Effect of states and mixtures in HMM model and Connected word Recognition of Profoundly deaf and hard of hearing speech

C. Jeyalakshmi [1], Dr. V. Krishnamurthi [2], Dr. A. Revathy [3]

[1]Department of ECE,Trichy Engineering College,Trichy,Tamilnadu,India.
[3]Department of ECE,Saranathan College of Engineering,Trichy,Tamilnadu,India.
[1] lakshmi.jeya67@yahoo.com
[3] revathidhanabal@rediffmail.com
[2]Department of ECE, Former HOD, Anna university Chennai,Tamilnadu,India.
[2] profvkmurthi@yahoo.co.in

***Abstract*— It is a challenge for many years that how to fix the no. of states and no. of mixtures when HMM models are used for speech recognition. In this paper we have analysed that for hearing impaired speech that is partially intelligible to people who are speaking to them frequently and it is not understandable by the unfamiliar listeners. They suffer in many aspects like education and in public places to converse with the normal speakers. Since speech is unique most of the time normal speech itself could not be understand by others. If we develop the speech recognizer for their speech it will convert their unintelligible speech into intelligible speech. Speaker dependent connected digit recognition for this task using HTK tool kit is done and the average recognition accuracy obtained is 93%. Totally 10 speakers out of which 3 are hard of hearing and 7 are profoundly deaf are considered for this experiment. Then for isolated words, no. of mixtures are varied from 3 to 10 for each set of states such as 6, 7, 8, 9, 10 and recognition accuracy is verified for each case. When we varied beyond that there is no any significant change in recognition accuracy and so it is concluded that we can have mixture and state value as 10 for small vocabulary and the recognition performance for all types of feature is comparable to that of normal speech recognition. But irrespective of the state higher recognition is achieved at 8 or 9 or 10 mixer value for different type of feature and it can be concluded that, if we have the mixer value as 8 , 9 or 10 we can get reasonable results.**

Keyword- Mel frequency cepstral coefficients (MFCC), Perceptual linear prediction coefficients(PLP), Linear Prediction cepstral coefficients(LPCC), Hidden Markov model (HMM), Hidden Markov Model tool kit (HTK).

I. INTRODUCTION

Most of the studies concentrated in assisting the deaf to recognize the speech of the normal for education and other purposes. But for communication it is necessary to recognize their speeches also, even though it is tough to understand for the unfamiliar listeners other than their teachers and parents. Hearing is the primary sensory channel through which children acquire speech and language and if it is affected they cannot hear the sounds so they cannot speak even though their vocal apparatus is good. Unlike the eye, the ear is always open so that speech may be heard whenever it is present, moreover it is non-directional. The listener is not required to face the speaker in order to receive a message, so that much of what we learn is behind us or out of our ear. So that it is surprising that learning of speech and language will be imperfect when hearing is significantly impaired at birth which is known as congential hearing impairment. Articulation of speech sounds, co articulation of their combinations, nuances of speech rhythm, modulation of loudness of voice, and the complexities of spoken language will not be learned adequately if hearing impairment is severe. Even when hearing is severely impaired it may still be utilized in acquiring and teaching speech. Hearing aids and lip-reading are more effective in face-to-face communication [1].

Measurement of hearing is done by audiometer to measure the level of hearing. The audiometer delivers to both ear, tones of known frequency and intensity to which the listener is expected to respond. When the listener can just detect the tone and respond, that level is called the listener’s hearing–threshold level for that tone. Threshold level responses for the different frequencies are commonly plotted on a chart called a pure-tone audiogram which is shown in figure 1. The vertical dimension of the form designates differences in intensity that reflect the loudness we might hear and horizontal dimension designates tones of different frequency. From top to bottom the perceived loudness of the tones is greater and it includes levels of 120 decibels and greater. These levels would be extremely loud to a person with normal hearing, about like standing near a jet airplane engine, but it might be merely a threshold level for some profoundly deaf listeners. As threshold levels are marked for each of the tones tested on one ear, they are connected by a line that displays the audiometric curve or configuration. This is also shown in figure 1.

