Recognition of Isolated Telugu Words with Soft and Hard Sounds using Hidden Markov Models

Shaik Shafee
Dept.of. ECE ,SVUCE,Tirupati

B.Anuradha
Dept.of. ECE ,SVUCE,Tirupati

ABSTRACT
In this paper an attempt has been made to recognize isolated Telugu words consisting of hard and soft sounds using Hidden Markov Models (HMM). In Telugu language ,the alphabets ka (క), cha (చ), ta (ట), tha (త), pa(ప) are called hard(paruSha- పరుష) and ga(గ), ja(జ), da(డ), dha(ద), ba(బ) are called soft (saraLa- సరళ) consonants [1]. While recognizing such sounds, the chance of recognizing hard sounds as soft and soft sounds as hard is more as they are pronounced with almost equal structure of vocal activity. For Example, T (త), may be recognized as D (ద), and P (ప) as B (బ). In case of Telugu word Thammudu means younger brother may be recognized as Dammudu which has different meaning. Different words of this kind being applied to HMM and analyze the success/error ratio using HMMs [2,4].

Keywords
Telugu Speech, Isolated word Recognition, Hard and Soft Sounds in Telugu, Vector Quantization, HMM, and MATLAB.

1. INTRODUCTION
ASR (Automatic Speech Recognition) is the process of recognizing human speech. Recognition may be of isolated word to continuous speech recognition, speaker-dependent to speaker-independent recognition, and from a small vocabulary to a large vocabulary [3]. In this paper we have chosen isolated word recognition on a small vocabulary of Telugu words with Hard and Soft Sounds. Each Telugu word is modeled by a distinct HMM and the unknown recognized word among this small vocabulary applied to HMM Recognition system and analyze the recognition success/error rate for different count of centroid HMM models.

2. HIDDEN MARKOV MODELS

2.1 Model Parameters
HMM is a stochastic approach tool representing probability distributions over observation sequence. This approach works on two properties : a) The observation vectors 'Ot' at time 't' was generated by some process whose State 'St' is hidden from the observer and ,(b) The State at 'St' dependent only on State 'St-1' and independent of all prior states . There are three fundamental problems arise in speech recognition using HMM namely (1) The probability of observation Sequence ({O1,O2,O3,…….,OT}) for the given HMM model (λ ={A,B, π}) that is P(O/ λ), (2) Determination of single best state sequence given the Observation sequence O, and the Model (λ ={A,B, π}) (3) And adjustment of HMM model (λ ={A,B, π}) to maximize recognition probability. 'A' is said to be State transition matrix, 'B' is said to be Observation symbol probability matrix, and 'π' is the initial state distribution vector. Isolated word recognition process includes three major steps namely Framing of HMM model to each isolated word, Train the Collected Isolated words, and finally recognize the unknown word by Maximum likelihood of HMM models [2, 3, and 4].

2.2 Feature Extraction and Vector Quantization
All the isolated word speech samples which are to be recognized will be collected through different speakers. Then the features being extracted from the collected word speech sequences. Generally Feature Extraction means collecting of LPC parameters and Cepstral coefficients [3]. Each of the sampled Speech signal is blocked into frames of N samples (If speech signal collected over 16 Khz then block of 20 ms i.e N=320 sample in each frame) and consecutive frames are spaced by 'deltaN=80 samples' apart. That means a word of 1.6 secs have around 200 frames with each frame ends at samples 320,400, 480,…etc..Then each frame is multiplied with hamming window of size 'N' and LPC parameters computed for each frame by using Levinson-Durbin recursion. Generally LPC order will be taken as Sampling frequency+2 (i.e 16+2=18 in the case of 16 khz sampling frequency). Then converting these LPC parameters to Q- Cepstral coefficients (with Q=18).

Cepstral Coefficients:

$$W(i) = 1+(Q/2)\mathrm{Sin}(\pi i/Q), \text{ (for } i=1,2,3\ldots,Q \text{)} \quad \text{---------(1)}$$

Now each occurrence of the words with T frames have 'T' feature vector constitutes {y1, y2, y3,…….,yT} observation vectors and 36 weighted cepstral coefficients for each frame for Q=18).

