International Journal of Engineering Sciences & Emerging Technologies, June 2014.
ISSN: 2231 – 6604 Volume 6, Issue 6, pp: 468-474 ©IJESET

VOWEL ARTICULATION TRAINING SYSTEM FOR HEARING IMPAIRED CHILDREN AS AN ASSISTIVE TOOL

Rupali.S.Kad[1], R.P.Mudhalwadkar[2]
[1]Department of Electronics &Telecommunication, PCCOE, Pune University, M.S., India
[2]Department of Instrumentation & Control Engineering,
COEP, Pune University, M.S., India

ABSTRACT

A vowel articulation training system for hearing impaired children which uses a MATLAB based GUI interfaced with microcontroller has been developed. The system provides display which gives visual information for spoken vowels i.e. whether pronounced vowel is correct y or not. We discuss the development of vowel training system for hearing impaired children specifically children aged between 5 and 10 years whose mother tongue is marathi. Formats emerged from vocal tract depends on the position of jaws, tongue and other and the shape of your mouth opening. Vowels in English are determined by how much the mouth is opened, and where the tongue constricts the passage through the mouth: front, back or in between parts of the vocal tract and also how you position your tongue. Formant range is different for same vowel for different ascent. We have collected 750 vowel samples from normal speakers. We discuss the methodology of collection of vowel database & vowel recognition results using the lpc method. The system typically provides correct identification of over 80% of steady state vowels spoken by both gender speakers

KEYWORDS: *Feature extraction, LPC, Vowel Database, GUI*

I. INTRODUCTION

Hearing is the sense by which humans acquire their spoken words. Through the language a child learns about the surrounding world, about him/herself and intellectual development takes place. The hearing impairment can range from a mild to profound loss of hearing or deafness. Hearing impairment in children causes a major problem in language development and requires special attention**.** Proposed system is an assistive tool for hearing impaired children .A vowel is a speech sound made by the vocal cords. The letters of the English alphabet are either vowels or consonants or both. Vowels form the basic block for word formation. It is the main constituent block for word pronunciation. A vowel sound comes from the lungs, through the vocal cords, and is not blocked, so there is no friction. All English words have vowels. Each spoken word is created out of the phonetic combination of a limited set of vowel and consonant .Presentation of speech signal in the frequency domain are of great importance in studying the nature of speech signal and its acoustic properties .The prominent part of speech signal spectrum belongs to the formants corresponds to the vocal tract. Formants are exactly resonant frequencies of vocal tract when pronouncing vowel.. Acoustic resonances in the vocal tract can produce peaks in the spectral envelope of the output sound. In speech science, the word 'formant' is used to describe either the spectral peak or the resonance that gives rise to it. Many researchers have worked in this regard. Some commercial software is also available in the market for speech recognition, but mainly in American English or other European languages.
This paper is divided into six sections. Section 1 gives Introduction. Section 2 deals with details of formation of vowels database. Section 3 focuses on system implementation, Section 4 covers result section 5 deals conclusion followed by references.

II. VOWEL DATABASE FORMATION

Database was formed from a total 50 individuals consisting children from both gender from Municipal schools and apartments. The speakers were children who had no obvious speech defects. The recordings were done using a microphone and a laptop with a sampling frequency of 8000Hz. The vowels were recorded using Omni directional microphone using the sound wave recorder. The samples were recorded in closed room where background noise was not present. The speakers were seating in front of the direction of the microphone with the distance of about 1-3 cm..

- Vowels: a, e, i, o, u
- Female vowel samples:675
- Male vowel samples :575

Database was formed with the samples of 23 male and 27 female normal speakers of 5-10 yrs age. Mother tongue of both the speakers was Marathi. Each speaker was asked to speak the 5 vowels with 5 utterances of each vowel. Total 25 utterances of the vowels were recorded for each speaker

III. SYSTEM IMPLEMENTATION

The vowel recognition system consists of vowel recognition process , GUI development & hardware environment.

3.1. Vowel recognition process

The main steps involved in vowel recognition process are recording of vowel, preprocessing of signal &feature extraction.

