International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 5 | ISSN : 2456-3307

Speech Recognition System for Different Kannada Dialects

Hemakumar G[1], Punithavalli M[2], Thippeswamy K[3]

[1]Department of Computer Science, Govt. College for Women (Autonomous), Mandya, Karnataka, India
[2]Department of Computer Application, Bharathiar University, Coimbatore, Tamil Nadu, India
[3] Department of Computer Science & Eng., Visvesvaraya Technology University, Post-Graduation Centre, Mysore, Karnataka, India.

ABSTRACT

In this paper discuss on pronunciation variations occurs in different Kannada dialects, designing language model, building acoustic models, finally recognition of Kannada dialect speech. Algorithm designed for recognition of isolated Kannada word and continuous Kannada speech made by different dialects speakers. The novelty of algorithm is in handling multiple Kannada dialects speaker's speech recorded by mini-microphone, headphone and cell phones. Robustness of the algorithm in handling different Kannada dialects speech and handling little noisy waves. Here speech waves recorded at natural environment. Here classification of speech models based on speaker's dialects and inside the dialects sub classes designed according to acoustic features. During recognition, breadth first matching technique and then inside that dialect class depth first matching techniques implemented. Here speech recognition designed using MFCC and coefficients of real cepstrum features and compared the performance. In these experiment real cepstrum coefficients, features produced better recognition rate while dealing with multiple dialects of same language. All computations made using mat lab.

Keywords: Speech recognition, Kannada dialect, Language model, Normal parameters and Speech enhancement.

I. INTRODUCTION

Speech is the primary way for human beings to communicate. Therefore, it is only natural to use speech as the primary method to input information into computational devices. Automatic speech recognition (ASR) is the technique where human speech can use has input for computational devices. So ASR systems shall be designed for general using systems, hence it need to support multiple speakers speech and capable to adapt all the speakers variations. The variations may occur in speech styles, pitch and anatomy that make each speaker unique. In addition, things like background noise, utterances, and dialects can negatively affect the interpretation of speech. Even words that sound alike can create problems for ASR systems. In this paper, for the experiments, consider the multiple speaker of different Kannada dialects and they are native speakers of that particular dialects and non-native of those dialects.

In paper [1] discuss on some variations within the speech signal which make the ASR task difficult. According to them, speech variation occurs by foreign and regional accents, speaker characteristics, speaking rate and style, age of speaker and speaker's emotions will reduces the performances of speech recognizer. In paper [2], discuss the pronunciation variants for handling non-native speakers' speech variability. To recognition of non-native speech, they used the adaptation of models, French phonemes on foreign speech data and compared multiple foreign accents. In paper [3] discuss the problem in inter-speaker variability. Here they have mentioned relations between speech production, perception and auditory processing. They followed the approach of universal warping to normalize the acoustic representations of speaker which may leads in unify many of the vowel normalization approaches. Paper [4] discuss about Kullback–Leibler GMM (Gaussian mixture model) algorithms, which can improve the dialect classification for text-independent spontaneous speech in Chinese, Arabic and Spanish languages. Paper [5] designed speaker-adaptive ASR on large collection of spoken records done by Czech radio. Here they integrated the speaker diarization and adaptation methods to achieve a low real-time factor (RTF). In paper [6], discuss the problems occur in ASR performance when moving from SD (speaker-dependent) to SI (speaker-independent) conditions for connectionist hidden markov model (HMM) and artificial neural network (ANN) systems in the context of LVCSR (large vocabulary continuous speech recognition).

The enhancement of speech signal is required in improving the intelligibility and overall perceptual quality of degraded speech signal using audio signal processing techniques. The enhancement of speech signal is required due to the signal degraded by noise or reduction of noise in signal. Those noises are like Gaussian white noise, pink, red and gray noises occur during recording time. Then noise which can occur by multi speakers during recording time is most challenging task to handle. Those noises shall remove before speech signal segmentation and feature extraction. Paper [7], discuss about speech signal enhancement using short-time discrete Fourier transform domains, which is corrupted by noise. In paper [8], discuss on filters, which allow explicit control of the trade-off between noise reduction and speech distortion via the chosen rank of the signal subspace. Paper [9], proposed the Bayesian STSA (short-time spectral amplitude) speech enhancement algorithms under a stochastic deterministic speech models which makes provision for the inclusion of a priori information by considering a non-zero mean. In paper [10] focused on reduction of the noise effects on speech recognition (SR) systems. Here in the first stage, normalization of clean speech and noisy speech is done using the cepstral time series. In the second stage, the contrast normalization of temporal modulation spectrum is done in order to reduce the artifacts due to noise, while preserving the information in the speech modulation events (edges). In paper [11], proposed the DOLPHIN (dominance based locational and power-spectral characteristics integration), which integrates the SCA (spatial clustering approach) and a FMA (factorial model approach) for enhancement. Paper [12], discuss regarding the measurement of enhancement considered a wide range of distortions introduced by four types of real-world noise at two signal-to-noise ratio levels by four classes of speech enhancement algorithms namely spectral subtractive, subspace, statistical based model, and Wiener algorithm.