2.3 K-means algorithm
Let 'K' vectors be initiated from the training feature vectors called clusters. Let each vector in training set belong to a cluster, then Euclidian distance calculated between the initiated cluster vector (C_1, C_2, C_3, ……., $C_{k)}$ and the training set vector as:

$$d(x,Ck) = \mathrm{Sqrt}[(x-Ck)(x-Ck)T] \text{ , for } 1 \le k \le K \quad \text{----------(2)}$$

Find the closest clusters to the training set vectors as:

$$k^* = \mathrm{Arg}\ \mathrm{Min}_k[d(x,Ck)] \quad \text{----------(3)}$$

Re-compute the Clusters Ck ($1 \le k \le K$) by taking the mean of the vectors that belongs to the centroids. This is done for every Ck. This iteration process continued till the

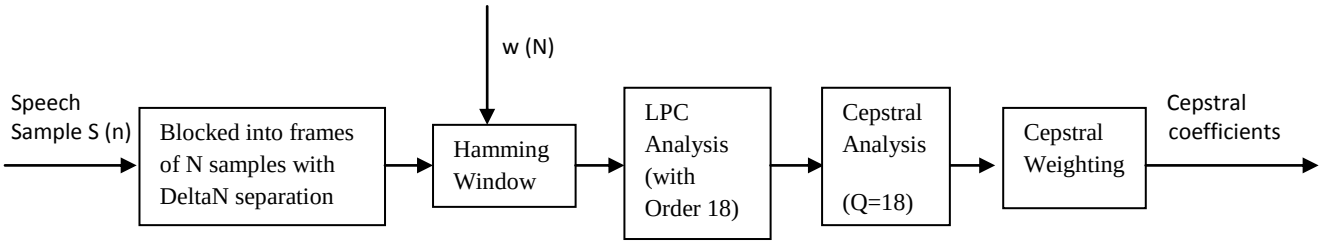

Fig1: Block diagram for Feature extraction (LPC and Cepstral coefficients)

Cluster Centroids adjusted such that all the training vectors belong to any of this cluster centroid with minimum distortion.
By using k-means vector quantization clustering algorithm, a discrete observation sequence i.e. code book is formed with k integer values in the range of 1<k<K.

2.4 Training the HMM

Now all the feature vectors of word sequence are quantized to any of the cluster vectors Ck ($1 \leq k \leq K$). Now a dedicated HMM (λ = {A, B, pi (1), S for yk, 1<k<K) Parameters be estimated characterized by state equation and observation sequence to each of the word. 'S' is the number of hidden states in the model. Let us assume 'xt' is the state random process, K is number of distinct observation symbols per state i.e number of 'yk' in quantized code book.

State probability vector at time't' is: pi (t) = A pi (t-1)
State transition matrix: A_{SxS}= aij,
{aij=a(i/j)=P(xt=i)/P(xt-1=j)}
Observation Probability Matrix: B_{KxS}= b_{ki},
{b_{ki}=b(k/i)=P(yt=k)/P(xt=i)}

Training problem, adjust the model parameters to maximize the likelihood :

$$\lambda^* = \max\{P(O/\lambda)\} \quad \text{----------(4)}$$

By forward/ backward recursion algorithms the above training problem adjusted and the HMM model is set for each of the isolated word.

2.5 Recognition of Isolated words

The unknown isolated word will be recognized from the HMM models constructed from the training samples. The features of the unknown word speech sequence will be extracted and quantized through the same process which is used while framing the HMM model. Then apply these quantized sets of the isolated word to each of the HMM model of all the words and calculate the probability of occurring the word: ($P(Y/\lambda_i)$, for i=1,2,3.....). Based on the maximum value of ($P(Y/\lambda_i)$, for i^{th} HMM model will be guessed as i^{th} word will be the recognized word among the available HMM word speech models. In this paper 20 Telugu isolated words used to frame HMM models.