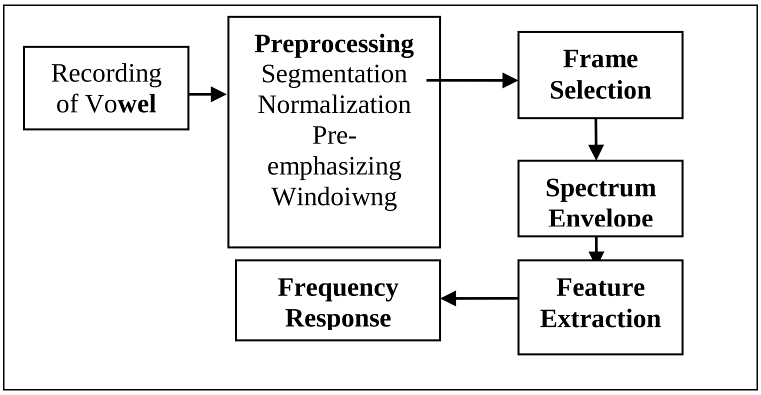

Figure 1. Vowel recognition process

3.1.1.Recording of vowel: All vowel samples were recorded with omnidirectional microphone.To avoid noise while recording the sample, the optimal distance between microphone and speaker is found out.To do this powerspectrum of the different audio signal which was output of audio amplifier at different distances were observed.

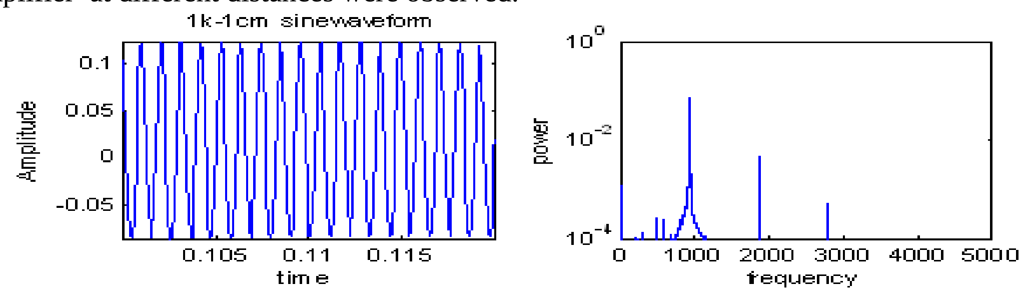

Figure 2. Power spectrum of 1KHz Audio signal at 1cm distance

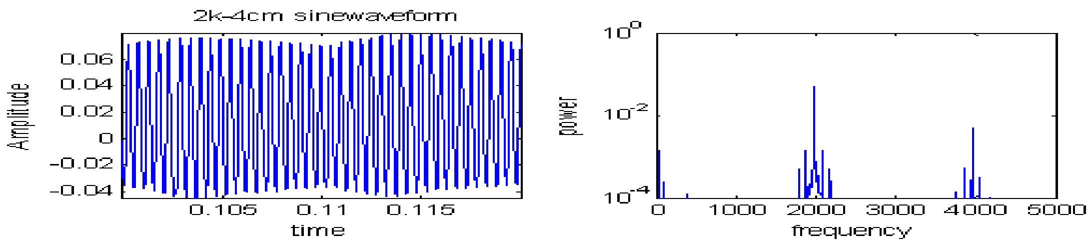

figure 3. Power spectrum of 2KHz Audio signal at 4cm distance

Fig 3 shows distortion at fundamental as compared to Fig 2. The speakers were seating in front of the direction of the microphone with the distance of about 1-3 cm

3.1.2. Pre Processing : Preprocessing is a process of segregation of voiced region from the silence/unvoiced portion of the captured signal. Extraction of the voiced part of the speech signal by marking and/or removing the silence and unvoiced region leads to substantial reduction in computational complexity at later stages[4]. A program was developed to extract the vowel portions of the signal based on energy computation. First, the voice signal was recorded using a laptop and a microphone using a sampling frequency of 8000Hz. The DC portion of the which were introduced by the recording equipment is removed and the resultant signal was then normalized. DC portion carries no useful information. Moreover, can carry disturbing information during energy calculation.Normalization is carried out to minimize inter speaker variation due to physiological or anatomical difference.[2]The start and endpoint locations were determined using only energy method.