Paper [13], explained on the paradigm of statistic in speech recognition for phonetic and phonological knowledge sources. They discuss on computational phonology and rigorous mathematical models like Bayesian analysis, non-stationary time series, statistical estimation theory, dynamic system theory and nonlinear function approximation (neural network) theory. Paper [14] discusses on discriminative approach in lexicon optimization, which directly contributes error reduction in speech recognition by considering not only linguistic constraints problems but also acoustic–phonetic confusability.

The Kannada language dialect classification mainly based on the social stratum and geographical regions. Paper [15], discuss on social (caste based) dialects of Kannada language. That is high caste dialect (the Brahmi), middle caste dialect (the non-Brahmin) and low caste dialect (the Harijan). Paper [16, 17], made a study on Kannada language and culture of Karnataka and clearly discussed about Kannada language, culture of Karnataka state. In Kannada language identified as 20 dialects. In the proposed model considered geographical regional (Karnataka state) dialects majorly classified into four group's namely southern dialect, coastal dialect, northern dialect and eastern dialects. Here selected native speakers from all four regional dialects speaker uttering word and sentence to design the Kannada dialects speech signal database for the purpose of training and testing. Here standard Kannada style means the language mainly used by Medias and educated peoples. In the present situation the Mysuru-Bengaluru region Kannada speaking style is considered has standard Kannada dialect by all Medias, educated people and in administration section. Nevertheless, in the writing of Kannada language by all the sector or dialect speakers are using same structure of Kannada script and style of grammar.

This paper discuss on difference between the accent and dialect, then how the syllables are changes during the pronunciation of words by different Kannada dialect speakers and finally designing the automatic Kannada speech recognition (AKSR) for different dialect, here discussed on designing language model for automatic speech recognition, designing the acoustic model and recognition.

In the area of speech recognition, MFCC and LPCC features are most commonly using techniques has mentioned in survey paper [18]. In paper [21], discusses that cepstral features derived from the DPS (differential power spectrum) will improve the robustness of a speech recognizer in presence of noise at background. These robust features are computed from the speech signal of a given voiced/unvoiced part frames through the following four steps. Firstly, the short-time power spectrum of speech signal, computed through the fast Fourier transform algorithm. Secondly, DPS obtained by differentiating the power spectrum with respect to frequency of that frames. Thirdly, the magnitude of DPS is projected from linear frequency to the mel scale and smoothed by a filter bank. Finally, the outputs of the filter bank, transformed to cepstral coefficients by the discrete cosine transform after a nonlinear transformation. Here ASR designed by extracting features from MFCC, Coefficients of Real Cepstrum, and then testing done for different Kannada dialects speech, and comparative results are recorded. For same Kannada speech database the AKSR is designed using 3-state continuous Hidden Markov Model (HMM), Gaussian Mixture Model (GMM), Vector Quantization-Hidden Markov Model (VQ-HMM), and Learning Vector Quantization (LVQ) network techniques. Then proposed model compared with those AKSR systems. In this paper, used coefficients of real cepstrum features to train and test all AKSR models designed by ourselves for same speech data. These speech recognition systems will recognize only the trained set of Kannada isolated words and continuous Kannada dialects speech by the native speaker's dialects like southern region of Karnataka dialect, coastal region of Karnataka dialect, northern region of Karnataka dialect, eastern region of Karnataka dialect and non-native Kannada speaker's style. The speech models designed for Kannada dialects speaking at South-eastern region of Karnataka.

This paper organized into six different sections; in section first discuss about introduction and literature survey. Section second deals with the pronunciation variations occurs at different dialects related to Kannada language. In section, three discuss the methodology to designed language model for speech recognition. Section 4 gives information on speech enhancement and acoustic model building for different dialect Kannada speech. Section 5 discuss on automatic Kannada dialects speech recognition for native speakers, non-native Kannada speakers. Section 6 gives the details of experimental results and discussions. Section seven deals with conclusion.

