Journal of Engg. Research **Vol. 5 No. (2) June 2017 pp.** *110-128*

مهمة صعبة في التعرف على الكلام للصم باستخدام نماذج السمع العادي بلغة التميل الكلاسيكية

أ. ريفازي* و سي. جيالاكشمي**

* مدرسة المعدات الكهربائية والإلكترونية، جامعة ساسترا، تانجور، تاميلنادو، الهند.

** إدارة اللجنة الاقتصادية لأوروبا، تاميلنادو كلية الهندسة، تريشي، K.Ramakrishnan، الهند.

الخلاصة

تناقش هذه الورقة وضع نظام التعرف على الكلام قوية للاعتراف بالكلام الأطفال المصابين بضعف السمع مع أية ضرورة لجمع كمية كبيرة من عينات الكلام منها. أنها عادة استخدام لغة الإشارة للتواصل مع الآخرين على الرغم من أن لديهم هيكل صوتي مثالية لإنتاج الكلام في شكل سلسلة أصوات. لذلك، من الضروري وضع نظام للاعتراف بكلماتهم وبخاصة في اللغة الأم ونظرنا التميل تحدث استماع ضعف الأطفال في عملنا. عادة مهمة صعبة الاعتراف بالخطب مشوهة المتكلمين ضعاف السمع . ولكن هنا يتم تحليل أداء النظام عن طريق تطبيق الخطب ضعيفي السمع مباشرة على النماذج التي تم تطويرها باستخدام الخطب من مكبرات الصوت العادية. يستخدم هذا العمل ميل التردد الإدراك الحسي سيبستروم التنبئي الخطي كميزة في النهاية الأمامية وك يعني تجميع في النهاية الخلفية للاعتراف بخطب الصم ويمكن عرض الخطاب المعترف بها تلفظ بمكبرات الصوت العادية كإخراج مرئي أو إخراج الصوت التي يمكن أن تكون مفهومة الواضح قبل الآخرين. النهج المتبع إزاء هذه المسألة ريال التصفية كأسلوب ما قبل معالجة إضافية ويتم تحليل الارتباط بين الكلام ريال تصفيتها والكلام العادي نظيفة لتحديد الكلام العادي نظيفة المقابلة من المجموعة من عدة خطب العادي. سوف استمع إلى خطب ضعاف السمع وتفسير واضح بمكبرات الصوت العادية، ويمكن استخدام هذا النظام للمساعدة على السمع لتحسين وضعهم في المجتمع .

A challenging task in recognizing the speech of the Hearing impaired using normal hearing models in classical Tamil language

A.Revathi* and C.Jeyalakshmi**

* *School of EEE, Sastra university,Tanjore,Tamilnadu,India.*

***Department of ECE, K.Ramakrishnan college of Engineering, Trichy,Tamilnadu, India.*

***Corresponding author: lakshmi.jeya67@yahoo.com*

ABSTRACT

This paper discusses the development of robust speech recognition system for recognizing the speech of the children with hearing impairment with no necessity to collect large amount of speech samples from them. They normally use sign language to communicate with others though they have perfect vocal structure to produce speech in the form of sequence of sounds. So, it is necessary to develop the system for recognizing their speeches especially in the native language and we have considered Tamil speaking hearing impaired children in our work. It is normally a challenging task to recognize the distorted speeches of hearing impaired speakers. But here the performance of the system is analyzed by applying the hearing impaired speeches directly on the models developed using the speeches of the normal speakers. This work uses Mel frequency perceptual linear predictive cepstrum as a feature at the front end and K means clustering at the back end for recognizing the speeches of the hearing impaired and the recognized speech uttered by normal speakers can be displayed as a visual output or audio output which can be clearly understandable by others. The approach to this issue employed RLS filtering as an additional pre-processing technique and correlation analysis is done between the RLS filtered speech and clean normal speech to select the corresponding clean normal speech from the set of several normal speeches. Hearing impaired speeches will be heard and interpreted clearly by the normal speakers and the system can be used to help the hearing impaired to improve their status in society.

Keywords: Hearing impaired(HI);Mel frequency perceptual linear predictive cepstrum (MFPLPC);normal hearing (NH); recursive least square (RLS) filtering;vector quantization**.**

INTRODUCTION

Hearing impairment ranks third on the list of chronic health conditions of older adults, after arthritis and hypertension (Craig Newman et al.,2004).Disease, damage or deformity of the cochlear hair cells is a common factor to cause hearing impairment or deafness. Loudness of sounds perceived by people in the range (0–25dB), (25–40dB), (40–55dB), (55–70dB) are classified as Normal hearing sensitivity, Mild hearing loss, Moderate hearing loss and moderately severe hearing loss respectively. Loudness of sounds perceived by people in the range (70–90dB) is classified as severe hearing loss and in the range above 90 dB is classified as profoundly deaf. Loudness of sounds perceived by humans in the range between mild to severe hearing loss is defined as hard of hearing. Among the challenges faced by hearing impaired, three important issues are education, employment and lack of basic communication skills (Jeyalakshmi C, et al., 2014).

Though the nasal and oral cavities of the HI are perfect, they cannot produce sounds since they cannot hear,asthe coordination is done by brain to understand language and speech production. A problem in the ear or damage in brain activities due to accident, stroke, orbirth defect may be the causes of sensori-neural hearing loss and people with these defects may have problems in producing speech. The assessment based on the degree of hearing ability classifies them as profoundly deaf and hard of hearing. Early detection and identification of deafness would pave the way for providing proper training methods like speech therapy to the hearing impaired to produce sounds. If the hearing loss is not detected at the right time, they feel isolated in society & family and face social stigma in life.Hearing impaired persons may understand the speech of the normal speakers by using the appropriate devices already available and most of the research has been carried out to recognize the normal speeches by HI (Brain C.J. Moore& Yoshinori Yamada et al., 2003,2000). Speeches are not pronounced in a similar manner by HI. Speeches of HI is a kind of disordered or distorted speech like dysartharic, cerebral palsy and stuttered speech and lot of works have been carried out to recognize their speech (Prasad D Polur & Gerald E Miller,;Lim sin chee et al., 2005,20062009). HMM/ANN Hidden markov model/Artifical neural network hybrid structure for a dysarthric speech (isolated word) recognition system had been investigated by them. The hybrid structure was able to provide a higher mean recognition rate (97%) than a pure HMM structure (92.2%). A small size vocabulary spoken by three cerebral palsy subjects had been utilized and the results demonstrated that the hybrid model structure had been quite robust in its ability to handle the large variability and non-conformity of dysarthric speech.

