕೆಲಸ ಏನೇ ಇರಲಿ	Ninna mundina kelasa ēnē irali	Whatever your next job
8.	ಆದರೆ ಬಡವರ ಜೀವನದ ಬೆಳಕು ಏನು	Ādare baḍavara jīvanada beḷaku ēnu	But what is the light of the life of the poor
9.	ಅವರ ಜೀವನ ಬರಿ ಅಸತ್ಯ ಕಥೆ	Avara jīvana bari asatya kathe	His life is simply untrue
10.	ಆದರೆ ನನ್ನ ಕಥೆ ಇಲ್ಲಿ ಮುಖ್ಯ	Ādare nanna kathe illi mukhya	But my story is important here
11.	ಕೊನೆಯಲ್ಲಿ ಮತ್ತೊಮ್ಮೆ ಸುದ್ದಿಗಳ ವಿವರವಿದೆ	Koneyalli mattomme vaartegala vivara	At the end is the detail of the news once again
12.	ಮತ್ತೊಮ್ಮೆ ಸಂಸತ್ತಿನ ಸದನಗಳಲಿ	Mattomme samsattina sadanagalalli	Once again in the houses of parliament
13.	ವಾರ್ತೆಗಳ ವಿವರ ಆರಂಭ	Vaartegala vivara aarambha	The beginning of the news detail
14.	ಮಾತಿನ ಮುಖ್ಯ ಅಂಶಗಳು	Bhaasanada mukhyaamshagalu	The main elements of talk
15.	ಸಂಸತ್ತಿನ ಮೇಲಿನ ನಿರ್ಣಯದ ಚರ್ಚೆ	Samsattina meelina nirnayada carce	The debate on the resolution on parliament
16.	ವಾರ್ತೆಗಳ ಚರ್ಚೆ ವಿವರ ಆರಂಭ	Varthegala carce vivara aarambha	Beginning of the discussion of the news
17.	ಸದನಗಳಲ್ಲಿಂದು ಉಭಯ ಸಂಸತಿನ	Sadanagalallindu ubhaya samsattina	Dual parliament in the house
18.	ರಾಷ್ಟ್ರಪತಿಗಳ ಮೇಲಿನ ವಂದನೆ	Raastrapatigala meelina vandane	Salute to the president
19.	ಭಾಷಣಕ್ಕೆ ಸಲ್ಲಿಸಬೇಕಾದ ವಿವರಗಳು	Bhaasanakke sallisuva vivara	Details to submit to the speech
20.	ನಿರ್ಣಯದ ಚರ್ಚೆ ಆರಂಭವಾಗಿದೆ	Nirnayada carce aarambhavaagide	The resolution debate has begun

REFERENCES

1. Bharali, S.S. & Kalita, S.K. “Speech recognition with reference to Assamese language using novel fusion technique”, Int J Speech Technol 2018, 21: 251. https://doi.org/10.1007/s10772-018-9501-1
2. Hassan, F., Khan, M. S. A., Kotwal, M. R. A., & Huda, M. N. Gender independent bangia automatic speech recognition. In International Conference on Informatics, Electronics & Vision (ICIEV-2012).

3. Muslima, U., & Islam, M. B. Experimental framework for melscaled LP based Bangla speech recognition. In 2013 IEEE 16th international conference on computer and information technology (ICCIT), Khulna 2014, (pp. 56–59).
4. Pruthi, T., Saksena, S., & Das, P. K. Swaranjali: Isolated word recognition for Hindi language using VQ and HMM. In international conference on multimedia processing and systems (ICMPS), Chennai, 2000.
5. Kumar, K., & Aggarwal, R. K. Hindi speech recognition system using HTK. International Journal of Computing and Business Research, 2011, 2(2), 2229–6166.
6. Kumar, K., Aggarwal, R. K., & Jain, A. A Hindi speech recognition system for connected words using HTK. International Journal of Computational Systems Engineering, 2012, 1(1), 25–32.
7. Kurian, C., & Balakrishnan, K. Speech recognition of Malayalam numbers. In IEEE World Congress on Nature & Biologically Inspired Computing, 2009. NaBIC 2009, Coimbatore (pp. 1475–1479).
8. Bansal, P., Dev, A., & Jain, S. B. Automatic speaker identification using vector quantization. Asian Journal of Information Technology, 2007, 6(9), 938–942
9. Balleda, J., Murthy, H. A., & Nagarajan, T. Language identification from short segments of speech. In Interspeech 2000, Beijing.
10. Kumar, R., & Singh, M. Spoken isolated word recognition of Punjabi language using dynamic time warp technique. In Information systems for Indian languages. Berlin: Springer, 2011, (pp. 301–301)
11. Senoussaoui, M., Kenny, P., Dehak, N., & Dumouchel, P. An I-vector extractor suitable for speaker recognition with both microphone and telephone speech. In Odyssey, Brno 2010.
12. Dipanjan Nandi, Debadatta Pati, K. Sreenivasa Rao, Implicit processing of LP residual for language identification, Computer Speech & Language, Volume 41, 2017, Pages 68-87, ISSN 0885-2308, https://doi.org/10.1016/j.csl.2016.06.002
13. Kim, C., & Stern, R. M. (2012). Power-normalized cepstral coefficients (PNCC) for robust speech recognition. In 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 4101–4104). IEEE. https ://doi.org/10.1109/ICASSP.2012.62888 20.
14. Li, J., Deng, L., Gong, Y., & Haeb-Umbach, R. (2014). An overview of noise-robust automatic speech recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22(4), 745–777. https ://doi.org/10.1109/TASLP .2014.23046 37.
15. Mukherjee, H., Obaidullah, S. M., Santosh, K. C., Phadikar, S., & Roy, K. (2018). Line spectral frequency-based features and extreme learning machine for voice activity detection from audio signal. International Journal of Speech Technology. https ://doi.org/10.1007/s1077 2-018-9525-6.
16. Bouguelia, M.-R., Nowaczyk, S., Santosh, K. C., & Verikas, A. (2017). Agreeing to disagree: Active learning with noisy labels without crowdsourcing. International Journal of Machine Learning and Cybernetics, 9(8), 1307–1319. https ://doi.org/10.1007/s13042-017-0645-0.