II. PRONUCIATION VARIATIONS

An accent can be defined as the phonetic (In Kannada language the phonemes and syllables are same) traits of an individual's native language carried over into a second non-native language. A dialect is a variation of a same language spoken in a particular community or in a particular geographical region. An accent is the way that individual persons or group of people makes sound while pronouncing the words. For example in English, the word 'b***a***throom' is pronounced by some group of peoples using short vowel /a/ and other group of peoples pronounce using long vowel /A/, this is due to the person's accent. The accent can be learnt by the speakers and pronounce as it utter by particular person or sections of people. In Kannada language each phoneme, are bounded with unit of fixed amount of time. In Kannada language, the speakers should careful, while uttering any phoneme sounds, namely short vowels (a, i, e, o, u), long vowels (A, I, E, O, U), unaspirated plosive (p, b, t, d, k, g, T, D) sounds, aspirated plosive (ph, bh, th, dh, kh, gh, Th, Dh) sounds, unaspirated affricates (c, j), aspirated affricates (ch, jh) sounds, fricatives (h, s, sh, Sh) sounds and Nasals (m, n, N, nY, nG) sounds. In Kannada language, any interchange of phonemes leads in change of meaning of word and function of word in sentences.

Kannada dialect describes both person's accent as well as the way of grammatical features used by the speaker and vocabulary of speaker. The dialect cannot be learnt easily and speak by other native language speakers without knowing the taste of that particular regional dialect. In table 1, few Kannada example words are listed to show how the syllables (phonemes) are changing during pronunciation made by each native dialect. It also gives the information of how the words uttered in standard style of Kannada and other four major regional dialects. Here standard dialect means the Kannada language mainly used by medias and educated society. The standard Kannada dialect is most popular in Mysuru-Bengaluru region and comparison of other major dialect classified according to Karnataka state geographical regional into eastern, northern, coastal and south parts. These changes occur during the speaking (pronouncing) not in writing form. In Kannada language the written form is more or less same in all the dialects, the changes is only during pronouncing the words or natural speaking.

TABLE I

SHOWING THE PRONUNCIATION VARIATIONS OCCUR WHILE UTTERING THE SAME KANNADA WORDS BY DIFFERENT KANNADA DIALECT NATIVE SPEAKER'S.

Standard Dialect Style	Eastern dialect style	Northern Dialect Style	Coastal Dialect Style	South Dialect Style
/hakki/	/akki/	/həkki/	/(h) əkki/	/akki/
/haNNu/	/aNNu/	/həNNə/	/(h) ənnu/	/aNNu/
/hUvu/	/Uva/	/hUvə/	/(h)Ugu/	/U/
/viSha/	/isa/	/isa/	/vica/	/isa/
/varSha/	/varsa/	/varsə/	/varca/	/varsa/
/shAyi/	-	/sAyi/	/cAyi/	/syAyi/
/AkAsha/	/AkAsa/	/AkAsa/	/AkAca/	-
/vEsha/	/Esa/	/Ese/	/vɛ:ca/	/Esa/
/sante /	/sante /	/santi/	/cante/	/santé/
/siTTu/	/siTTu/	/siTTə/	/ciTTu/	/siTTu/
/sAsive/	/sAsve/	/sAsi/	/cAcəmi/	/sAsve/
/kaDale/	/kaDLe/	/kaDLi/	/kaDle/	/kaLLe/
/beraLu/	/baLLu/	/bəLLə/	/bellu/	/beLLu/
/aDike/	/aDke/	/aDki/	/əDike/	/aDke/
/oLage/	/vaLge/	/vaLgə/	/ɔlage/	/vaLge/
/takkaDi/	/takkaDi/	/takkDi/	/takkəDi/	/takkəDi/
/kattale/	/kattle/	/kattla/	/kattəle/	/kattle/
/kuruDa/	/kuLDa/	/kuDDa/	/kuyDa/	/kuNDa/
/eraDu/	/yaLDu/	/yaDDu/	/eyDu/	/EDu/
/karaLu/	/koLLu/	/kəLLə/	/kəllu/	/kaLLu/
/siDilu/	/siDLu/	/siDLə/	/ceDlu/	/siLLu/
/saDilu/	/soDLa/	/səDLə/	/səDilu/	/saLLu/
/mogge/	/magge/	/məggi/	/mɔke/	/moggu/
/onake/	/vaNke/	/vaNki/	/ɔnəke/	/vanke/
/eNTu/	/eNTu/	/yaNTu/	/enTu/	/eNTu/
/ghaNTe/	/gaNTe/	/ghəNTi/	/ganTe/	/gaNTe/