A new speech recognition system using visual features and HMM have been developed for HI (Xu Wang et al.,2008). Based on global optimization, a new genetic algorithm (GA) for training HMM had been described. Six Chinese vowels had been considered as the experimental data and ten hearing impaired speakers had been chosen as the testees. Among the 600 experimental data, 400 were used for training and 200 used for testing. This improved HMM had reached 91.47% as recognition rate, compared to 88.96% accuracy for classic HMM.The development of a flexible speech to text recognition system for cerebral palsy disabled is presented (Mohd Hafidz & Mohamad Jamil et al., 2011). It is a system, where the stored speech references in the database can be adapted flexibility according to speech of cerebral palsy disabled. The development algorithms include speech detection triggering, zero crossing rate (ZCR) for the endpoint detection, Mel-Frequency cepstral coefficients (MFCC) for the feature extraction, and dynamic time warping (DTW) for the pattern classification.The automatic speech recognition system for people with dysarthria speech disorder based on both speech and visual components have been developed (Elham S, et al., 2014). The Hidden Markov model (HMM) classifier is applied on the combined feature vector of acoustic and visual components. Hidden Markov model toolkit (HTK) is used for training and testing the proposed speech recognition. The number of speakers used in this experiment is 5 speakers for a set of 155 words. The average recognition accuracy is 58% for audio only features. Results of the system indicate that visual features are highly effective and can improve the accuracy 7.91% for speaker dependent experiments and 3% for speaker independent experiments. A new form of augmentative and alternative communication (AAC) device for people with severe speech impairment–the voice-input voice-output communication aid (VIVOCA)-is described (Mark et al., 2013). The VIVOCA recognizes the disordered speech of the user and builds messages, which are converted into synthetic speech. A novel methodology for building small vocabulary, speaker-dependent automatic speech recognizers with reduced amounts of training data, was applied. In this study HMMs with 11 states and 9 speakers participated with a maximum of 47 words of input vocabulary size. The models were trained using the HMM toolkit with the Baum-Welch algorithm. Experiments showed that this method is successful in generating good recognition performance (mean accuracy 96%) on highly disordered speech, even when recognition perplexity is increased.

Implementation of the present system in real time for understanding the speeches of the HI persons by the normal people may really help the parents, especially to understand the speeches of their children and attention may be properly paid or proper guidance may be given to them to lead a normal life.

Recognition of HI speech is carried out by the use of HMM and MFCC features using hearing impaired speech models and recognition accuracy of 84% is achieved (Jeyalakshmi C, et al., 2015). However, there are two difficulties with this approach that make the speech recognition for HI both interesting and challenging. First, the amount of data that one has to work is limited. Such a limitation may be inherent characteristics of data collection process. The second difficulty is that the speeches of HI are often disordered, distorted or corrupted by background noise or contaminated with artefacts. Due to difficulty in collecting large amount of datafrom HI children, models are developed using the speeches of NH children in this work. Testing is done with speeches of the HI. This work emphasizes the effectiveness of RLS filtering (Revathi&Venkataramani,2012) in ensuring the development of robust system for recognizing the speeches of HI with use of HI speeches only for testing.

Description about database used

Performance of the speech recognition system is usually evaluated in terms of accuracy and speed. Vocalizations for the hearing impaired persons vary in terms of accent, pronunciation, articulation, roughness, nasality, pitch, volume and speed. For this study, speech samples are collected from 10 deaf children in the age group of 10-14 years from Maharishi Vidhya Mandir service centre for the hearing impaired, Tiruchirappalli. Deaf children are not able to speak at the beginning; they are given training subsequently to follow the facial expressions and throat vibrations in one language, i.e. native language. They can learn by visually looking at the words written on the board. Among ten children considered in this work, nine are profoundly deaf and one is hard of hearing. If proper training is given to HI children to speak, they would be able to study along with normal children later on. In this work, 10 isolated digits i.e.poojyam to onbadhu (0 to 9) and commonly spoken 10 simple sentences in Tamil are considered for testing. Each child is required to utter the digit or the sentence six times and 60 speeches spoken by the HI children are considered for testing. Similarly, speeches are recorded from ten NH speakers and each speaker is required to utter digit or simple sentences twenty times. Speeches of NH speakers are used to develop training models. Speeches of HI speakers are considered for testing. Table 1 indicates the list of simple sentences in Tamil language.

Table 1. List of simple sentences in Tamil language

Sl.No.	Transliterated form of Tamil sentences	Meaning in English
1	Elundhu nil	Stand up
2	Engeutkaar	Sit here
3	Ennasaapittaai	What have you eaten?
4	Kaalaivanakkam	Good morning
5	Nandraagapadi	Study well
6	odivaa	Come here
7	Pallividumurai	School holiday
8	SaththamPodaathey	Don't make noise
9	Seekkirameludhu	Write quickly
10	Vayadhuenna	What is your age?

Speech analysis - Normal Hearing and Hearing Impaired

By way of analyzing the speeches of the hearing impaired, it is evident that their speeches are highly distorted and they take more time to pronounce the simple sentences. If their speeches are heard instantaneously, it is very difficult to interpret their speeches. Hence, more data are needed to separate out differences between talkers(Harry Levitt 1971).The language skills of these children are on an average severely retarded.Their speech production and speech reception could be of limited use, their vocabulary, grammar and reading show great deficiencies as compared to normal children (J. M. Picjlett 1969). The rate of speech production of the hearing impaired is said to be slower. The normal subjects produce 3.3 syllables per second (Pickett JM & Costam A 1968) but the deaf produce 2.0 syllables per second (Nickerson RS et al., 1974). Experiments have indicated that the NH speakers are aware of the articulator movements associated with the speech sounds (Ruscello DM et al., 1980). Figure1indicates the speech signal of the hearing impaired speaker with its spectrogram. It clearly elucidates the characteristics of the speaker and the difficulty faced by them in pronouncing simple sentence and pauses are introduced unnecessarily between syllables. It also reveals that the amount of time taken by them to utter this sentence is high.

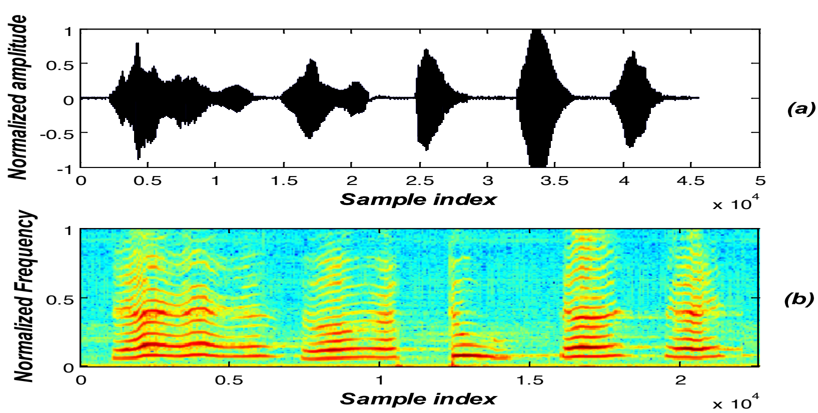

Fig.1. Analysis on HI speech 'elundunil' a) Speech signal b) Its spectrogram

Figure 2 indicates the speech signal of the normal speaker with its spectrogram. It clearly reveals that they have not faced troubles in pronouncing the sentence and the amount of time taken by them to spell out the sentence is less.