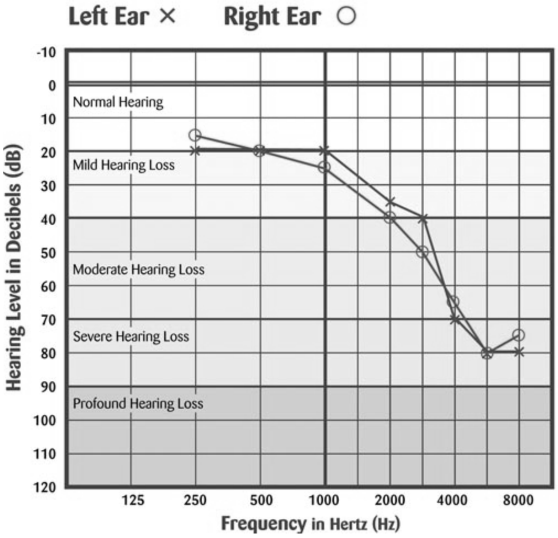

Fig.1 Audiogram and audiometric curve for left and right ear

Since the audiometric curve has to be given at a single decibel level which is usually the average level of 500Hz, 1000Hz, and 2000Hz commonly called the speech frequencies. Speech is very complex in frequency compared to a pure tone, containing combinations of frequencies ranging from 100Hz to 3000Hz. The most speech information is carried in a range of frequencies where the ear is most sensitive, between 400 Hz and 3000Hz.

The process of deliberately formulating accurate syntax and careful articulation tends to slow the rate of speech of deaf children, often causing them to produce speech in disjointed, word-by-word segments. Due to this their speech sound is unnatural. Similarly pitch determination is very likely to be affected by the severe hearing loss; perception of intonation through pitch changes may not be available to all deaf children, even with amplification. If we consider the formants, although there may be several formants associated with a particular sound, the first two or three are most important and vowels can be recognized when only the first two formants are audible. These formants will not be audible to the deaf child unless the speech signal is amplified. Instead of amplifying the entire speech spectrum, by selective frequency amplification we can make at least the F1 and F2 to be audible to the deaf listener. So that we can use the formants for classification of deaf and normal speaker [2]. Consonants are more diverse in their acoustic characteristics than the vowels. Hearing the formant transition of the first two or three formants is sufficient for the consonant recognition. Whereas consonants are characterized by vocal tract constriction, high frequency components, and often aperiodicity, vowels are characterized by sustained voicing, lack of constriction, and a dominant lower frequency formant structure. [3]

Their speech development is depend on the age of onset of hearing loss, severity and type of hearing impairment, the age at which the hearing loss was detected and the age at which guidance was sought for, from the speech therapist and audiologist[4]. Though their pronunciation is very poor and difficult to understand, having MFCC features[5] and Hidden markov model [6], with 7 mixtures isolated digit Tamil speech recognition system using HTK tool kit for recognition produced 92.4% accuracy for the hearing impaired [7]. In this study we have varied the states and mixtures up to 10 and the performance is evaluated. The figure 2 shows [8] a plot of average word error rate versus no. of states N for the case of isolated digits. It can be seen that the error is somewhat insensitive to N and achieving local minimum at N=6.

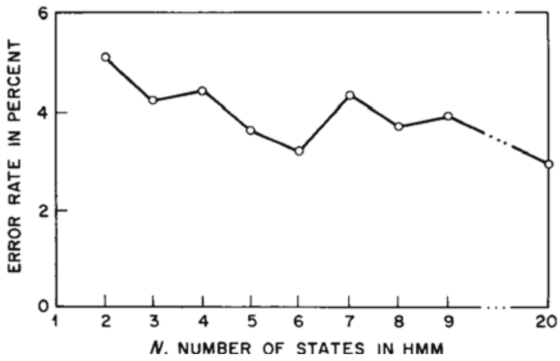

Fig.2 Average word error rate verses no. of states

In general, states can be fixed as 6 according to the above or depend on the no. of phonemes in a word it can be fixed. Each word has different phonemes so that no. of states is kept different for all the words.

When generating models for each word using HMM we have to mention the no. of states and mixtures and for no. of mixtures it is trial and error only. In order to optimize the no. of states and mixtures for short vocabulary, in this study we have varied them and then investigated the performance of recognition. We have also extended the recognition for connected words using HTK.

II. EXPERIMENTS AND METHOD

A. Deaf Speech Data

Hearing impaired can able to follow only one language mostly their native language since it is tough for them to follow different facial expressions and throat vibrations. In this work we have taken 10 Tamil isolated digits and connected words from the students of Maharishi Vidhya Mandir service centre for the hearing impaired, Tiruchirappalli. Children have spoken each word 20 times and their speeches are recorded using high quality microphone. 15 speeches of 10 speakers are considered for training and remaining 5 speeches of 10 speakers are considered for testing.