17. Davis, S., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Transactions on Acoustics, Speech, and Signal Processing, 28(4), 357–366. https ://doi.org/10.1109/TASSP .1980.1163420.
18. Farooq, O., Datta, S., & Shrotriya, M. C. (2010). Wavelet sub-band based temporal features for robust Hindi phoneme recognition. International Journal of Wavelets, Multiresolution and Information Processing, 08(06), 847–859. https ://doi.org/10.1142/S0219691310003845.
19. Ganchev, T., Fakotakis, N., & Kokkinakis, G. (2005). Comparative evaluation of various MFCC implementations on the speaker verification task. In Proceedings of the SPECOM (pp. 191–194).
20. Farooq, O., & Datta, S. (2001). Mel filter-like admissible wavelet packet structure for speech recognition. IEEE Signal Processing Letters, 8(7), 196–198. https ://doi.org/10.1109/97.928676.
21. Grigoryan, A. M. (2005). Fourier transform representation by frequency-time wavelets. IEEE Transactions on Signal Processing, 53(7), 2489–2497. https ://doi.org/10.1109/TSP.2005.849180.
22. Biswas, A., Sahu, P. K., Bhowmick, A., & Chandra, M. (2014a). Feature extraction technique using ERB like wavelet sub-band periodic and aperiodic decomposition for TIMIT phoneme recognition. International Journal of Speech Technology, 17(4), 389–399. https ://doi.org/10.1007/s1077 2-014-9236-6.
23. Biswas, A., Sahu, P. K., & Chandra, M. (2016). Admissible wavelet packet sub-band based harmonic energy features using ANOVA fusion techniques for Hindi phoneme recognition. IET Signal Processing, 10(8), 902–911. https ://doi.org/10.1049/iiet-spr.2015.0488.
24. Steffen, P., Heller, P. N., Gopinath, R. A., & Burrus, C. S. (1993). Theory of regular M-band wavelet bases. IEEE Transactions on Signal Processing, 41(12), 3497–3511. https ://doi.org/10.1109/78.258088.
25. Vetterli, M., & Herley, C. (1992). Wavelets and filter banks: Theory and design. IEEE Transactions on Signal Processing, 40(9), 2207–2232. https ://doi.org/10.1109/78.15722 1.
26. Lin, T., Xu, S., Shi, Q., & Hao, P. (2006b). An algebraic construction of orthonormal M-band wavelets with perfect reconstruction. Applied Mathematics and Computation, 172(2), 717–730. https ://doi.org/10.1016/j.amc.2004.11.025.
27. Pollock, S., & Cascio, IL (2007). Non-dyadic wavelet analysis. In Optimisation, econometric and financial analysis (pp. 167–203). Berlin: Springer. https ://doi.org/10.1007/3-540-36626 -1_9.
28. Chiu, C.-C., Chuang, C.-M., & Hsu, C.-Y. (2009). Discrete wavelet transform applied on personal identity verification with ECG signal. International Journal of Wavelets, Multiresolution and Information Processing, 07(03), 341–355. https ://doi.org/10.1142/S0219 69130 90029 57.
29. Rajoub, B., Alshamali, A., & Al-Fahoum, A. S. (2002). An efficient coding algorithm for the compression of ECG signals using the wavelet transform. IEEE Trans-

actions on Biomedical Engineering, 49(4), 355–362. https ://doi.org/10.1109/10.99116 3.