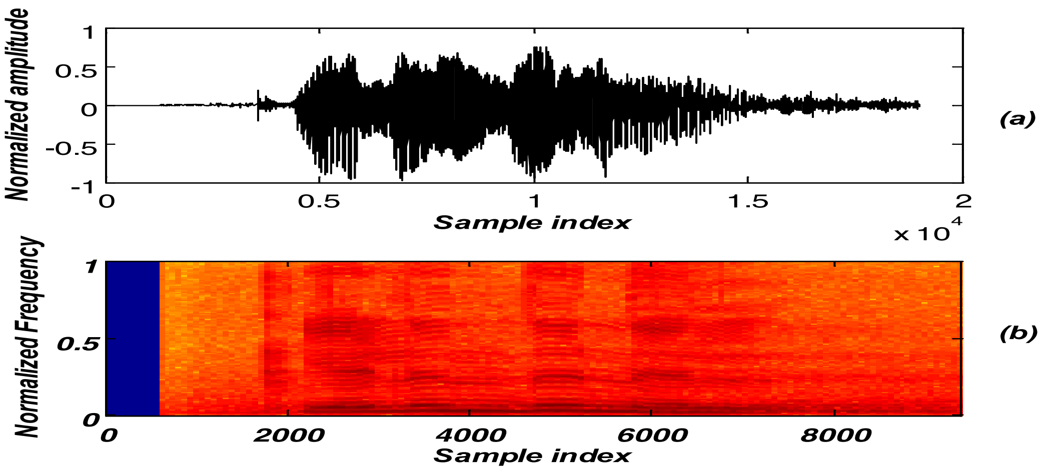

Fig.2. Analysis on NH speech 'elundunil' a) Speech signal b) Its spectrogram

From Figures 1 and 2, it is evident that there is huge variation in pronouncing the words in speeches uttered by NH speakers and HI speakers in terms of duration, energy and spectral components. Due to this reason, recognition of the speech is difficult in the case of the HI (Jeyalakshmi C, et al., 2014). If the HI speeches are applied directly on normal models, there is no guarantee that the accuracy could be reasonably good and it is practically not possible. Alternatively, we can use the additional preprocessing technique to obtain clean normal speech by way of doing effective comparison between preprocessed HI speeches and clean NH speeches. In this work, RLS filtering is used as additional preprocessing technique to obtain clean normal speech by comparing the Recursive least square RLS filtered HI speech with clean speeches taken for analysis. Energy is computed for filtered speeches and clean speeches taken for analysis and energy differences are found out respectively. Among the NH speeches, the best NH speech is chosen based on the minimum energy difference. These normal hearing speeches are considered as representative of HI speeches from which, features are extracted and applied to the models developed using speeches of NH speakers. So, the system behaves like NH Vs. NH and the system becomes robust for recognizing the speeches of HI.

Speech recognition system using MF-PLP and K means clustering technique

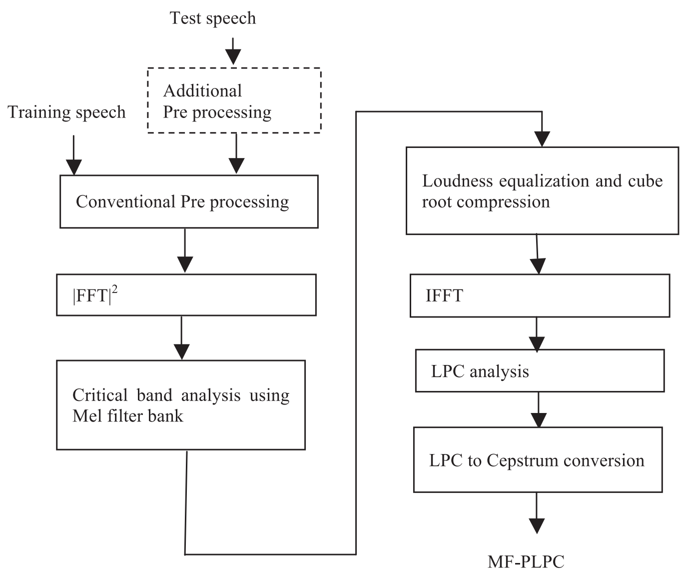

Fig.3. MF-PLP feature extraction model

In general, for speech recognition applications such as normal and emotional speech recognition, MFCC features are used (Murty KSR &Yegnanarayana B,;Tin Lay New et al.,2006.2003). This speech recognition system for HI is developed using MF-PLPC as a feature and K means clustering for developing training models. The detailed procedure used for MF-PLPC (Revathi& Venkataramani2011)feature extraction is shown in Figure 3.

Training and testing based on clustering technique

The way in which L training vectors can be clustered into a set of M code book vectors is by K-means clustering algorithm (Rabiner, & Juang 1993). Block diagram for K-means clustering using vector quantization (VQ) is shown in Figure 4.

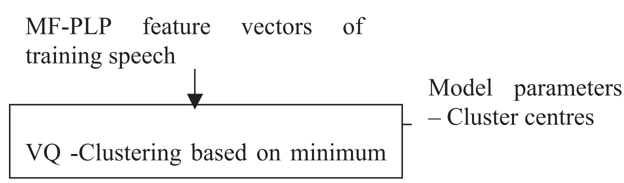

Fig.4. Basic block diagram of the basic VQ training structure

Figures 5 and 6 indicate the hierarchical binary cluster tree for the training data of the sentences 'elundunil' and 'ennasapitai'.

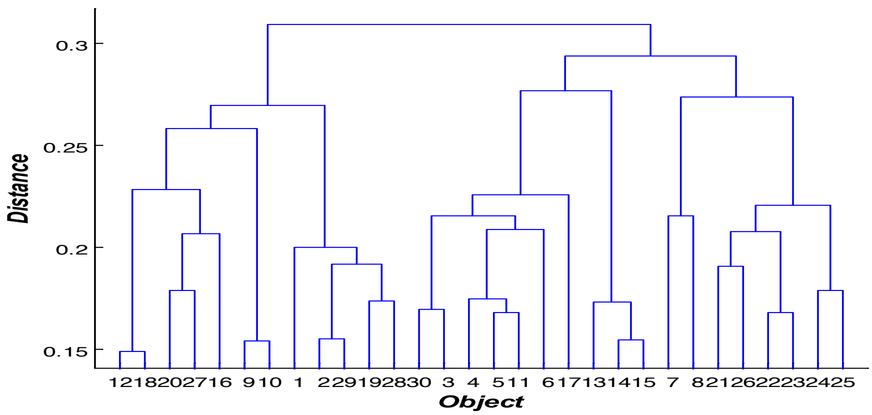

Fig.5. Dendrogram plot – Hierarchical binary cluster tree for 30 training vectors for the sentence 'elundunil'