B. Model Generation

Hidden Markov models [9] are widely used for automatic speech recognition because they have a powerful algorithm in estimating the model parameters and achieve a high performance. Once a structure of the model is given the model parameters are obtained automatically by applying training data. In HMM for each state, there is an output probability distribution of an acoustic vector, and each iteration is associated with a state-transition probability. These probabilities are called the model parameters and can be estimated effectively by using Baum Welch algorithm [10] [11]. Here we have used HTK toolkit for building Hidden Markov Models. However, HTK is primarily designed for building HMM-based speech processing tools, in particular recognizers.

Here the models are created for each word and whenever unknown speech is given, its correct transcription is got at the output. There are two major processing stages involved. First, the HTK training tools [12] are used to estimate the parameters of a set of HMMs using training utterances and their associated transcriptions. Secondly, unknown utterances are transcribed using the HTK recognition tools. HRest can be used for normal isolated word training which causes the parameters of the given HMM to be re-estimated repeatedly using the data files until either a maximum iteration limit is reached or the re-estimation converges.

Initially Tamil isolated digits from poojam (0) to ombadu (9) is taken and maximum phoneme among these words are 6 and the mixture value is varied from 3 to 10. By considering LPCC, MFCC, PLP [13] [14] features, word models are created and tested. Then set of words are added with the isolated digit to get a maximum phoneme of 7,8,9 and 10 and by varying the mixture value again from 3 to 10 the recognition performance is evaluated using HTK.

III. RESULTS AND DISCUSSION

The recognition performance of isolated digits from poojam (0) to (9) whose maximum states are 6 and mixture from 3 to 10 is shown in Table I.

TABLE I. RECOGNITION PERFORMANCE FOR MAXIMUM PHONEME SET OF 6.

Mix ture value	Fea ture	Recognition accuracy of Hearing impaired Speech utterance along with states									
		poojam 6	**ondru 4**	**irandu 6**	**moondru 5**	**naangu 5**	**aindhu 4**	**aaru 3**	**yelu 3**	**yettu 4**	**ombadu 6**
3	LPCC	92.5	77.5	75	70	70	70	65	50	77.5	95
	PLP	95	75	82.5	92.5	82.5	80	82.5	85	90	95
	MFCC	92.5	77.5	85	80	82.5	82.5	75	92.5	87.5	95
4	LPCC	92.5	80	82.5	80	90	80	62.5	85	82.5	97.5
	PLP	92.5	80	77.5	92.5	85	95	85	95	90	100
	MFCC	95	87.5	87.5	82.5	82.5	92.5	85	92.5	92.5	97.5
5	LPCC	90	92.5	70	87.5	75	77.5	77.5	77.5	87.5	100
	PLP	90	85	85	87.5	80	85	82.5	82.5	82.5	100
	MFCC	95	87.5	90	90	85	87.5	90	82.5	90	95
6	LPCC	90	92.5	85	87.5	80	87.5	85	92.5	92.5	97.5
	PLP	95	85	82.5	97.5	85	90	85	95	97.5	97.5
	MFCC	95	92.5	87.5	92.5	77.5	92.5	85	92.5	95	97.5

7	LPCC	92.5	90	82.5	95	77.5	90	87.5	90	97.5	97.5
	PLP	95	90	85	92.5	80	95	97.5	92.5	97.5	100
	MFCC	92.5	85	90	97.5	97.5	95	97.5	95	97.5	97.5
8	LPCC	90	90	85	97.5	82.5	92.5	90	97.5	100	97.5
	PLP	92.5	87.5	87.5	97.5	85	92.5	97.5	95	97.5	97.5
	MFCC	92.5	87.5	87.5	95	87.5	92.5	100	95	97.5	100
9	LPCC	92.5	95	82.5	97.5	85	100	92.5	92.5	100	97.5
	PLP	97.5	87.5	85	95	87.5	95	97.5	95	95	100
	MFCC	95	85	87.5	97.5	90	92.5	97.5	97.5	97.5	100
10	LPCC	92.5	95	82.5	92.5	87.5	95	90	90	95	95
	PLP	97.5	85	92.5	100	90	100	90	95	95	97.5
	MFCC	92.5	92.5	92.5	95	90	97.5	100	95	97.5	100