III. Building Language Model

Language model is required in recognition of the correct possibility sequences of the syllables to form words and correct words sequences possibilities in formation of sentences, hence designed Kannada language model. Here to design the Kannada language model, randomly selected 10,000 Kannada sentences of different word length (minimum lengths of sentences consists of 2 words and maximum length of sentences consists of 45 words) from Hampi text corpus. While designing the statistical language model, first computed the text to word frequency for the Kannada text data, and then built the unique word list, which considered has vocabulary of the language model and stored in the language model. Then computed the individual words sequences from the text data. Here computed the bigram and trigram models as shown in fig 2 and stored in the language model. So final statistical language model has containing list of unique words, bigram, trigram of words and probability of the sequences of words in which it can occur is computed. Here consecutive 2 and 3 word sequences probability is computed and stored.

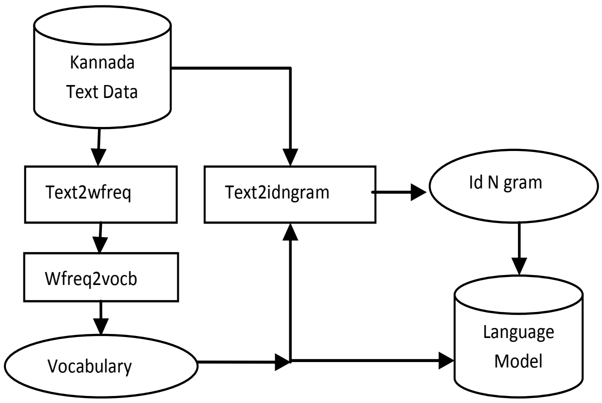


Figure.1 Show the methodology to design the Language model. In language model all Kannada language scripts are written and stored in Roman letters to reduce the memory size of corpus.

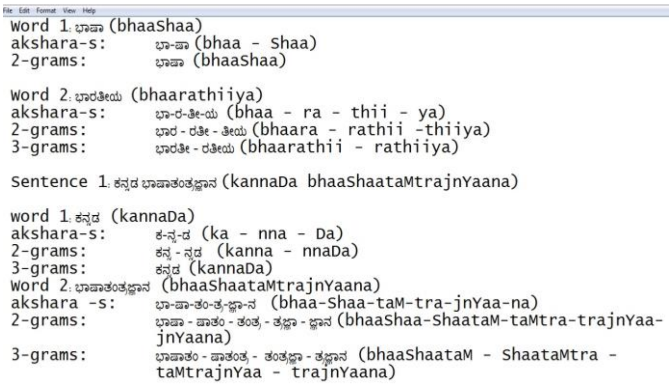
In language model probability for two word sequences and three word sequences is computed and stored in language model which is consider has bi-gram (bi-word) and tri-gram (tri-word) models respectively. To illustrate how the probability computed, taken one Kannada sentence has example and shown how computation of the language model done during the training or testing phases (Kannada language sentence written in roman letters).

Ex: - "idannu saripaDisuvudu tuMbaa agatyavaagide"

From the above mentioned Kannada example sentence there are of 4 words, which can be consider has KW_1 = "idannu", KW_2 = "saripaDisuvudu", KW_3 = "tuMbaa", and KW_4 = "agatyavaagide". Unigram is computed for each word by computing $P(KW_1 \mid \text{null or white space})$, $P(KW_2 \mid KW_1)$, $P(KW_3 \mid KW_2)$ and $P(KW_4 \mid KW_3)$. Bi-grams is computed $P(KW_3 \mid KW_1, KW_2)$ and $P(KW_4 \mid KW_2, KW_3)$. Tri-grams are computed by $P(KW_4 \mid KW_1, KW_2, KW_3)$. Where P is a probability, which computed by using methods

$$P_{MLE}(kw_n|kw_1,\dots,kw_{n-1}) = \frac{Count(kw_1,\dots,kw_n)}{Count(kw_1,\dots,kw_{n-1})} \qquad equ\ (3.1)$$