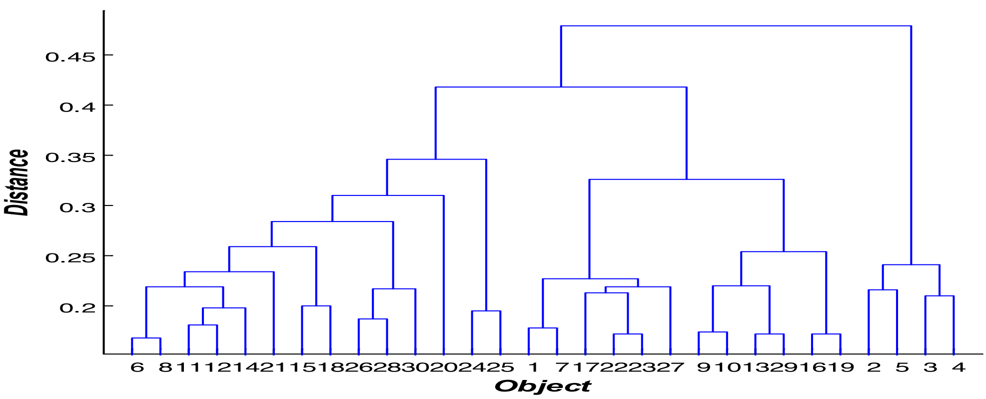

Fig.6. Dendrogram plot – Hierarchical binary cluster tree for 30 training vectors for the sentence 'ennasapitai'

Dendrogram plots reveal the production of distinct set of clusters for speeches. Speech recognition system involves extraction of features from the training data building VQ codebook models for all digits (0-9) and simple sentences and testing each utterance against a certain number of speech models to detect the identity of the speech of that utterance from among the speech models. Speaker independent speech recognition system is evaluated using the data for isolated digits and simple sentences. Ten NH speakers of any age are involved in the process of creating database for training. They are required to utter the isolated digits or simple sentences in Tamil and speeches are recorded using high quality microphone. The limitation with the development of training models using the speeches of HI is that it is difficult to create a large speech database for HI. An alternative is to use the HI speeches only for testing and speeches of NH for developing training models. For creating a training model, speeches of NH speakers corresponding to the respective isolated digit or continuous speeches are concatenated and the resultant signal is pre-processed to eliminate silence and low energy frames and pre-emphasized using a difference operator. Hamming window is applied on differenced speech frames of 16msecs duration with overlapping of 8 msecs. Then, MF-PLPC feature is extracted (Hynek Hermansky et al., 1986, 1991, 1994). For each speech VQ codebook model is developed, based on K-means clustering procedure with perceptual features as input. In this algorithm, there is a mapping from L training vectors into M clusters. Each block is normalized to unit magnitude, before giving as input to the model. Testing procedure used for testing the hearing impaired speeches is shown in Figure 7. RLS filtering is done on HI test speeches with normal speech (NH-1 to NH-10) considered as reference speech. Energy is computed for the normal speeches and RLS filtered speeches (FS-1 to FS-10) and the energy difference is found out. NH speechthatshows minimum energy difference is appropriately selected. This is illustrated in Table 2.

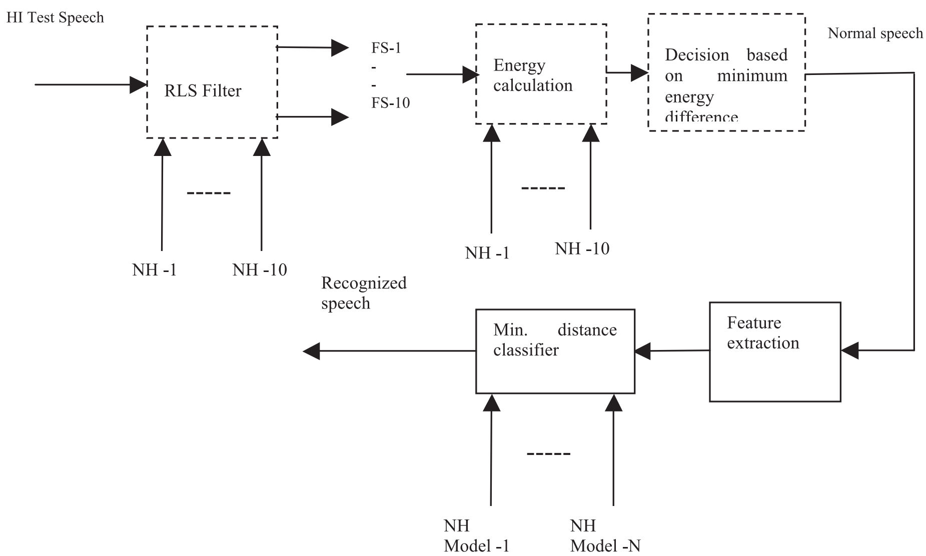

Fig.7. Block diagram for classification of HI test speeches

Table 2. Selection of normal speech based on energy difference

HI test Speech index	NH speech index for RLS filter of order 4									
	1	2	3	4	5	6	7	8	9	10
1	10.3	10.5	10.2	9.8	10.2	11.1	10.1	9.9	10.1	8.5
2	11.9	11.0	9.6	10.4	14.2	10.6	8.5	9.9	11.1	10.3
3	13.7	13.2	13.7	12.1	12.4	11.1	12.6	12.0	10.0	9.3
4	11.4	11.7	8.4	9.2	12.5	10.6	10.0	10.0	9.0	8.5
5	12.3	13.8	13.1	13.4	12.6	13.1	13.9	12.3	11.4	10.3
6	11.6	10.6	8.5	10.3	11.0	9.4	9.9	9.9	10.4	9.1
7	9.1	8.8	11.2	11.3	13.5	10.2	10.4	9.6	10.3	10.4
8	11.7	13.0	12.4	10.7	14.3	12.3	10.8	12.6	10.2	10.7
9	10.8	9.9	11.9	11.9	14.5	11.4	11.3	10.1	12.0	11.7
10	10.8	9.9	11.9	11.9	14.5	11.4	11.3	10.1	12.0	11.7
11	12.6	10.8	11.0	11.3	12.4	10.3	10.3	11.1	12.7	11.9
12	10.2	10.5	12.3	11.6	12.1	9.6	10.2	9.8	10.0	11.2
13	21.7	23.0	49.9	29.6	49.4	27.8	23.4	23.5	19.3	16.2
14	11.1	11.7	10.8	11.3	18.0	11.2	9.4	11.3	10.3	9.9
15	9.4	11.4	10.6	9.8	16.7	10.6	9.0	10.6	10.5	10.2
16	9.4	11.4	9.8	10.2	15.6	12.1	10.0	9.1	7.8	9.6
17	6.7	9.6	10.0	10.3	15.0	10.8	9.1	8.4	9.1	8.3
18	26.0	26.3	49.1	29.6	49.9	27.4	26.5	26.9	19.1	16.8
19	14.3	13.8	11.2	15.2	14.4	14.8	13.3	12.9	15.3	15.4
20	16.3	16.2	15.3	15.9	13.9	16.0	15.8	16.5	12.3	13.7
21	15.4	15.8	14.3	12.7	12.8	13.2	13.0	15.1	15.2	15.9
22	15.4	15.8	14.3	12.7	12.8	13.2	13.0	15.1	15.2	15.9
23	14.9	15.4	13.5	14.1	15.7	14.4	14.9	14.6	14.9	16.8
24	14.3	14.7	12.2	14.1	15.6	15.0	15.0	13.8	13.9	15.8
25	10.6	11.7	13.6	12.7	10.1	11.0	13.0	12.8	11.3	10.3
26	11.2	12.4	10.5	10.8	13.2	11.6	12.4	12.2	10.9	11.7
27	12.4	13.2	13.6	13.0	11.3	12.4	12.7	13.0	13.1	10.3
28	11.3	11.3	13.2	11.7	13.2	11.4	11.8	12.0	11.4	10.5
29	13.5	11.9	11.4	12.5	14.5	11.9	10.5	11.6	12.8	10.9
30	12.6	12.2	10.6	12.2	15.6	13.4	12.5	13.1	12.0	12.4
31	9.2	12.0	11.7	9.8	10.8	10.6	8.5	11.3	11.9	13.5
32	9.1	11.7	12.4	10.3	12.8	11.1	9.3	10.5	12.7	13.4
33	10.7	12.5	8.9	9.7	9.5	11.2	10.3	11.2	11.3	12.3
34	10.7	11.1	12.4	10.0	11.3	11.5	8.6	10.1	11.3	13.5