TABLE II. OVERALL RECOGNITION PERFORMANCE

Type of Features	**Average recognition accuracy in %**							
	3 mixture	**4 mixture**	**5 mixture**	**6 mixture**	**7 mixture**	**8 mixture**	**9 mixture**	**10 mixture**
LPCC	74.25	83.25	83.5	89	90	92.25	93.5	91.5
PLP	86	89.25	86	91	92.5	93	93.5	94.25
MFCC	85	89.5	89.25	90.75	94.5	93.5	94	95.25

Similarly a set of three words, instead of last three isolated digits are included consecutively to get a maximum phonemes of 7,8,9,10 and the above procedure is repeated and the results are tabulated correspondingly.

In order to extend our investigation, 10 connected words which are frequently uttered by the deaf students are taken into consideration and the recognition performance is validated.

The average recognition performance is also shown in Table II. Instead of individual word it is the overall recognition performance for all the words.

In table I for LPCC feature and for the mixture value 3 the performance is very much degraded and when we increase the mixture value it is increasing for all the features but not gradually. i.e. the variation is not linear so that we have analysed by taking for how many words for a particular mixer, the recognition is maximum and it is also not linear. When considering the average recognition performance LPCC is maximum at mixture value 9 and for 10 it is decreasing but for PLP and MFCC it is maximum for m=10.

Now we have added newly three words namely vanavil, vilayadu, sapidu which has the phoneme value 7, by replacing the words yelu,yettu,ombadu. Models are created for this new data by considering 3 features and again the performance is verified and it is given in Table III and Table IV shows the average recognition performance.

TABLE III RECOGNITION PERFORMANCE FOR A MAXIMUM PHONEME SET OF 7

Mix ture value	Fea ture	Recognition accuracy of Hearing impaired Speech utterance along with states									
		Poo jam 6	**ondru 4**	**Iran du 6**	**Moon Dru 5**	**naangu 5**	**Ain dhu 4**	**aaru 3**	**vanavil 7**	**Vila yadu 7**	**Sapi du 7**
3	LPCC	92.5	80	65	65	72.5	65	60	55	55	95
	PLP	95	65	75	77.5	60	82.5	75	85	82.5	90
	MFCC	95	87.5	80	77.5	65	92.5	65	77.5	87.5	95
4	LPCC	95	87.5	77.5	72.5	75	82.5	85	87.5	82.5	95
	PLP	97.5	82.5	82.5	92.5	67.5	90	85	87.5	92.5	95
	MFCC	95	85	92.5	87.5	67.5	87.5	85	85	90	95
5	LPCC	95	82.5	82.5	82.5	75	90	90	87.5	82.5	95
	PLP	92.5	80	92.5	90	85	90	87.5	85	92.5	92.5
	MFCC	95	90	87.5	85	75	92.5	92.5	77.5	95	95
6	LPCC	95	90	85	82.5	75	82.5	82.5	85	85	95
	PLP	95	80	95.5	90	80	90	90	82.5	95	92.5
	MFCC	95	92.5	92.5	87.5	90	90	82.5	87.5	95	95
7	LPCC	90	92.5	87.5	77.5	85	90	75	87.5	82.5	95
	PLP	100	85	95	90	87.5	97.5	82.5	85	92.5	95
	MFCC	97.5	90	95	92.5	90	97.5	90	87.5	95	95
8	LPCC	92.5	90	85	82.5	87.5	92.5	85	87.5	87.5	97.5
	PLP	100	90	97.5	92.5	87.5	95	87.5	85	95	92.5
	MFCC	97.5	87.5	85	92.5	90	97.5	95	87.5	95	92.5
9	LPCC	92.5	92.5	87.5	82.5	82.5	87.5	90	90	87.5	95.5
	PLP	97.5	87.5	95	97.5	90	95	85	87.5	95	95
	MFCC	97.5	90	90	97.5	87.5	100	90	87.5	95	95
10	LPCC	97.5	85	87.5	82.5	82.5	90	85	87.5	97.5	95
	PLP	97.5	90	97.5	100	87.5	100	97.5	87.5	95	95
	MFCC	97.5	90	92.5	90	92.5	100	90	85	95	95

TABLE IV OVERALL RECOGNITION PERFORMANCE

Type of Features	Average recognition accuracy in %							
	3 mixture	**4 mixture**	**5 mixture**	**6 mixture**	**7 mixture**	**8 mixture**	**9 mixture**	**10 mixture**
LPCC	70.5	84	86.25	85.75	86.25	88.75	88.8	89
PLP	78.75	87.25	88.75	89.05	91	92.25	92.5	94.75
MFCC	82.25	87	88.5	90.75	93	92	93	92.75

For each word the recognition is different according to the vowels and consonants present in the word. Here, only for LPCC with 3 mixtures recognition is poor but for the words poojam and sapidu it is high due to the soft consonants present in the word.

