An Interactive Speech Therapy Session using Linear Predictive Coding in Matlab and Arduino

R.Vijayalakshmi
PG scholar, Department of EEE
S.A. Engineering College
Chennai, India
viji1192@gmail.com

S.Priya
Professor and Head, Department of EEE
S.A. Engineering College
Chennai, India
priyasakthikumar@gmail.com

***Abstract*— Speech Therapy has become an efficient tool to bring back proper speech for patients suffering from various speech disorders. Patients are more benefitted when the speech therapy session is interactive where there is a visible change in the environment. Hence, the proposed system aims at manipulating devices when the user input is correct and also indicates if the user input is incorrect. Speech recognition has been done using the concept if Linear predictive coding and Arduino Uno board is used for hardware interface.**

***Index Terms*— Speech therapy, Arduino, Linear Predictive Coding, Speech processing, ATmega328 microcontroller**

I. INTRODUCTION

Speech plays an important role in our lives. When speech is impaired it creates a problem in day to day activities. Speech disorders occur due to various physical and psychological reasons. Speech disorders may be caused due to problems in the parts responsible for speech production (Eg. Stuttering) , other medical conditions like Stroke, Cleft lip, Autism, and language disorders like Selective Mutism.

Speech therapy helps people overcome their difficulty in speaking and pronunciation. When we automate the speech therapy session, it reaches out to wider audience and benefits a larger section of the society. The heart of the proposed system is the speech recognition module. Speech recognition is done in many ways using various methods.

The technique to convert oral lectures into text is analysed in [1]. This proves beneficial for students in the classroom. Speech recognition mediated lecture acquisition namely real time captioning and postlecture transcription were evaluated. These methodologies consist of speech recognition engine, error correction methods, transcription display methods and transcription distribution methods. It was useful for students who could not take notes efficiently.

Audio-visual speech recognition where noise is removed is explained in [2]. The audio and video part of the speech is processed separately to remove noise for efficient operation. Problems may occur due to environmental audio noise or video compression. Baye's theorem has been using for noise processing.

The study done in [3] shows that first and second order differences of harmony features also play an important role in speech emotion recognition. The study uses mel frequency cepstral coefficients in a Chinese language database.

Feature extraction for noise robust automatic speech recognition is discussed in [4]. This concept concentrates on estimation of minimum mean square error. Here, minimum mean square error of mel frequency cepstral features is taken. The drawback of the system is that these methods relied on stationary noise sources.

The technique mentioned in [5] combines active learning and semi supervised learning techniques. It proves to be better than the two methods separately. Tests were conducted using initial supervised training instances and two different training sets. Small data was only used in this initiative.

Speech controlled devices form the main idea in [6]. Speech recognition has been done in matlab and using arduino and zigbee the devices are interfaced. Here, a direct application of speech recognition using Mel frequency cepstrum and dynamic time warping is discussed.

The concept of augmentative and alternative communication is presented in [7]. AAC is done using communication boards and devices for electronic communication. AAC devices consist of blocks which communicate wirelessly. It is used for patients with severe motor disabilities. The problem with this system is that it contains multiple modules.

Accuracy of characterisation methods for automatic detection of multiple speech disorders is evaluated in [8]. The following impairments considers in this study are: dysphonia, Parkinson's disease, laryngeal pathologies, and cleft lip and palatte. Here, spectral and

ISBN No.978-1-4673-9545-8

cepstral features are used to model the voice spectrum. The conclusion of the study is that, each type of speech impairment requires different features to be modelled.

From speech, emotional content is extracted using mel frequency cepstral coefficients in [9]. Speech signals are divided into frames for smooth transition in this process. The frame length and scroll time is important for emotion recognition. Emotions have been classified using support vector machine and k-nearest neighbour algorithm. It was found that usage of support vector machine was more beneficial.

Voice related quality is discussed in [10]. Here, patients with total laryngectomy are subjected to speech therapy. Voice handicap index questionnaires, focus groups, online surveys are used. The analysis of VHI yields the following conclusions: oesophageal and electrolarynx speakers have moderate voice handicap and tracheoesophageal speakers and patients without voice rehabilitation have severe handicap. The analysis proves the necessity to improve current vocal assistive methods.

One of the main concerns in speech processing is noise filtering. The study in [11] deals with speech filtering methods using adaptive centre weighted average (ACWA) filter. It was shown that various noises were removed effectively using this filter.

II. EXISTING SYSTEM

The existing system consists of only the software. It comprised speech recognition techniques and audio visual feedback. The user was asked to utter a word and the pronunciation was checked. The result was displayed and also uttered.

The system falls short in the fact that there is no hardware to make it more interactive. Hence we need to improve this technique by adding a hardware module whose working is governed by the correct pronunciation of the user.