35	9.7	12.2	11.0	9.6	12.7	10.7	8.7	11.2	12.3	13.9
36	13.0	10.7	11.3	11.5	9.6	11.2	10.3	11.1	11.3	13.2
37	11.7	10.9	10.6	10.2	10.7	11.6	9.7	11.4	9.7	11.5
38	11.3	8.5	9.2	11.6	11.5	11.4	9.3	9.3	11.1	9.6
39	11.7	11.4	11.0	9.3	11.6	10.9	9.6	11.6	11.2	11.4
40	10.4	10.1	10.2	7.5	7.0	8.1	7.2	8.2	11.5	9.3
41	12.3	8.0	10.6	9.6	8.8	9.5	7.5	8.7	12.4	9.2
42	11.5	10.1	10.5	11.1	9.5	12.2	11.3	9.2	11.9	10.2
43	7.7	8.8	9.0	8.0	12.7	9.0	8.1	8.0	9.9	13.3
44	8.7	9.3	11.6	9.0	11.3	9.8	8.6	9.5	11.6	13.8
45	9.5	8.6	11.0	8.8	12.3	9.0	8.1	8.2	10.0	11.1
46	10.0	8.3	11.0	7.8	13.8	10.1	8.2	8.4	10.3	13.4
47	8.8	8.2	10.1	9.0	10.2	8.8	7.3	7.6	9.8	11.0
48	9.7	8.5	10.2	7.8	11.5	9.1	7.9	8.0	9.2	12.5
49	11.4	12.5	14.1	12.4	13.6	13.1	12.4	11.9	12.8	9.8
50	12.5	11.5	14.9	9.7	11.3	11.1	10.6	10.1	13.7	9.8
51	9.8	10.2	11.0	12.7	15.0	13.0	12.1	9.0	11.3	9.6
52	12.3	9.8	13.2	9.8	13.2	10.3	9.3	9.3	11.9	9.3
53	12.3	13.3	11.5	12.3	17.5	12.3	11.5	12.5	10.2	10.1
54	12.7	11.8	14.1	8.6	16.7	11.0	8.4	10.6	9.6	11.6
55	7.6	8.9	6.8	9.4	7.0	8.5	7.8	8.7	8.7	10.5
56	8.4	9.9	7.1	8.9	8.0	9.9	8.0	8.3	9.2	11.0
57	8.3	10.2	7.7	9.1	11.0	8.7	8.0	9.1	9.5	9.1
58	9.6	11.4	9.1	9.0	9.5	9.9	8.7	10.1	10.2	11.4
59	8.3	6.9	7.8	8.0	11.6	7.8	6.5	5.6	8.7	9.8
60	9.1	11.1	8.2	8.8	12.3	9.8	8.6	10.2	9.6	11.9

The index of NH speech being chosen based on minimum energy difference is shown as bold values in Table 2. Then, the chosen NH speech is pre-processed for removal of silence and low energy frames. Pre-emphasis is done to spectrally flatten the signal and frame blocking is done to convert the signal into overlapped frames of 16msecs duration with 8msecs overlapping. Then, perceptual features are extracted for test speech. Test data can be either isolated digit or simple sentence from the database of HI. MF-PLP features extracted from each test utterance are fed to the claimant models (NH model -1 to NH model-N). Then the minimum distance is found between each test vector and centroids of clusters. Average of minimum distances for each speech model is determined. The test utterance best matches with a speech model which has minimum of averages (Revathi&Venkataramani,2009).

RESULTS AND DISCUSSIONS

Speech recognition rate is the number of correct choices over the total number of test speeches. Performance of the system for isolated digits and simple sentences using K means clustering technique is assessed by varying the cluster size as16,32,64,128,256,512 and 1024 and by varying the RLS filter order. The recognition accuracy is compared by applying the hearing impaired test speeches on each cluster model. Figure 8 indicates the features used for developing the training model for the simple sentence "elundunil'.

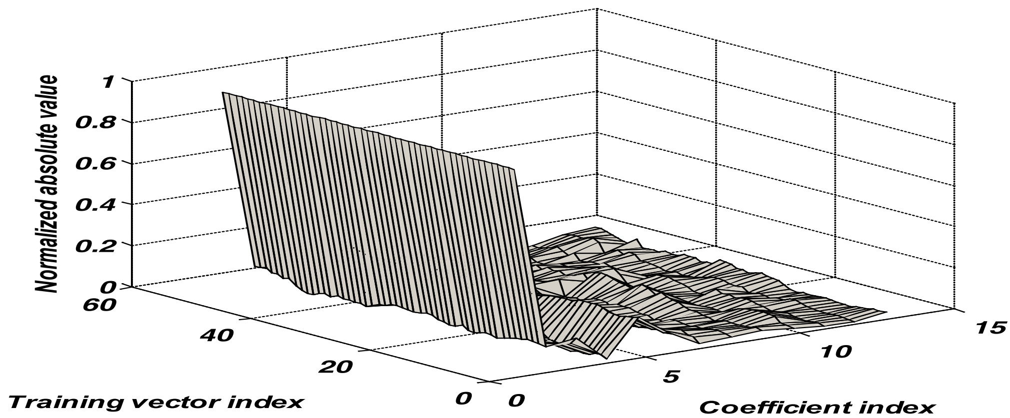

Fig.8. MF-PLPC features for the simple sentence "elundunil' – training

Figures 9 and 10 depict the features of the test speech for the simple sentence "elundunil' without and with RLS filter.