To know the performance for a maximum phoneme set of 8, we have introduced 3 words namely seekiram, vanakkam, vidumurai and models are generated and tested. The recognition performance is tabulated in Table V. The overall recognition performance is also verified and it is given in Table VI.

TABLE V RECOGNITION PERFORMANCE FOR A MAXIMUM PHONEME SET OF 8

Mixture value	Feature	Recognition accuracy of Hearing impaired Speech utterance along with states									
		poojam 6	**ondru 4**	**irandu 6**	**moondru 5**	**naangu 5**	**aindhu 4**	**aaru 3**	**seekiram 8**	**Vanak kam 8**	**Vidu murai 8**
3	LPCC	92.5	70	82.5	75	82.5	82.5	82.5	90	92.5	95
	PLP	95	80	77.5	85	67.5	85	87.5	87.5	87.5	95
	MFCC	97.5	97.5	87.5	85	72.5	77.5	75	67.5	92.5	95
4	LPCC	92.5	85	90	90	80	90	85	85	95	95
	PLP	97.5	90	85	95	82.5	95	80	92.5	90	95
	MFCC	97.5	85	90	92.5	72.5	82.5	85	92.5	92.5	95
5	LPCC	95	97.5	92.5	90	85	82.5	80	85	95	95
	PLP	100	87.5	87.5	95	77.5	95	97.5	95	90	95
	MFCC	97.5	95	92.5	97.5	87.5	95	92.5	92.5	95	95
6	LPCC	97.5	95	85	90	90	95	80	92.5	95	95
	PLP	100	87.5	92.5	97.5	90	92.5	97.5	95	90	95
	MFCC	97.5	82.5	82.5	92.5	82.5	92.5	95	92.5	85	95
7	LPCC	97.5	97.5	90	90	82.5	97.5	85	95	95	95
	PLP	97.5	90	95	95	90	87.5	95	95	87.5	95
	MFCC	100	87.5	87.5	92.5	87.5	97.5	97.5	95	90	95
8	LPCC	100	100	92.5	92.5	95	95	85	92.5	95	95
	PLP	100	90	95	95	95	95	97.5	95	90	95
	MFCC	97.5	90	87.5	95	87.5	97.5	92.5	92.5	87.5	95
9	LPCC	97.5	100	87.5	90	90	95	90	92.5	95	95
	PLP	97.5	90	92.5	97.5	95	97.5	97.5	95	90	95
	MFCC	97.5	92.5	92.5	95	97.5	100	100	95	90	95
10	LPCC	97.5	90	92.5	87.5	90	97.5	87.5	95	95	95
	PLP	97.5	100	95	97.5	97.5	97.5	100	95	90	95
	MFCC	97.5	92.5	92.5	92.5	90	97.5	92.5	92.5	87.5	95

TABLE VI OVERALL RECOGNITION PERFORMANCE

Type of Features	Average recognition accuracy in %							
	3 mixture	**4 mixture**	**5 mixture**	**6 mixture**	**7 mixture**	**8 mixture**	**9 mixture**	**10 mixture**
LPCC	84.5	88.75	89.75	91.5	92.5	94.25	93.25	92.75
PLP	84.75	90.25	92	93.75	92.75	94.75	94.75	96.5
MFCC	84.75	88.5	94	89.75	93	92.25	95.5	93

The performance of recognition is maximum at 8 ,9 and 10 mixture for LPCC, MFCC, PLP. Also there is a slight variations in the performance and among the three features LPCC only has low accuracy.

From the Table V it is clear that for the words vidumurai and poojam, accuracy is above 90% irrespective of the mixture values and for other words, except for mixture value 3,4,5 the accuracy is above 80%.

We couldn't fix the mixture value for a particular feature since it is not gradually increasing. Even though it does not produce abrupt changes, there is 2 or 3% changes in the recognition accuracy. So for the mixture value between 8,9 and 10 the performance of the system is reasonable.