3.1.3 Feature Extraction*:*LPC provides a good model of speech signal. LPC is an analytically tractable model. The method of LPC is mathematically precise and simple and straightforward to implement in either software or hardware. The computation involved in the LPC processing is considerably less than that required for an all digital implementation of bank of filter models. LPC analyzes the speech signal by estimating the formants, removing their effects from the speech signal, and estimating the intensity and frequency of the remaining buzz. The process of removing the formants is called inverse filtering, and the remaining signal after the subtraction of the filtered modeled signal is called the residue.[1]

The basic steps of LPC processor include the following:

1. *Preemphasis*: The digitized speech signal, *s*(*n*), is put through a low order digital system, to spectrally flatten the signal and to make it less susceptible to finite precision effects later in the signal processing. The output of the preemphasizer network, , is related to the input to the network, *s*(*n*) , by difference equation:

$$\tilde{s}(n) = s(n) - \tilde{a}s(n-1)$$

2. Frame Blocking: The output of preemphasis step, $\tilde{s}(n)$, is blocked into frames of N samples, with adjacent frames being separated by M samples. If x $_l$(n) l is then frame of speech, and there are L frames within entire speech signal, then

$$x_l(n) = \tilde{s}(Ml + n)$$

3 *Windowing*: After frame blocking, the next step is to window each individual frame so as to minimize the signal discontinuities at the beginning and end of each frame. If we define the window as *w*(*n*), 0 ≤ *n* ≤ *N*– 1, then the result of windowing is the signal:

$$\tilde{x}_l(n) = x_l(n)w(n)$$

where $0 \le n \le N-1$

Typical window is the Hamming window, which has the form

$$w(n) = 0.54 - 0.46\cos\left[\frac{2\pi n}{N-1}\right] \qquad 0 \le n \le N-1$$

4. *Autocorrelation Analysis*: The next step is to auto correlate each frame of windowed signal in order to give

$$r_l(m) = \sum_{n=0}^{N-1-m} \tilde{x}_l(n)\tilde{x}_l(n+m) \qquad m = 0,1,\ldots,p$$

where the highest autocorrelation value, p, is the order of the LPC analysis

5. *LPC Analysis*: The next processing step is the LPC analysis, which converts each frame of p + 1autocorrelations into LPC parameter set by using Durbin's method. This can formally be given as the following algorithm:

$$E^{(0)} = r(0)$$

$$k_i = \frac{r(i) - \sum_{j=1}^{i-1} \alpha_j^{i-1} r(|i-j|)}{E^{i-1}} \qquad 1 \le i \le p$$

$$\alpha_i^{(i)} = k_i$$

$$\alpha_j^{(i)} = \alpha_j^{(i-1)} - k_i \alpha_{i-j}^{(i-1)} \qquad 1 \le j \le i\text{-}1$$

$$E^{(i)} = (1 - k_i^2) E^{i-1}$$

By solving above equations recursively for $i = 1,2,\ldots,p$, the LPC coefficient, a_m , is given as $a_m = \alpha_m^{(p)}$

6. LPC Parameter Conversion to Cepstral Coefficients: LPC cepstral coefficients, is a very important LPC parameter set, which can be derived directly from the LPC coefficient set. The recursion used is

$$c_m = \sum_{k=m-p}^{m-1} \left(\frac{k}{m}\right) \cdot c_k \cdot a_{m-k} \qquad m > p$$

The LPC coefficients are the features that are extracted from voice signal and these coefficients are used to plot frequency response. Formants are determined from the same.