3. TELUGU WORDS WITH HARD AND SOFT SOUNDS

Telugu is the second most spoken language in INDIA after Hindi. Telugu written from left to Right as like in English. The modern Telugu contains Persian and Arabic expressions. Some English words also being used in Telugu Language. For example ,the foreign English words Collector being used as kalkataru (కల్కటరు) and assistant being used as ashiShtaaMtu (అశిష్టాంటు), In the same way the Arabic word jawab means Answer being used as javaabu (జవాబు) in Telugu [1].

The pronunciation of some Telugu consonants like Cha (ఛ), and Ja (ఝు), ca (స), and za (జ) is peculiar. The Hard sounds Cha (ఛ), and Ja (ఝు) are sometimes softened into ca (స) and za (జ). In ancient pronunciation, Telugu uses the Soft sounds alone: like chinna (చిన్న) and cheppu (చెప్పు) were pronounced as cinna (సిన్న) and ceppu (సెప్పు). The letter la (ల) soft sound may sometimes mispronounced as harsh sound La(ళ) with tongue upward. The letters ta (ట), da (డ), Na (ణ) are harder and the sounds tha (త), dha(ద) ,na (న) are softer.

In this paper we focus on Recognition of Telugu Isolated words consisting of hard and soft sounds. In Telugu many words of Hard and Soft sounds with almost same Vocal activity in pronouncing have different meaning altogether. For Example the words 'paata – పాట' (hard 'T') means 'song' and 'paatha – పాత' (soft 'T') means 'Old' have the same Vocal activity and the only difference is the Hard and soft 'T' which changes the meaning altogether. In the same way the word 'PaMdi – పండి' (hard 'D') means 'having fruited' and 'paMdhi- పంది' (soft 'D') means 'Pig' have the same vocal activity have different meaning just because of hard and soft sounds. So such words need to be recognized correctly with hard and soft sounds. If the word recognized as soft in place of hard the meaning of total sentence may be changed. In this paper we attempted to recognize such words by having the vocabulary consists of both the sounds. The words that are taken for experiment have been mentioned in the tables below.

Table 1: Telugu words ending with Hard and Soft 'T'

Hard T (ta)		**Soft T (tha)**	
word	**Meaning**	**Word**	**Meaning**
poatu (పోటు)	stab	pothu (పోతు)	Beast
koata (కోట)	fort	koatha (కోత)	cutting
paata (పాట)	song	paatha (పాత)	old
koati (కోటి)	Ten millions	koathi (కోతి)	monkey
moota	bundle	mootha	cover

(మూట)		(మూత)	
katti (కట్టి)	binding	kaththi (కత్తి)	sword
thaata (తాట)	Brake of a tree	thaatha (తాత)	Grand father
pottu (పొట్టు)	husk	poththu (పొత్తు)	friendship
potti (పొట్టి)	short	potthi (పొత్తి)	rag
viMta (వింట)	bow	viMtha (వింత)	Strange

Table 2: Telugu words ending with Hard and Soft 'D'

Hard 'D' (da)		**Soft D(dha)**	
Word	**Meaning**	**Word**	**Meaning**
podi (పొడి)	powder	Podhi (పొది)	pouch
paadu (పాడు)	sing	paadhu (పాదు)	Bed of tree
niMda (నిండ)	fully	niMdha (నింద)	blame
padi (పడి)	Having fallen	padhi (పది)	ten
paMdi (పండి)	Having fruited	paMdhi (పంది)	pig
dhiddi (దిడ్డి)	door	Dhidhdhi (దిద్ది)	correcting
vaadu (వాడు)	he	vaadhu (వాదు)	quarrel
gadda (గడ్డ)	boil	gadhdha (గద్ద)	vulture
buddi (బుడ్డి)	bottle	budhdhi (బుద్ది)	sense
poda (పొడ)	speck	podha (పొద)	bush