$$P_{MLE}(tuMbaa \mid \text{"idannu"}, \quad \text{"}saripaDisuvdu\text{"}) = \frac{Count(\text{"idannu"}, \text{"saripaDisuvdu"}, \text{"}tuMbaa\text{"})}{Count(\text{"idannu"}, \text{"}saripaDisuvdu\text{"})} \quad equ(3.2)$$



```
Word 1: ಭಾಷಾ (bhaaShaa)
akshara-s:      ಭಾ-ಷಾ (bhaa - Shaa)
2-grams:        ಭಾಷಾ (bhaaShaa)

Word 2: ಭಾರತೀಯ (bhaarathiiya)
akshara-s:      ಭಾ-ರ-ತೀ-ಯ (bhaa - ra - thii - ya)
2-grams:        ಭಾರ - ರತೀ - ತೀಯ (bhaara - rathii -thiiya)
3-grams:        ಭಾರತೀ - ರತೀಯ (bhaarathii - rathiiya)

Sentence 1: ಕನ್ನಡ ಭಾಷಾತಂತ್ರಜ್ಞಾನ (kannaDa bhaaShaataMtrajnYaana)

word 1: ಕನ್ನಡ (kannaDa)
akshara-s:      ಕ-ನ್ನ-ಡ (ka - nna - Da)
2-grams:        ಕನ್ನ - ನ್ನಡ (kanna - nnaDa)
3-grams:        ಕನ್ನಡ (kannaDa)
Word 2: ಭಾಷಾತಂತ್ರಜ್ಞಾನ (bhaaShaataMtrajnYaana)
akshara -s:     ಭಾ-ಷಾ-ತಂ-ತ್ರ-ಜ್ಞಾ-ನ (bhaa-Shaa-taM-tra-jnYaa-na)
2-grams:        ಭಾಷಾ - ಷಾತಂ - ತಂತ್ರ - ತ್ರಜ್ಞಾ - ಜ್ಞಾನ (bhaaShaa-ShaataM-taMtra-trajnYaa-
                jnYaana)
3-grams:        ಭಾಷಾತಂ - ಷಾತಂತ್ರ - ತಂತ್ರಜ್ಞಾ - ತ್ರಜ್ಞಾನ (bhaaShaataM - ShaataMtra -
                taMtrajnYaa - trajnYaana)
```

Figure 2. Show the Kannada word and sentence split into akshara (syllable), bi-grams and tri-grams form the words. This is the way computed the sequence of akshara (syllable) occurs in the words.

IV. BUILDING ACOUSTIC MODEL

The proposed model works in offline mode. Here all speech wave files are recorded and stored in folders. Then one by one, wave files passed for the training. The stored wave files are recorded through single channel mini-microphone, headphone and cell-phone, and the sampling rates are not fixed during recording time (Intentionally recorded). Due to this reason while converting the analog signal into digital signal, first check for sampling rate of signal and then resample the signals into 16 KHz, 16-bit mono PCM (pulse code modulation). Then digital values are stored in single dimension array. The DC components removed in those digital values. Then wave is made to pass through low order digital system in order to spectrally flatten the signal and to prevent the precision effects from affecting the analysis. Then go for noise removal and enhancement of speech wave quality, apply the GMSK (Gaussian minimum shift keying) filtering technique. The designed low pass FIR Gaussian pulse shaping filter, which returns the filter coefficients as output vector.

Then next step is to detect the voiced, unvoiced and silence region in the wave (outputted vector data). Paper [19] discusses the problem in windowing for the frequency analysis in short time. The human hearing is relatively insensitive to short time phase distortion of the speech wave. There is also no apparent reason for the using of symmetric windows, which gives a linear phase response. Here the framing for every 20 millisecond with an overlapping of 7 milliseconds is done to find the voice, unvoiced and silence region in signal. Then using same size of frame, hamming window is applied. The frame magnitude and short time energy is computed. Then compute dynamic threshold for each frame by using the time domain and frequency domain of the wave [20]. Using that threshold frame verified whether it is silence, voiced or unvoiced region. If it is silence region, that frame is dropped in this stage only and moved to next frame to check whether it consist of voice or unvoiced speech region.

Next stage is feature extraction stage from voiced and unvoiced speech wave region. Here LPC method is used to compute LPC coefficients. Then real cepstrum coefficients are computed from those LPC Coefficients. The real cepstrum of a wave RC(n) is calculated by determining the natural logarithm of magnitude of the Fourier transform of RC(n), then obtaining the resulting sequence through inverse Fourier transform. The returned sequence is a real-valued vector of size equal to the input vector; here the input order of C is trailed for 16, hence for each voiced/unvoiced frame the outputted feature vector size will be 16. The coefficients of real cepstrum are in dimension of C*f matrices for each syllable / sub-word. Here f is number of frame occurred in that syllable or sub-word. This matrix will be passed into k-means algorithm by keeping k=3 and outputted values are passed into 3 state Baum–Welch algorithm and each syllable or sub-word is trained. Then re-estimation of Baum-Welch parameters is done.