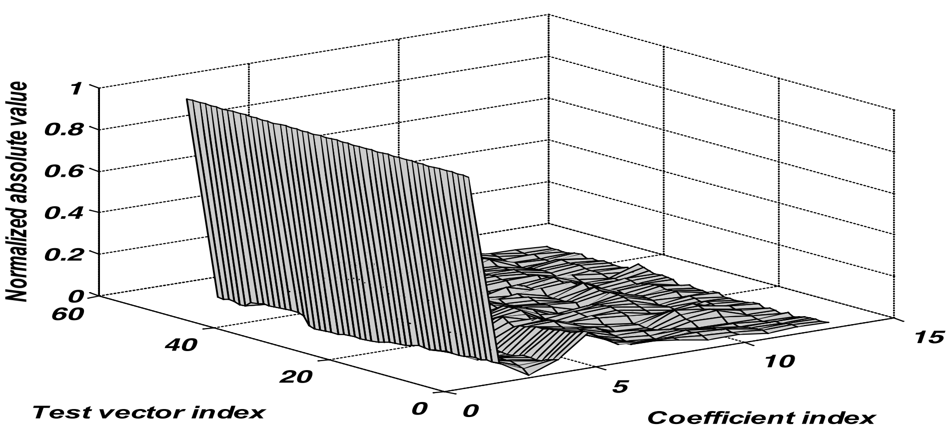

Fig.9. MF-PLPC Features of the test speech for "elundunil' without RLS

From the Figures 8 and 10, it is evident that there is more correlation between the features of the training and test data for the simple sentence 'elundunil' when RLS filtering is performed. Figures11,12 and 13 indicate the efficiency of the testing procedure using RLS filtering adopted for classification with respect to the selection based on minimum of average of minimum distances computed between the features of the test data and the models of cluster size 256 for the first three simple sentences taken for study in this work.

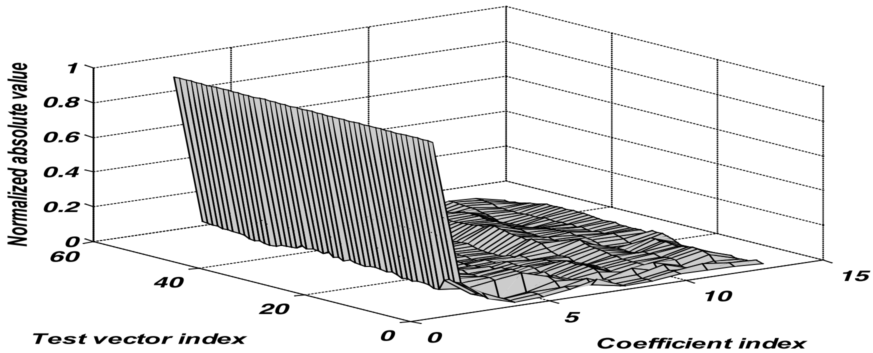

Fig.10. MF-PLPC Features for test speech for 'elundunil' with RLS

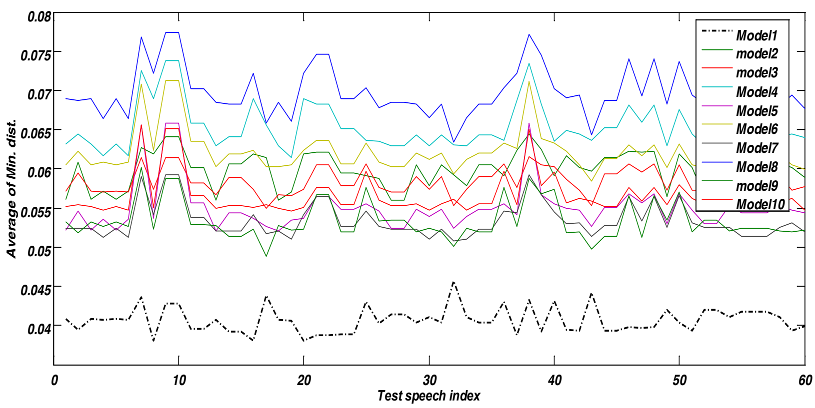

Fig.11. Classification based on minimum of averages for the first simple sentence 'elundunil' – 256 clusters

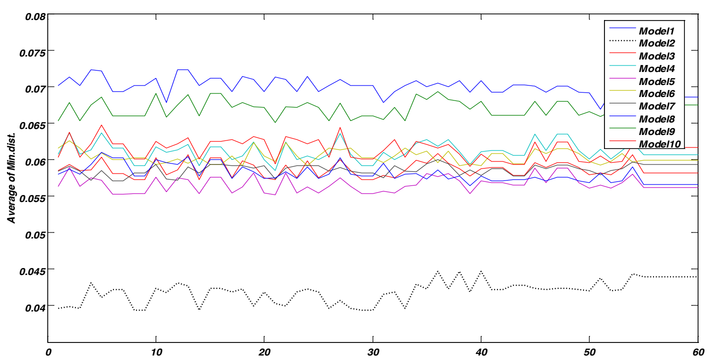

Fig.12. Classification based on minimum of averages for the second simple sentence'engeutkar' – 256 clusters

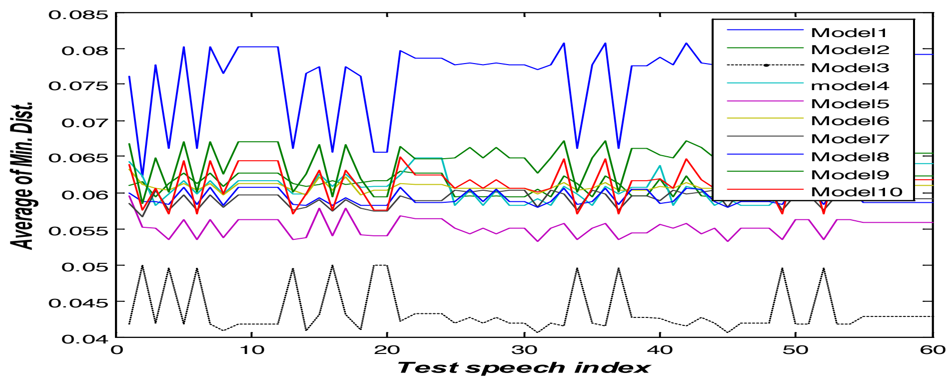

Fig.13. Classification based on minimum of averages for the third simple sentence 'ennasapitai' – 256 clusters

Performance of the system for recognition of isolated digits using various clustering models without RLS filtering and with RLS filter of order 4and 2 is given in Table 3 and Table 4.Performance of the system for recognition of simple sentences using various clustering models without RLS filtering and with RLS filter of order 4 and 2 is given in Table 5 and Table 6.