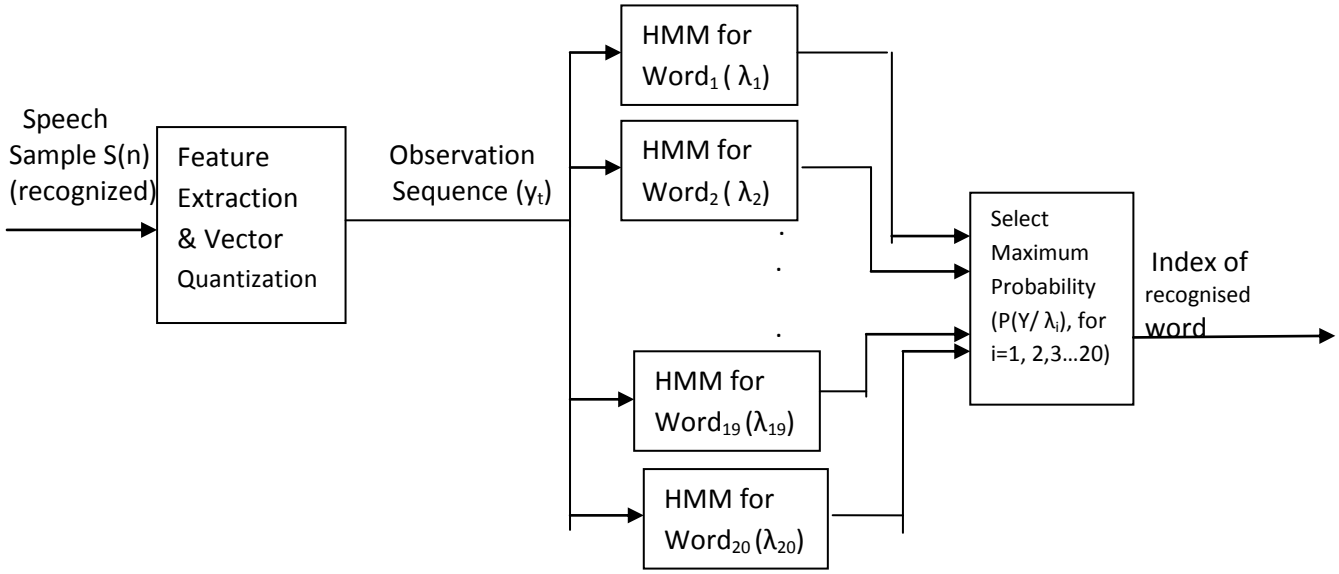

Fig.2: Isolated word recognition using HMM

4. EXPERIMENTAL RESULTS

The Speech samples for 20 Telugu words tabulated above which consisting both hard and soft sounds have been collected from 10 different speakers for 2 times (4 male speakers, 4 female speakers and 2 child speakers). Means Each Telugu word has 2 sets of 10 different samples. One Set used for modeling the HMM and the other set used for testing the recognition success/error rate [5]. 20 HMM models for 20 different words, one HMM for one isolated word framed by using 10 samples of each word. Now another set of 20 words with 10 samples of each word applied to these framed HMM models for testing. The word selected randomly from each set of words and applied to all the HMM models and calculate the probability of occurrence for each model, then the maximum probability of occurrence guessed as recognized word. All the words from the testing set being applied to HMM models and the results are tabulated. 4 different cases are analyzed for different count of centroid HMMs and different frame lengths and overlaps [5, 6].

From the results it is observed that the HMMs need to be framed with more centroids while recognizing and differentiating hard and soft Telugu sounds. Means the HMMS which are capable recognize the normal words needs to be framed with more centroids if there is a need to differentiate hard and soft Telugu isolated sounds