The estimated parameter $P(O \mid \lambda)$ is optimized. The standard Lagrange optimization is setup using Lagrange multipliers; then P is maximized by [22, 23]

$$\pi_i = \frac{\pi_i(\partial P/\partial \pi_i)}{\sum_{k=1 \ldots N} \pi_k(\partial P/\partial \pi_k)} \qquad equ\ (4.1)$$

$$T_{ij} = \frac{T_{ij}(\partial L/\partial T_{ij})}{\sum_{k=1 \ldots N} T_{ik}(\partial L/\partial T_{ik})} \qquad equ\ (4.2)$$

$$O_j(k) = \frac{O_j(k)(\partial P/\partial O_j(k))}{\sum_{f=1 \ldots M} O_j(l)(\partial L/\partial O_j(l))} \qquad equ\ (4.3)$$

Then normal-fit is applied for 2 consecutive HMM parameter $\lambda(T, O, \pi)$ and Normal parameters are computed. Her the trained two consecutive $\lambda(T, O, \pi)$ are considered has sample data. So, there will be a samples data from $x_1 \ldots x_n$ for this a normal parameter $N(\widehat{\mu}, \widehat{\sigma^2})$ is computed by using the

$$\hat{\mu} = \bar{d} \equiv \frac{1}{n}\sum_{i=1}^{n} d_i \qquad \text{equ(4.4)}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2 \qquad \text{equ (4.5)}$$

The labeled $\hat{\mu}$ and $\widehat{\sigma^2}$ value will be classified according to acoustic classes and then stored. Those data are stored in acoustic model has representatives of syllables or sub-words in that particular dialect classes. In language model designed bi-syllable language model for each isolated word and tri-syllable language model for continuous speech. Here created the corresponding mapping between acoustic values and syllables kept in language model and then data is stored in acoustic-syllable model table.

V. Testing / Recognition

In this phase, first testing speech wave is converted into digital values by resample into 16 KHz, 16-bit PCM. Then computed for the silence, voiced / unvoiced region and passed for feature extraction. Here real cepstrum coefficients features extracted from the each voiced and unvoiced frame. Then HMM parameters are computed, then for 2 and 3 consecutive HMM parameters values are combined and then normal fit parameters are computed. The normal fit values $\boldsymbol{\hat{\mu}}$ and $\boldsymbol{\widehat{\sigma^2}}$ are matched with trained set of normal fit values by keeping threshold values in each dialect classes using breadth first matching technique. When dialect class found, then again used depth first matching technique to match the consecutive sequence values in each acoustic sub-classes. Then computed probability of bi-syllable and tri-syllable language models are used for the recognition of isolated word and continuous Kannada dialect speech respectively. Then sentence recognition is done by concatenation of tri-syllable language mode, then decision is taken and outputted the top ranked matched word / sentence has recognition.

VI. Experimental Results and Discussion

To design speech recognition for different Kannada dialects there were no Kannada speech corpus, hence needed to design ourselves Kannada speech database for isolated word and continuous speech. While designing Kannada speech waves database for training set, here selected the speakers from 5 districts namely Mysuru, Bengaluru, Mandya, Chamarajnagar and Ramanagar districts lies in south-eastern part of Karnataka state. Here selected the native Kannada dialect speakers and non-native Kannada speakers. Then designed Kannada speech recognition for different dialect.

In this experiment showed that how the syllables (Phonemes) will changes during uttering the words by different dialect speakers. In the table 1 with few Kannada word examples illustrated the changes of sounds by all major Kannada dialects. Here clearly mentioned the difference between dialect to dialect of Kannada languages. In Kannada language, uttering the same word by different geographical region people will make some changes in syllables (phonemes) during pronouncing the same words and while making sentences they will also use different word phrases, this becomes dialects for Kannada languages.

In this paper experimentation are done on recognition of isolated Kannada words uttered by different dialect speaker’s using 3-state continuous HMM, GMM, VQ-HMM, LVQ network methods and compared with our proposed model for the same speech database. To experiment programs are written in mat lab and run on Intel Core i5 with 2.30 GHz processor speed and 3 GB of RAM (Here observed that only 20% of CPU utilization during the time of running the program). In table 2 showed the comparisons of MFCC and Real Cepstrum Coefficient features for the different dialect Kannada word recognition. Here the speech models are built for south-eastern Kannada dialects and compared for four native Kannada dialect speakers’ speech, then average accuracy rate is taken. In this experiment, observed that features extracted by real cepstrum coefficient worked better than MFCC features while dealing with different dialects speech waves. Here accuracy rate is measured for recognition of 294 unique words uttering by above mentioned native Kannada dialects speakers.