3.2.Graphical User Interface Development

Graphical user interface (GUI) is a type of user interface that allows users to interact with electronic devices through graphical icons and visual indicators such as labels or text. This MATLAB application is self-contained MATLAB programs with GUI front ends that automate a task or calculation. The GUI typically contains controls such as menus, toolbars, buttons, and sliders. MATLAB based GUI is developed to record the .wav file or load the .wav file from destination. Toggle buttons are used to select vowel to be analyzed. If utterance of vowel is correct it is displayed on vowel text box otherwise message for incorrect utterance is displayed .Structure of graphical user interface is shown in figure 4.

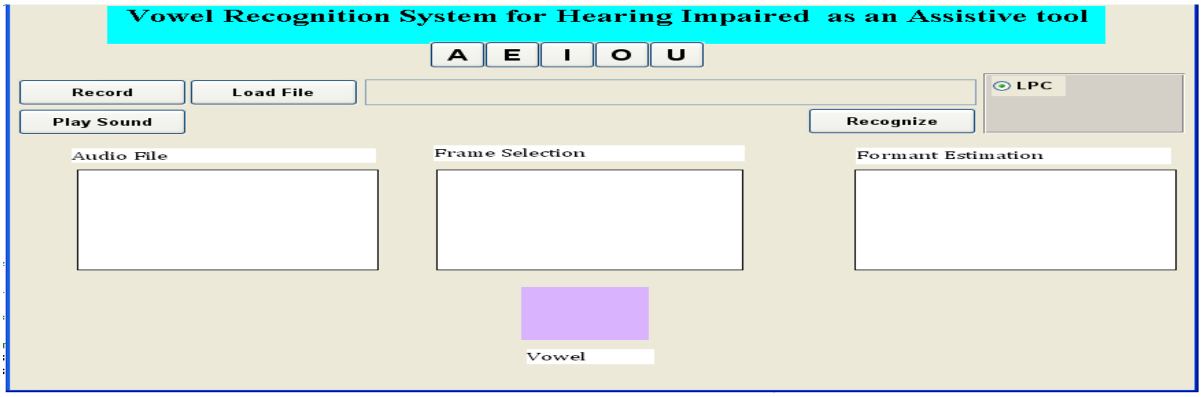

Figure 4. Graphical User Interface

ISSN: 2231 – 6604 Volume 6, Issue 6, pp: 468-474 ©IJESET

3.3.Hardware Enviornment

The hardware environment for this paper consists of a P89V51 microcontroller, a PC, a RS232 driver/receiver, and a DB-9 serial cable. The Philips microcontroller development board is interfaced with PC. The PC is used to write user specified embedded programs to be executed by the Philips microcontroller. Furthermore, the PC hosts an interactive GUI for the user to record and load audio file and visualize pronounced vowel. The microcontroller and the PC communicate using a serial interface. In this paper, we use a P89V51RD2BN, 40-pin, 8-bit CMOS FLASH dual inline package IC. To facilitate serial communication between PIC and PC, we interface a RS232 driver/receiver with the P89V51RD2BN. The effectiveness of our MATLAB -based GUI environment to interact with PIC microcontroller is demonstrated by exporting analyzed vowel of speaker from a MATLAB GUI interfaced to the PC