Table 3. Isolated digit recognition - Testing of HI speeches on NH models for RLS filter of order 4

Digit	%RA													
	Without RLS for various cluster sizes							With RLS for various cluster sizes						
	1024	512	256	128	64	32	16	1024	512	256	128	64	32	16
Ondru	11.7	13.3	11.7	1.7	10	1.7	18.3	100	100	100	100	100	90	90
Irandu	0	0	1.7	0	0	0	0	100	100	100	100	100	100	100
Moondru	10	0	0	0	0	0	0	100	100	100	100	100	100	100
Naangu	0	0	90	5	90	36.7	90	100	100	100	100	100	94	89
Aindu	0	0	0	0	0	0	0	100	100	100	100	100	100	100
Aaru	0	0	0	0	0	78.3	0	100	100	100	100	100	100	100
Yelu	0	0	0	0	0	0	0	100	100	100	100	100	100	15
Yettu	97	0	0	0	0	0	0	100	100	100	100	100	100	99
Onbadu	1.7	0	3.3	3.3	6.7	23.3	0	100	100	100	100	100	100	100
Poojyam	10	94	10	90	0	0	10	100	100	100	100	100	100	100

Table 4. Isolated digit recognition - Testing of HI speeches on NH models for RLS filter of order 2

Digit	**%RA**													
	Without RLS for various cluster sizes							**With RLS for various cluster sizes**						
	1024	**512**	**256**	**128**	**64**	**32**	**16**	**1024**	**512**	**256**	**128**	**64**	**32**	**16**
Ondru	10	13.3	11.7	1.7	10	1.7	20	100	100	100	100	100	77	77
Irandu	0	0	1.7	0	0	0	0	100	100	100	100	100	100	100
Moondru	10	0	0	0	0	0	0	100	100	100	100	100	100	100
Naangu	0	0	90	3.3	90	35	90	100	100	100	100	100	100	99
Aindu	0	0	0	0	0	0	0	100	100	100	100	100	100	100
Aaru	0	0	0	0	0	92	0	100	100	100	100	100	100	94
Yelu	0	0	0	0	0	0	1.7	100	100	100	100	100	85	15
Yettu	97	0	0	0	0	0	0	100	100	100	100	100	100	100
Onbadu	1.7	0	3.3	3.3	6.7	23.3	0	100	100	100	100	100	100	100
Poojyam	10	97	10	90	0	1.7	10	100	100	100	100	100	100	100

Table 5. Continuous speech recognition - Testing of HI speeches on NHmodels for RLS filter of order 4

Simple sentence	**%RA**													
	Without RLS for various cluster sizes							**With RLS for various cluster sizes**						
	1024	**512**	**256**	**128**	**64**	**32**	**16**	**1024**	**512**	**256**	**128**	**64**	**32**	**16**
Elundu nil	0	0	0	0	0	0	0	100	100	100	100	100	100	100
Engeutkar	0	3.3	3.3	0	1.7	1.7	3.3	100	100	100	100	100	100	100
Ennasapitai	0	0	37	15	5	5	8.3	100	100	100	100	100	100	100
Kalaivanakkam	0	0	1.7	0	0	0	0	100	100	100	100	100	100	90
Nandragapadi	1.7	1.7	0	0	3.3	3.3	0	100	100	100	100	100	100	100
Odiva	0	0	0	0	0	0	0	100	100	100	100	100	100	100
Pallividumurai	0	0	0	0	0	0	0	100	100	100	100	100	100	100
Sattampodathey	8.3	8.3	57	78.3	87	73.3	68.3	100	100	100	100	100	100	100
Seekirameludu	58.3	58.3	15	6.7	1.7	11.7	0	100	100	100	100	100	100	100
Vayaduenna	23.3	3.3	8.3	8.3	6.7	10	5	100	100	100	100	100	100	100

Table 6. Continuous speech recognition - Testing of HI speeches on NH models for RLS filter of order

Simple sentence	%RA													
	Without RLS for various cluster sizes							With RLS for various cluster sizes						
	1024	512	256	128	64	32	16	1024	512	256	128	64	32	16
Elundu nil	0	0	0	0	0	0	0	100	100	100	100	100	100	100
Engeutkar	0	3.3	3.3	0	0	0	3.3	100	100	100	100	100	100	100
Ennasapitai	0	0	25	11.7	5	5	6.7	100	100	100	100	100	100	100
Kalaivanakkam	1.7	1.7	1.7	3.3	1.7	0	3.3	100	100	100	100	100	100	90
Nandragapadi	1.7	1.7	0	0	3.3	3.3	0	100	100	100	100	100	100	100
Odiva	0	0	0	0	0	0	0	100	100	100	100	100	100	100
Pallividumurai	0	0	0	0	0	0	0	100	100	100	100	100	100	100
Sattampodathey	10	10	58.3	77	85	73.3	68.3	100	100	100	100	100	100	100
Seekirameludu	60	58.3	10	5	1.7	11.7	0	100	100	100	100	100	100	100
Vayaduenna	23.3	3.3	8.3	6.7	6.7	10	5	100	100	100	100	100	100	100

Table 7. Isolated digits and Continuous speech recognition - Testing of NH speeches on NH models

%RA													
Isolated Digits for various cluster sizes							Simple sentences for various cluster sizes						
1024	512	256	128	64	32	16	1024	512	256	128	64	32	16
100	100	100	100	100	90	50	100	100	100	100	100	100	100
100	100	100	100	100	100	100	100	100	100	100	100	100	100
100	100	100	100	100	100	60	100	100	100	100	100	100	100
100	100	100	100	100	70	40	100	100	100	100	100	100	90
100	100	100	100	100	100	100	100	100	100	100	100	100	100
100	100	100	100	100	100	90	100	100	100	100	100	100	100
90	90	90	80	70	30	70	100	100	100	100	100	100	100
100	100	100	100	100	100	90	100	100	100	100	100	100	100
100	100	100	100	100	100	100	100	100	100	100	100	100	100
100	100	100	100	100	100	100	100	100	100	100	100	100	100

Results in Table 7 indicate that the individual accuracy for the digit ‘yelu’ is relatively poor for all the clustering models. However, individual accuracy for recognition of simple sentences is 100% for all clustering models except the one with 16 clusters.

From these tables, it is understood that clustering model of 128 cluster size ensures 100% as accuracy for all isolated digits and simple sentences for the RLS filter of order 4 and 2. For testing one speech, this algorithm takes less than a second to display the text information of the speech as visual output or the speech to be heard as the audio output for the RLS filter of order 2 and clustering model of cluster size 32. These tables also reveal that the efficiency of the RLS filtering algorithm in ensuring the best accuracy, whereas, the application of the hearing impaired speeches directly on the NH models has ensured low accuracy. Based on the validation of results with and without RLS filtering, it is evident that the HI verses normal models with RLS filtering is providing the best accuracy for considering the 60 test speeches for each isolated digit and simple sentence in Tamil language.

Thus, the RLS filtering approach has ensured that the HI speeches are correctly recognized by the NH speakers and proper guidance may be given appropriately to the hearing impaired speakers. Hence, this system can be used as voice interactive response system for HI speakers and HI can interact with NH speakers without any hassle.