The table 3 shows the word recognition rate for others dialects. Here the accuracy rate is measured and compared by designing ASR using 3-state continuous HMM, GMM, VQ-HMM, LVQ Network methods and proposed model. It clearly shows that proposed model gives better results compare to other models. The average recognition of Kannada dialect showed 81.94% accuracy rate. In this experiment main goal was only to recognize the isolated Kannada word from the differnt speaker dialects. Table 4 shows the continuous Kannada speech recognition for different dialects. Here the accuracy rate is better than isolated word recognition because in continuous speech tri-syllable language models have help to increase in the accuracy rate.

TABLE II
COMPARISONS OF SPEECH RECOGNITION DESIGNED USING MFCC AND REAL CEPSTRUM COEFFICIENT FEATURES VALUES FOR DIFFERENT KANNADA DIALECTS SPEECH.

Speaker Dialects	Mel Frequency Cepstrum Coefficient	Real Cepstrum Coefficient
Southern Region	80.56%	81.75%
Coastal Region	82.33%	84.45%
Northern Region	85.91%	85.95%
Eastern Region	73.83%	75.64%

TABLE III
COMPARISON OF ISOLATED WORD RECOGNITION RATE FOR MAJOR KANNADA DIALECTS USING DIFFERENT TECHNIQUES.

Methods	South region	Coastal region	Northern Region	Eastern region
Proposed Model	81.75 %	84.45 %	85.95%	75.64%
HMM	76.75 %	78.06 %	75.20%	71.04%
GMM	78.72 %	79.27 %	77.97%	70.79%
VQ-HMM	75.30 %	65.63 %	67.97%	64.31%
LVQ Network	80.43 %	80.98 %	83.34%	73.12%

TABLE IV
CONTINUOUS SPEECH RECOGNITION RATE FOR MAJOR KANNADA DIALECTS

Methods	South region	Coastal region	Northern Region	Eastern region
Proposed Model	82.75%	85.45%	83.95%	75.64%
HMM	76.75%	79.06%	76.2%	75.04%
GMM	76.72%	77.27%	75.97%	71.79%
VQ-HMM	76.30%	70.03%	75.73%	75.1%
LVQ Network	79.12%	81.99%	79.47%	75.32%

This experiment shows that one speech model can be design to recognize the words uttered by any dialects speaker. In this paper, designed ASR model is experimented for recognizing the isolated Kannada words and continuous Kannada speech of different dialects. The proposed ASR model gives better accuracy rate compare to ASR designed using 3-state continuous HMM, GMM, VQ-HMM and LVQ Network techniques. All models are trained and tested for same Kannada speech database.

VII. CONCLUSION

In this experiment, it clearly shown that features extracted using Real Cepstrum Coefficients provide better accuracy rate compared to MFCC features while dealing with different dialects speech of same language. When classification of speech models are done according to dialects and inside each dialect classes based on acoustic models. Then combination of breath first and depth-first matching technique leads in fast matching in different dialect acoustic classes with better accuracy rate.

The designed ASR is capable to recognize all sort of dialect then it can be used in application level software like telephone dialing systems, data entry, filling online and offline forms, command and control for robotics, voice enabled web browsing, dictation machine and in automated machine translation.

VIII. ACKNOWLEDGMENT

Authors would like to thank Department of Kannada, Government College for Women, Mandya, for their cooperation and guiding Kannada dialect by giving the examples of Kannada words and analysing the syllables changing in words spoken by different dialect speakers.

Authors also thank for all speakers friends who have helped in designing Kannada dialect speech data. Authors would like thank all our friends, reviewers and editor board to preparing this paper.

IX. REFERENCES

[1]. Benzeguiba M et al., "Automatic Speech Recognition and Intrinsic Speech Variation", Proc of IEEE ICASSP, pp. 1021-1024, 2006.

[2]. Katarina Bartkov A, "On using units trained on foreign data for improved multiple accent speech recognition", Speech Communication, Vol. 49, Issues 10–11, pp. 836-846, Oct–Nov 2007.

[3]. Umesh S, "Studies on inter-speaker variability in speech and its application in automatic speech recognition", Indian Academy of Sciences, Sadhana Vol. 36, Part 5, pp. 853–883, October 2011.

[4]. Yun Lei and John H. L. Hansen, "Dialect Classification via Text-Independent Training and Testing for Arabic, Spanish, and Chinese", IEEE Transactions On Audio, Speech, And Language Processing, Vol. 19, No. 1, pp. 85-96, Jan 2011.

[5]. Petr Cerva et al., "Speaker-adaptive speech recognition using speaker diarization for improved transcription of large spoken archives", Speech Communication, vol 55, Issue 10, pp. 033-046, Nov–Dec 2013.