IV. EXPERIMENTAL/SIMULATION RESULTS

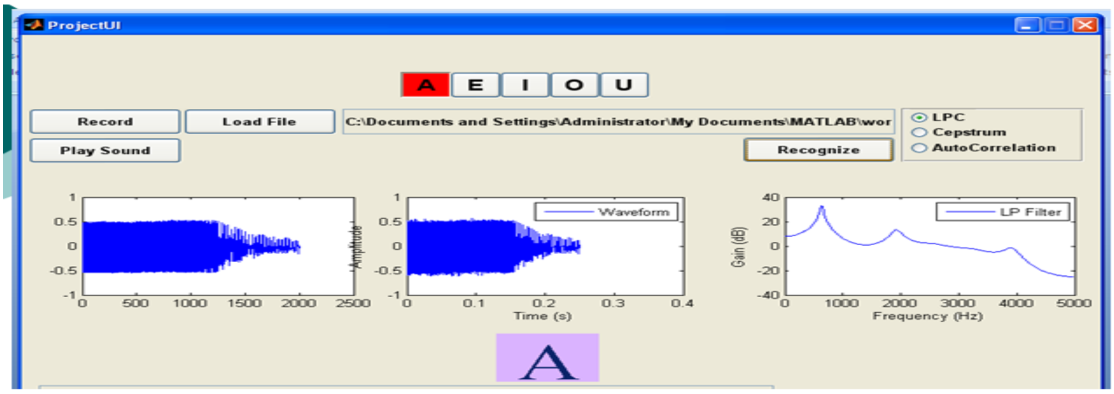

Figure 5.Recognition of vowel /a.

.wav file of vowels are analyzed .Formant ranges for vowels are determined.Fig.5&Fig.6 shows that vowel /a& vowel /e are pronounced correctly.

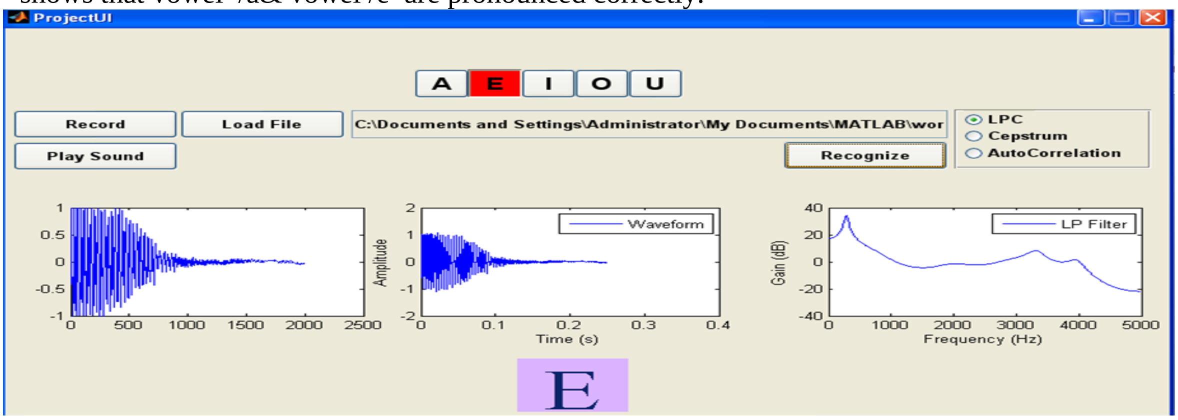

Figure 6. Recognition of vowel /e

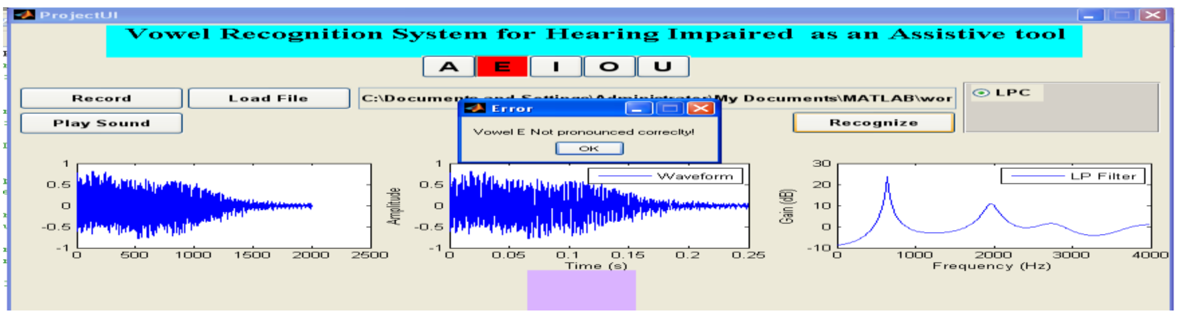

Figure 7: Incorrect Pronunciation of vowel /e

Fig 7. Shows result of utterance of vowel /e of hearing impaired child. Since formants are different as that of normal speaker system shows displays that vowel is not pronounced correctly.