CONCLUSION

In this paper, performance of the speech recognition system for recognizing the speeches of HI is assessed and analyzed. This is effectively done by applying their speeches to RLS filtering for extracting the appropriate clean NH speech by first finding the energy difference between the HI speech and ten normal speeches. Then the selection is done based on the comparison between energy difference values and the one with minimum energy difference is selected and subsequently features are extracted. These features are applied to the models of NH speakers and the classification is done for isolated digits and simple sentences. It ensures 100% as overall accuracy for cluster models of size 64 and RLS filter of order 2 for isolated digit recognition. Overall accuracy is 100% for recognizing simple sentences by using cluster model of size 32 and RLS filter of order 2. So, RLS filter of order 2can be taken to minimize the time required to select the clean normal speech among the set of corresponding normal speeches. The main difficulty in training HI persons to speak for collecting speech samples is avoided in our work, since speeches of HI are used only for testing. Results indicate that this speech recognition system for HI would enable the HI to communicate with others like normal people. Though, the clustering technique is considered as an old technique, it is proved that it has provided good results. They could be able to operate all voice operated devices without requiring assistance from others. Ultimately, social status of the HI would be greatly improved and they behave socially with everyone like normal persons. It is hoped that this kind of work would facilitate reservation in education and jobs, financial aid and social inclusion for HI citizens. It further enhances their social status and removes stigma and facilitates them to lead a normal life. Even deaf people who excel at lip reading could make use of this system as a translator, when it comes to meetings or group conversations.

REFERENCES

Brain C.J.Moore, 2003. Speech processing for the hearing impaired: successes, failures, and implications for speech mechanisms, Speech communication; **41**: 81-91.

Craig W. Newman, Sharon A.Sandridge, Hearing loss is often undiscovered but screening is easy, Audiology Research Laboratory, Department of Otolaryngology and Communicative Disorders, The Cleveland Clinic Foundation.

Elham S. SalamaReda A. El-khoribi Mahmoud E. Shoman, 2014. Audio-visual speech recognition for people with speech disorders', International Journal of Computer Applications, 96 : 2.

Harry Levitt, Member, 1971. Acoustic analysis of deaf speech using digital processing techniques, IEEE Fall Electronics Conference, Chicago.

HynekHermansky,Kazuhiro Tsuga, Shozo Makino, HisashiWakita, 1986, Perceptually based processing in automatic speech recognition. In: IEEE int. conf. on Acoustics, speech and signal processing; Tokyo: pp.1971-1974.

HynekHermansky, Nelson Margon, ArunaBayya, Phil Kohn, 1991. The challenge of Inverse E: The RASTA PLP method. In: Twenty fifth IEEE Asilomar conf. on signals, systems and computers; November 1991; Pacific Grove, CA, USA: pp.800- 804.

HynekHermansky, Nelson Morgan, RASTA processing of speech, 1994. IEEE transactions on speech and audio processing 2: 578-589.

Jeyalakshmi,C., Revathi, A&Krishnamurthi, V. 2014. Development of speech recognition system in native language for hearing impaired, Journal of Engineering research Kuwait Journal of Science and Engineering,2:81-99.

Jeyalakshmi, C., Revathi, A&Krishnamurthi, V. 2015. Investigation of voice disorders and recognizing the speech of children with hearing impairment in classical Tamil language. Int. J. of Biomedical Engineering and Technology,17:356–370.

Lim Sin Chee, Ooi Chia Ai, M.Hariharan&SazaliYaacob, 2009. MFCC based recognition of repetitions and prolongations in stuttered speech using K-NN and LDA, Proceedings of SCOReD,vol.16-18, (IEEE), pp 416–419.

Mark S. Hawley, Stuart P. Cunningham, Phil D. Green, Pam Enderby, Rebecca Palmer, SiddharthSehgal, and Peter O'Neill, 2013, 'A Voice-Input Voice-Output Communication Aid for People With Severe Speech Impairment', IEEE transactions on neural systems and rehabilitation engineering, vol. 21, no. 1, pp.23-30.

MohdHafidzMohamadJamil, Al-Haddad,SAR &CheeKyun Ng, 2011, 'A Flexible Speech Recognition System for Cerebral Palsy Disabled', International conference on Informatics Engineering and information science, Springer, vol. 251, pp. 42-55.

Murty, K.S.R., &Yegnanarayana. B, 2006, Combining evidence from residual phase and MFCC features for speaker recognition, IEEE Signal Processing Letters,13:1, 52–55.

Nickerson RS, Stevens KN, Boothroyd A Rollins A, 1974, Some Observations on Timing in the Speech of Deaf and Hearing Speakers; Report No. 2905.

Prasad D Polur, Gerald E Miller, 2005, Effect of high frequency spectral components in computer recognition of dysarthric speech based on a Mel-cepstral stochastic model, Journal of Rehabilitation research and development 42: 363-372.

Prasad D Polur, Gerald E Miller, 2006, Investigation of an HMM/ANN hybrid structure in pattern recognition application using cepstral analysis of dysarthric (distorted) speech signals. Medical engineering & Physics 28: 741-748.

Picjlett J. M., Some Applications of Speech Analysis to Communication Aids for the Deaf 1969,IEEE Transactions on Audio and Electro acoustics 17: 283-289.

Pickett JM, Costam A, 1968, A visual speech trainer with simplified indication to vowel spectrum. American Annals of the Deaf , 113: 253–258.

Rabiner, L. R., and B. H. Juang, 1993, Fundamentals of Speech Recognition, Prentice Hall, New Jersey.

Ruscello DM, Sholtis DM, Moreau VK, 1980, Adult's awareness of certain articulatory gestures. Perceptual andMotor Skills,50: 1156–1158.

Revathi A, Venkataramani Y, 2012, Evaluate multi-speaker isolated word recognition using concatenated perceptual feature. EE Times-India ; 1-10.

Revathi A, Venkataramani Y, 2011, Speaker independent continuous speech and isolated digit recognition using VQ and HMM. International Conference on Communications and Signal Processing (ICCSP); Calicut: pp.198-202.

Revathi A, Venkataramani Y, 2009, Perceptual features based isolated digit and continuous speech recognition using iterative clustering approach. In: First International Conference on Networks and Communications; 27-29; Chennai: pp. 155-160.

Tin Lay Nwe, Say Wei Foo, Liyanage C, De Silva, 2003, Speech emotion recognition using Hidden markov models. Speech communication, 41, 603-623.

Xu Wang, Zhiyan Han, JianWang&MingtaoGuo, 2008, 'Speech Recognition System Based on Visual Feature for the Hearing Impaired' Proceedings of fourth International Conference on Natural Computation, vol.2, pp. 543-546.

Yoshinori Yamada, Hector Javkin, Karan Youdelman, 2000, Assistive speech technology for persons with speech impairments, Speech communication, 30: 179-187.

Submitted: 10/Feb/2016
Revised : 6/July/2016
Accepted : 28/Sept/2016