For a maximum phoneme set of 9 the words considered are pathimoondru, pathinangu, pathinindu and the performance is tabulated in Table VII,VIII .

TABLE VII. RECOGNITION PERFORMANCE FOR A MAXIMUM PHONEME SET OF 9

Mix ture value	Feature	Recognition accuracy of Hearing impaired Speech utterance along with states									
		poojam 6	ondru 4	irandu 6	Moon dru 5	naangu 5	aindhu 4	aaru 3	Pathi moondru 9	Pathi nangu 9	pathi nindu 9
3	LPCC	97.5	85	80	82.5	80	82.5	80	95	97.5	90
	PLP	92.5	90	82.5	90	87.5	82.5	82.5	97.5	97.5	100
	MFCC	95	90	87.5	75	67.5	80	70	97.5	97.5	97.5
4	LPCC	92.5	92.5	70	87.5	80	90	90	97.5	97.5	90
	PLP	97.5	87.5	85	95	87.5	80	85	95	100	97.5
	MFCC	95	87.5	87.5	90	85	92.5	90	100	100	97.5
5	LPCC	95	90	90	87.5	87.5	87.5	77.5	95	90	92.5
	PLP	95	85	90	95	82.5	87.5	92.5	97.5	100	97.5
	MFCC	95	85	87.5	90	80	95	95	97.5	100	92.5
6	LPCC	95	97.5	82.5	82.5	85	90	82.5	97.5	100	92.5
	PLP	95	90	92.5	97.5	80	90	95	100	100	100
	MFCC	97.5	85	87.5	95	85	97.5	95	97.5	100	100
7	LPCC	92.5	92.5	90	85	75	90	90	97.5	100	97.5
	PLP	95	87.5	95	97.5	82.5	92.5	92.5	100	100	97.5
	MFCC	97.5	87.5	87.5	97.5	92.5	97.5	95	100	100	100
8	LPCC	95	95	92.5	90	82.5	97.5	87.5	97.5	100	97.5
	PLP	100	100	100	100	100	100	100	100	100	100
	MFCC	100	82.5	90	97.5	95	97.5	97.5	100	100	97.5
9	LPCC	97.5	92.5	87.5	85	90	100	92.5	100	100	95
	PLP	97.5	87.5	92.5	97.5	97.5	100	95	100	100	97.5
	MFCC	95	85	85	95	87.5	97.5	92.5	100	100	97.5
10	LPCC	95	87.5	87.5	92.5	90	100	92.5	100	100	92.5
	PLP	97.5	85	92.5	97.5	92.5	100	97.5	100	100	97.5
	MFCC	97.5	90	87.5	95	97.5	97.5	95	100	100	97.5

TABLE VIII OVERALL RECOGNITION PERFORMANCE

Type of Features	Average recognition accuracy in %							
	3 mixture	4 mixture	5 mixture	6 mixture	7 mixture	8 mixture	9 mixture	10 mixture
LPCC	87	88.75	89.25	90.5	91	93.5	94	93.75
PLP	90.25	91	92.25	94	94	100	96.5	96
MFCC	87.75	92.5	91.75	94	95.5	95.75	93.5	95.75

Here by observing the results of average accuracy from 3 to 7 mixture value, it is gradually increasing and beyond that it varies as high and low. But for PLP only it varies by 4% and for LPCC, MFCC its variation is not dominant.

The results explicitly show that even for the mixture value 3 the recognition performance is comparatively high and for few times only it is around 70%. The word which has the state value 9 has nearly 100% accuracy.

Mostly for speech recognition MFCC features are preferred compared to other features due to its higher performance. But PLP features are also produces the results equally well and in this analysis it is higher than the MFCC.

Finally three words namely paththonbadhu, pathirikkai, pallikoodam are considered and the models are created having the mamimum phoneme set as 10 and performance is verified which is given in Table IX, X.