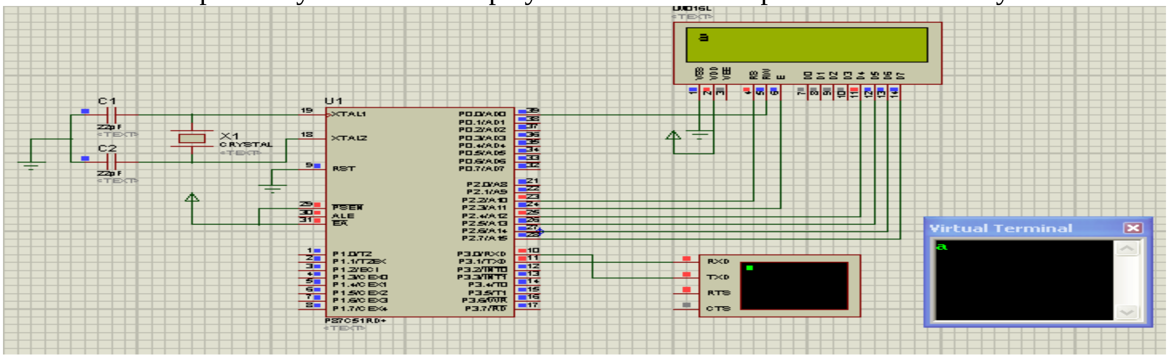

Figure 8: Simulation Result of Microcontroller PC interfacing

Fig.8 shows PROTEUS simulation result of microcontroller interfacing with PC

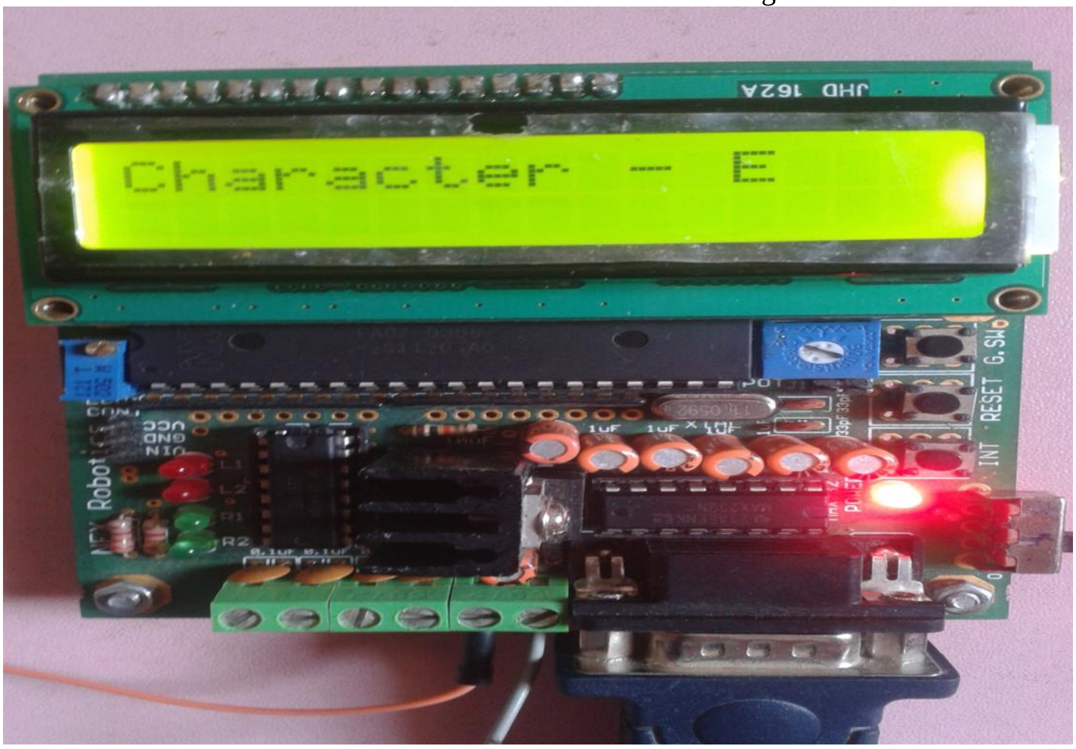

Figure 9: Microcontroller Board

.Fig.9 shows that vowel /e is pronounced correctly.