30. Tabibian, S., Akbari, A., & Nasersharif, B. (2015). Speech enhancement using a wavelet thresholding method based on symmetric Kullback–Leibler divergence. Signal Processing, 106, 184–197. https ://doi.org/10.1016/J.SIGPR O.2014.06.027.
31. Zao, L., Coelho, R., & Flandrin, P. (2014). Speech enhancement with EMD and hurst-based mode selection. IEEE Transactions on Audio, Speech and Language Processing, 22(5), 899–911. https://doi.org/10.1109/TASLP .2014.23125 41.
32. Adeli, H., Zhou, Z., & Dadmehr, N. (2003). Analysis of EEG records in an epileptic patient using wavelet transform. Journal of Neuroscience Methods, 123(1), 69–87. https ://doi.org/10.1016/S0165-0270(02)00340 -0.
33. Ocak, H. (2009). Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy. Expert Systems with Applications, 36(2), 2027–2036. https ://doi.org/10.1016/J.ESWA.2007.12.065.
34. Biswas, A., Sahu, P. K., & Chandra, M. (2014b). Admissible wavelet packet features based on human inner ear frequency response for Hindi consonant recognition. Computers & Electrical Engineering,40(4), 1111–1122. https ://doi.org/10.1016/J.COMPE LECENG.2014.01.008.
35. Leggetter, C. J., & Woodland, P. C. (1995). Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. Computer Speech & Language, 9(2), 171–185.
36. Gales, M. (2000). Cluster adaptive training of hidden Markov models. IEEE transactions on speech and audio processing, 8(4), 417–428.
37. Karpov, A., et al. (2014). Large vocabulary Russian speech recognition using syntactico-statistical language modeling. Speech Communication, 56, 213–228
38. Daubechies, Ingrid. Ten lectures on wavelets. Society for industrial and applied mathematics, 1992.
39. http://www.iitg.ac.in/cseweb/tts/tts/Assamese/transliteration/Indic%20Language%20 Transliteration%20Tool%20(IT3%20to%20UTF-8)_11.html
40. http://www.iitg.ac.in/samudravijaya/tutorials/ILSL_V3.2.pdf
41. Povey, Daniel, Arnab Ghoshal, Gilles Boulianne, Lukas Burget, Ondrej Glembek, Nagendra Goel, Mirko Hannemann et al. "The Kaldi speech recognition toolkit." In IEEE 2011 workshop on automatic speech recognition and understanding, no. CONF. IEEE Signal Processing Society, 2011.
42. Thimmaraja G. Yadava, H. S. Jayanna. "A spoken query system for the agricultural commodity prices and weather information access in Kannada language", International Journal of Speech Technology, 2017
43. "Performance of Isolated and Continuous Digit Recognition System using Kaldi Toolkit", International Journal of Recent Technology and Engineering, 2019
44. Thimmaraja Yadava G, H S Jayanna. "Creation and Comparison of Language and Acoustic Models Using Kaldi for Noisy and Enhanced Speech Data", International Journal of Intelligent Systems and Applications, 2018

45. P. S. Praveen Kumar, G. Thimmaraja Yadava, H. S. Jayanna. "Continuous Kannada Speech Recognition System Under Degraded Condition", Circuits, Systems, and Signal Processing, 2019
46. Biswas, Astik, P.K. Sahu, Anirban Bhowmick, and Mahesh Chandra. "Hindi phoneme classification using Wiener filtered wavelet packet decomposed periodic and aperiodic acoustic feature", Computers & Electrical Engineering, 2015.
47. Mahadevaswamy, D J Ravi, Performance of Isolated and Continuous Digit Recognition System using Kaldi Toolkit, 2019 International Journal of Recent Technology and Engineering
48. Mahadevaswamy and D. J. Ravi, "Performance analysis of adaptive wavelet denosing by speech discrimination and thresholding," 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT), Mysuru, 2016, pp. 173-178, doi: 10.1109/ICEECCOT.2016.7955209.
49. Mahadevaswamy and D. J. Ravi, "Performance analysis of adaptive wavelet denosing by speech discrimination and thresholding," 2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques *(ICEECCOT)*, Mysuru, 2016, pp. 173-178, doi: 10.1109/ICEECCOT.2016.7955209
50. Mahadevaswamy and D. J. Ravi, "Performance Analysis of LP Residual and Correlation Coefficients based Speech Seperation Front End," 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC), Mysore, 2017, pp. 328-332, doi: 10.1109/CTCEEC.2017.8455039
51. Geethashree A, D J Ravi, “Automatic Segmentation of Kannada Speech for Emotion Conversion”, Journal of Advanced Research in Dynamical and Control Systems.
52. Geethashree A, D J Ravi, “Modification of Prosody for Emotion Conversion using Gaussian Regression Model”, 2019 International Journal of Recent Technology and Engineering
53. Geethashree A., Ravi D.J. (2018) Kannada Emotional Speech Database: Design, Development and Evaluation. In: Guru D., Vasudev T., Chethan H., Kumar Y. (eds) Proceedings of International Conference on Cognition and Recognition. Lecture Notes in Networks and Systems, vol 14. Springer, Singapore.
54. Basavaiah, J., & Patil, C. M. (2020). Human activity detection and action recognition in videos using convolutional neural networks. Journal of Information and Communication Technology, 19(2), 157-183.
55. Basavaiah, J., & Anthony, A. A. (2020). Tomato Leaf Disease Classification using Multiple Feature Extraction Techniques. Wireless Personal Communications. doi:10.1007/s11277-020-07590-x