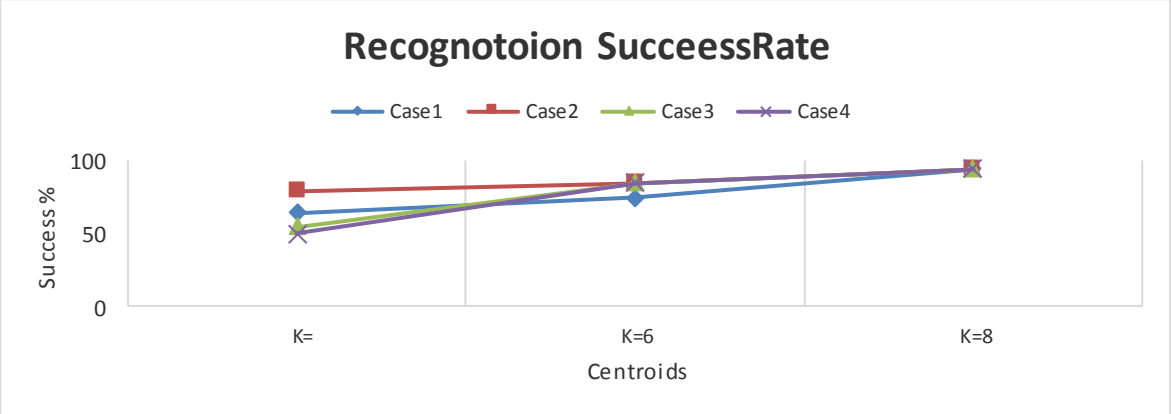

Fig.3: Success rate for HMMs of different count of cluster centroids

Case1: 10 ms frame and 5 ms overlap (N=160; deltaN=80)

Centroids	Distortion	No.words recognized	Success %
K=4	0.0234	13/20	65
K=6	0.0220	15/20	75
K=8	0.0210	19/20	95

Case2: 10 ms frame and 2.5 ms overlap (N=160; deltaN=40)

Centroids	Distortion	No.words recognized	Success %
K=4	0.0163	16/20	80
K=6	0.0153	17/20	85
K=8	0.0146	19/20	95

Case3: 20 ms frame and 5 ms overlap (N=320; deltaN=80)

Centroids	Distortion	No.words recognized	Success %
K=4	0.0227	11/20	55
K=6	0.0212	17/20	85
K=8	0.0201	19/20	95

Case4: 20 ms frame and 10 ms overlap (N=320; deltaN=160)

Centroids	Distortion	No.words recognized	Success %
K=4	0.0331	10/20	50
K=6	0.0314	17/20	85
K=8	0.0297	19/20	95

5. CONCLUSION

With the above experiments it is observed that by increasing the centroids in HMM model, the recognition success rate is increased but the mathematical computation also increased accordingly. It was also observed that the error occurred mainly between hard and soft sounds, i.e the Model recognizing 'koati' as 'koathi' etc for lesser cluster centroids. So while recognizing these kind of special words, the model needs to be designed with more centroids to recognize with high success rate.

6. FUTURE WORK

In this paper we focused on recognizing Telugu isolated words consisting of hard and soft sounds for less vocabulary using HMMs, the same needs to be analyzed for large vocabulary systems. For training and testing we used the same set of 10 different speakers, the success rate may be decreased if the different set of speakers used for training and testing .this issue may over come if large vocabulary (more number of speakers) is taken for training the HMMs. This work may be extended using ANN and other neural network approaches.

7. REFERENCES

[1] A Dictionary of the Mixed Dialects and foreign words used in Telugu by Charles Philip Brown -1854

[2] Hidden Markov Models for Speech Recognition B. H. Juang; L. R. Rabiner ,Technometrics, Vol. 33, No. 3. (Aug., 1991), pp. 251-272.

[3] Modelling asynchrony in automatic speech recognition using loosely coupled hidden Markov models H.J. Nock, S.J. YoungCambridge University Engineering Department, Trumpington Street, Cambridge CB2 1PZ, UK,Accepted 22 March 2002

[4] The Application of Hidden Markov Models in Speech Recognition by Mark Gales and Steve Young-Foundations and Trends in Signal Processing Vol. 1, No. 3 (2007) 195–304-(2008).

[5] Speech Recognition using HMM: Performance evolution in noisy environment by Mikael Nilsson and Marcus Ejnarsson, Blekinge Institute of Technology,March 2002.

[6] Speech Matlab exercises: IT University of Copenhagen Multimedia Technology Speech and IT Systems.

8. AUTHOR'S PROFILE

Shaik Shafee, Reasearch student, S.V. University College of Engineering, Dept. Of E.C.E, Tirupati, A.P, INDIA.

B.Anuradha, Professor , S.V. University College of Engineering, Dept. Of E.C.E, Tirupati, A.P, INDIA.