Table 1. Recognition of vowels

Vowel	*Number of samples*	*Samples Recognized*	*Samples Not recognized*	*Correct Recognition*
/a	135	109	26	80.7
/e	130	107	23	82.30
/i	135	105	30	77.77

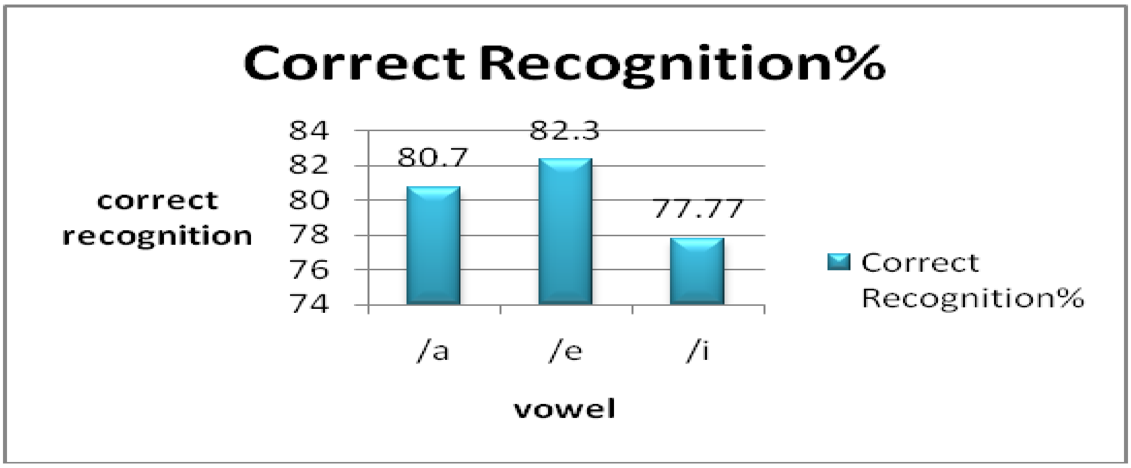

Figure 10: Recognition of vowels

V. CONCLUSION

From experimental results, it can be concluded that LPC can recognize the speech signal well. The highest correct recognition achieved is 82.30%. For further work, in order to get better recognition another recognition method such as ANN or neuro-fuzzy method can be applied in this system. The low cost and good performance of this system indicate that developed system will be useful in vowel training of hearing impaired children as an assistive tool .As compared to the common multimedia sound card which adds significant noise above system has potential for speech training at home for the hearing impaired children which are in the age group of 5-10yrs.

REFERENCES

[1]Shahrul Azmi M.Y., "*An improved feature extraction method for Malay vowel recognition based on spectrum data*," International Journal of Software Engineering and Its ApplicationsVol.8, No.1 (2014), pp.413-426

[2]Y. A. Alotaibi and A. Hussain, "Comparative analysis of Arabic vowels using formants and an automatic speech recognition system," International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 3, 2010.

[3] L. Rabiner and B. Juang, *Fundamentals of speech recognition*. Prentice Hall, 1993.

[4] Ayaz Keerio, Bhargav Kumar Mitra, Philip Birch,Rupert Young, and Chris Chatwin,*"On Preprocessing of Speech Signals,"* International Journal of Signal Processing vol.5,2009,pp.216-222

[5] Rabiner. L. R., Schafer. R. W., "Digital Processing of SpeechSignals", First Edition, Prentice-Hall.

[6] X. Huang, A. Acero and H. Hon, "*Spoken language processing: A guide to theory, algorithm, and system development",* Prentice Hall PTR Upper Saddle River, NJ, USA, **(2001)**.