TABLE IX RECOGNITION PERFORMANCE FOR A MAXIMUM PHONEME SET OF 10

Mix ture value	Feature	Recognition accuracy of Hearing impaired Speech utterance along with states									
		poojam 6	**ondru 4**	**irandu 6**	**moondru 5**	**naangu 5**	**aindhu 4**	**aaru 3**	**Paththon badhu 10**	**Pathi rikkai 10**	**Palli kooda m 10**
3	LPCC	92.5	85	70	80	80	85	77.5	97.5	97.5	100
	PLP	100	85	80	87.5	90	87.5	67.5	95	100	100
	MFCC	92.5	95	87.5	80	62.5	70	75	85	97.5	100
4	LPCC	95	92.5	87.5	87.5	90	85	82.5	95	100	100
	PLP	97.5	77.5	82.5	92.5	85	87.5	85	90	100	100
	MFCC	95	90	90	90	80	85	90	87.5	100	100
5	LPCC	92.5	87.5	82.5	82.5	90	90	85	92.5	97.5	100
	PLP	97.5	85	87.5	97.5	80	92.5	85	97.5	100	100
	MFCC	92.5	82.5	87.5	95	77.5	87.5	85	95	100	100
6	LPCC	92.5	95	80	87.5	90	90	85	97.5	100	100
	PLP	95	85	87.5	87.5	80	85	87.5	95	100	100
	MFCC	95	90	85	87.5	67.5	90	92.5	92.5	100	100
7	LPCC	92.5	92.5	85	95	87.5	92.5	85	97.5	97.5	100
	PLP	95	85	90	97.5	87.5	97.5	95	95	100	100
	MFCC	95	90	95	92.5	82.5	97.5	97.5	97.5	100	100
8	LPCC	95	90	90	90	85	90	92.5	92.5	97.5	100
	PLP	97.5	82.5	90	100	90	95	97.5	95	100	100
	MFCC	100	77.5	92.5	95	90	95	100	97.5	100	100
9	LPCC	95	87.5	90	92.5	82.5	97.5	90	97.5	100	100
	PLP	95	85	97.5	100	90	92.5	92.5	97.5	100	100
	MFCC	97.5	85	90	97.5	97.5	92.5	90	100	100	100
10	LPCC	95	95	92.5	92.5	87.5	97.5	92.5	97.5	100	100
	PLP	97.5	87.5	92.5	97.5	92.5	95	97.5	95	100	100
	MFCC	97.5	92.5	92.5	97.5	92.5	97.5	97.5	100	100	100

TABLE X OVERALL RECOGNITION PERFORMANCE

Type of Features	Average recognition accuracy in %							
	3 mixture	**4 mixture**	**5 mixture**	**6 mixture**	**7 mixture**	**8 mixture**	**9 mixture**	**10 mixture**
LPCC	86.5	91.5	90	91.75	92.5	92.25	93.25	95
PLP	89.25	89.75	92.25	90.25	94.25	94.75	95	95.5
MFCC	84.5	90.75	90.75	90	94.75	94.75	95	96.75

By considering the state and mixture value as 10 we can achive very good results for all kind of features. When we have the mixture value as 10 the recognition system has high similarity measures to compare with the trained model. It is very clear that the accuracy is high compared to other results except for 2 or 3 times at mixer value 3. Apart from that the speech recognition performance is very high for all other words at different mixture value by considering all the features. Inspite of the mixture value 3 the words which has the state value as 6 produces very good result.

So whatever may be the kind of feature coefficients except for 3 mixtures overall performance of the recognition system is 90% and above only. Moreover all the features produce more or less the same recognition accuracy when state and mixture value is10. Real time speech contains connected words and continuous speech so we have analysed the performance of the speech recognition system for 10 connected words using HTK and the recognition is comparable to that of isolated digits. The analysis considered 10 connected words which are frequently spoken by the deaf students in the class room and the recognition performance is given in figure 3. Each student uttered each word 20 times and from this 1500 is taken for training and 500 are taken for testing.

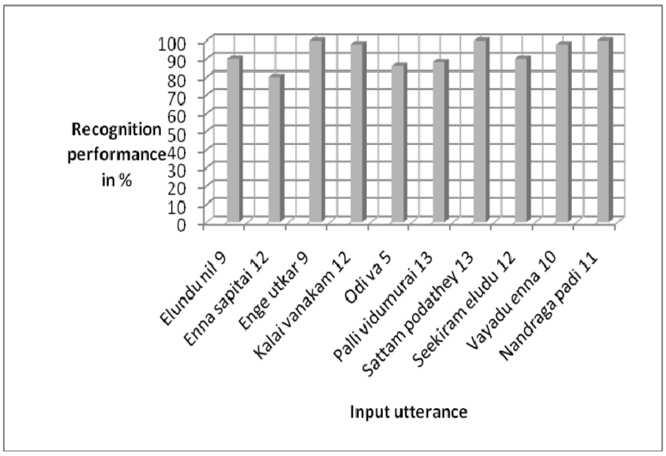

Figure 3. Connected word recognition performance

The average recognition performance is 93% and for the words Enge utkar, Odi va, palli vidumurai recognition is reduced due to hard consonants present in the word. The words are represented along with their states which has a maximum of 13 and according to our analysis we have taken the mixture as 8 and achieved the above performance.