[6]. Sabato Marco Siniscalchi et al, "Hermitian Polynomial for Speaker Adaptation of Connectionist Speech Recognition Systems", IEEE Transactions on Audio, Speech, And Language Processing, Vol. 21, No. 10, pp. 2151-2161, October 2013.

[7]. Martin Krawczyk and Timo Gerkmann, "STFT Phase Reconstruction in Voiced Speech for an Improved Single-Channel Speech Enhancement", IEEE/ACM Transactions on Audio, Speech and Language Processing, Vol. 22, No. 12, pp. 1931-1940, Dec 2014.

[8]. Jesper Rindom Jensen et al., "A Class of Optimal Rectangular Filtering Matrices for Single-Channel Signal Enhancement in the Time Domain", IEEE Transactions on Audio, Speech and Language Processing, Vol. 21, No. 12, pp. 2595-2606, Dec 2013.

[9]. Matthew McCallum et al., "Stochastic-Deterministic MMSE STFT Speech Enhancement with General A Priori Information", IEEE Transactions on Audio, Speech and Language Processing, Vol. 21, No. 7, pp. 1445-1457, July 2013.

[10]. X. Lu et al., "Temporal contrast normalization and edge-preserved smoothing of temporal modulation structures of speech for robust speech recognition", Speech Communication 52 (2010) page no. 1–11.

[11]. Tomohiro Nakatani et al., "Dominance Based Integration of Spatial and Spectral features for Speech Enhancement", IEEE Transactions on Audio, Speech and Language Processing, Vol. 21, No. 12, pp. 2516-2531, Dec 2013.

[12]. Yi Hu and Philipos C. Loizou, "Evaluation of Objective Quality Measures for Speech Enhancement", IEEE Transactions on Audio, Speech and Language Processing, Vol. 16, No. 1, pp. 229-238, Jan 2008.

[13]. Li Deng (1998), "A dynamic feature-based approach to the interface between phonology and phonetics for speech modeling and recognition", Speech Communication 24, pp. 299-323, 1998.

[14]. Mijit Ablimit, Tatsuya Kawahara, Askar Hamdulla, "Lexicon optimization based on discriminative learning for automatic speech recognition of agglutinative language", Speech Communication, vol 60, pp. 78-87, May 2014.

[15]. Bright W (2014), "Social Dialect and Language History". Available https://is.muni.cz/el/1421/jaro2014/AJ22089/Social_Dialect_and_Language_History.pdf

[16]. Jennifer Torres et al., "Kannada Manual: Language and Culture", Texas State University class of 2012.

[17]. Nag, S., "Early reading in Kannada the pace of acquisition of orthographic knowledge and phonemic awareness", Journal of Research in Reading, 30(1), pp. 7-22, 2007.

[18]. Hemakumar G. and Punitha P., "Speech Recognition Technology: A Survey on Indian Languages", International Journal of Information Science and Intelligent System, Vol. 2, No.4, 2013, pp. 1-38.

[19]. Robert Rozman and Dusan M. Kodek, "Using asymmetric windows in automatic speech recognition", Speech Communication 49, pp. 268–276, 2007.

[20]. Hemakumar G. and Punitha P., "Automatic Segmentation of Kannada Speech Signal into Syllables and Sub-words: Noised and Noiseless

Signals", International Journal of Scientific & Engineering Research, Volume 5, Issue 1, pp. 1707-1711, Jan-2014.

[21]. Jingdong Chen et al (2003), "Cepstrum derived from differentiated power spectrum for robust speech recognition", Speech Communication 41 (2003), page no 469–484.

[22]. Hemakumar G. and Punitha P., "Large Vocabulary Speech Recognition: Speaker Dependent and Speaker Independent", Proc of INDIA-2015, Published by Springer- Advances in Intelligent and Soft Computing, Vol. 339, pp 73-80.

[23]. Lalit R.Bahl, et.al, "Estimating Hidden Markov Model Parameters so as to maximize speech recognition Accuracy", IEEE Transactions on Audio, Speech and Language processing vol.1, no.1, 1993.

[24]. Satya Dharanipragada, et.al, "Gaussian mixture models with covariance's or Precisions in shared multiple subspaces", IEEE Transactions on Audio, Speech and Language Processing, vol.14, no.4, 2006.

[25]. Thangarajan R., Natarajan A. M. and Selvam M. "Syllable modeling in continuous speech recognition for Tamil language", International Journal for Speech Technology, vol. 12, pp.47 - 57, 2009.