IV. Conclusion

We know that the choice of topological configuration and the number of states in the model is generally a reflection of the apriori knowledge of the particular speech source to be modeled and is not in any way related to the mathematical tractability or implementation considerations. According to this we cannot generalize the no. of states using mathematical equation for speech recognition but for our deaf speech data taken if the states and mixtures are 10 the performance is higher. But the states are the no. of phonemes present in the word so we cannot make the states as fixed since for different word it is different. If the states are higher than ten we can concentrate only on mixtures. In general the Mixture distribution function is used to characterize the distribution of the observations in each state. If we increase the mixture value the performance will increase but we should not keep on increasing further then it will become over tuned. We can optimize the mixture value for our speech data, only on trial and error and here it is 8 to 10 and based on this, recognition performance for connected words have produced 93%. For large vocabulary we cannot use word models instead phone models are used which has single state excluding 2 dummy states, so we can vary the mixture value and optimize it for our speech data.

References

[1] Dr.Colin Brooks, Speech to text system for deaf, deafened and hard of hearing people, The Institution of Electrical Engineers, IEE-2000.

[2] C.Jeyalakshmi, V. krishnamurthi, A. Revathy, "*Deaf speech Assessment using digital processing techniques*", *Signal & Image Processing : An International journal (SIPIJ),* vol.1, no.1, sep2010.

[3] Danial forgetry,Diane Kewley-Port, and Larry E. Humes "*The relative importance of consonant and vowel segments to the recognition of words and sentences: Effects of age and hearing loss*", J Acoust Soc Am.,V- 132(3), pp.1667–1678, 2012.

[4] Nickerson R.S.,"*Characteristics of the Speech of the deaf persons*", *Volta Review,* V-77, pp.342-362, 1965.

[5] Murty, K.S.R., & Yegnanarayana, B, Combining evidence from residual phase and MFCC features for speaker recognition, *IEEE Signal Processing Letters,* vol.13, No.1, 2006, pp 52–55.

[6] C.Jeyalakshmi, V. krishnamurthi, A. Revathy, "*Transcribing Deaf and hard of hearing Speech Using Hidden Markov Model",* International Conference on Signal Processing, Communication, Computing and Networking Technologies (ICSCCN), IEEE conference publication, pp 326-331,2011.

[7] C.Jeyalakshmi, V.Krishnamurthi, A.Revathi "Building *robust hmm models for speech recognition of Hearing impaired*", *EE Times-India, eMedia Asia Ltd,* page 1-11,2012.

[8] L.R.Rabiner, *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition* - Proceedings of the IEEE, vol. 77, no. 8, Feb 1989.

[9] Picone J, "*Signal modeling techniques in speech recognition*", *Proceedings of the IEEE,* Vol.81,No.9, 1993, pp 1215–1247.

[10] B.H. Juang; L.R.Rabiner, Hidden Markov Models for speech recognition, Technometrics, Vol.33, No.3,1991, pp.251-272.

[11] Pujol P, Pol S, Nadeu C, Hagen A, Bourlard H, "Comparison and combination of features in a hybrid HMM/MLP and a HMM/GMM speech recognition system", *IEEE Transactions on Speech and Audio processing,* vol.13, Issue.1, page14-22, 2005.

[12] Steve Young, Gunnar Evermann, Thomas Hain, Dan Kershaw, Gareth Moore, Julian Odell, Dave Ollason, Dan Povey, Valtcho Valtchev, Phil Woodland, *The HTK Book,* version 3.2.1, copyright 1995-1999 Microsoft Corporation, copyright 2001-2002 Cambridge University Engineering department.

[13] Lim Sin Chee,Ooi Chia Ai, M.Hariharan and Sazali Yaacob, "*MFCC based recognition of repetitions and prolongations in stuttered speech using K-NN and LDA*", *Proc. of SCORED,* vol.16-18, (IEEE) pp 416–419,2009.

[14] H. Hermansky, "Perceptual linear predictive analysis of speech", *Journal of the Acoustical Society of America,* vol. 87, no.4, pp.1738- 1752, 1990.