III. PROPOSED SYSTEM

The proposed system consists of a software and hardware modules as shown in Fig.1

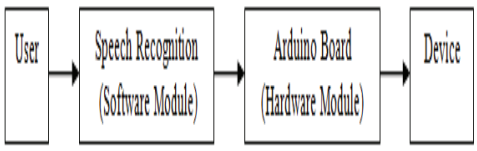

Fig.1.Block diagram of the proposed model

A. Software Module

The software module has been designed using MATLAB software. Its main function is speech recognition. Speech recognition has been done using Linear Predictive Coding and using Euclidean distance the error is calculated.

B. Linear predictive coding

This is a tool used in speech processing to represent the spectral envelope of a signal in compressed form. According to the concept of LPC, the present sample is the linear combination if past samples as depicted in the equation given below:

$$s(n) = \sum_{k=1}^{p} a_k \, s(n-k) + e(n) \quad (1)$$

where,

s(n) - present speech sample

ak - pth order linear predictive coefficient

s(n-k) - past speech sample

e(n) - prediction error

The linear predictive coder processes voiced and unvoiced sounds differently and hence gives accurate results. It is also reliable and efficient.

C. Euclidean Distance

Based on the Euclidean distance between the LPC coefficients of the input signal and the reference signal, we conclude if the user pronunciation is correct or not.

Euclidean distance is given by the formula:

$$D = \sqrt{\Sigma (a-b)^2} \quad (2)$$

where, a,b are two matrices with vectors as columns.

D. Hardware module

The hardware module consists of an Arduino development board which consists of Atmega 328 microcontroller and associated ports and peripherals. The LED (Light emitting diode) is assumed as the device which is operated based on the user's pronunciation. Arduino is an open-source platform which us user friendly and very efficient. It is easy to use for beginners and also efficient for advanced applications. It can be used for a wide range of applications with the advantage of being low cost.

The arduino integrated development environment contains text editor to compose programs and also houses options to upload the program onto the board. The Arduino Uno is a versatile board which connects to the PC via USB cable.It contains the Atmega 328 microcontroller and various input output ports. The ATmega328 belongs to low power Atmel 8-bit microcontroller family. It has advanced RISC architecture with 131 powerful instructions. The hardware setup is shown in Fig.2.

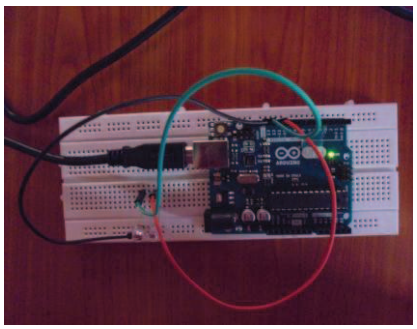

Fig.2. Hardware setup

IV. SIMULATION RESULTS

The simulation was carried out using the MATLAB software. In the system, the word to be pronounced is spoken out and the corresponding picture is displayed as shown in Fig.3.

Fig.3. Word to be pronounced by user

If the word has been pronounced correctly, the feedback is given in audio form and displayed as given in Fig.4. The green LED glows showing that the pronunciation is correct as shown in Fig.5.The arduino is programmed to give logic 1 to green LED if pronunciation is correct and give logic 1 to red LED if the pronunciation is wrong. Sending logic 1 to the corresponding pin depends on the result of the speech recognition module.

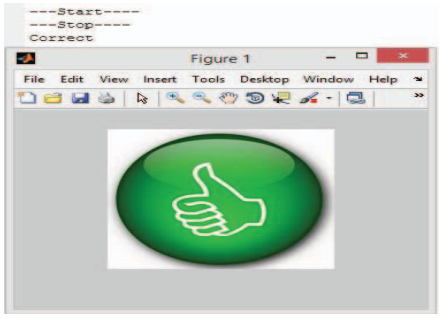

Fig.4. Indicates correct pronunciation

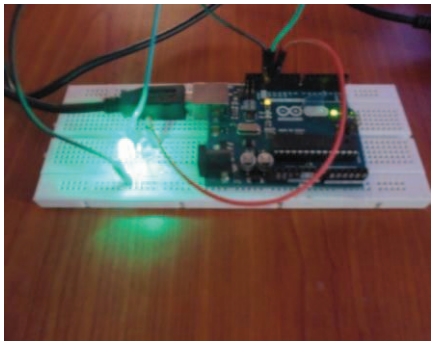

Fig.5.Green LED glows showing correct pronunciation

If the word has been pronounced incorrectly, again the feedback is given in audio form and displayed as given in Fig.6. The red LED glows showing that the pronunciation is incorrect as shown in Fig.7.

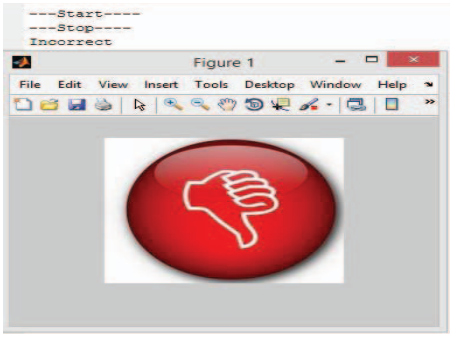

Fig.6. Indicates incorrect pronunciation

V. CONCLUSION

An interactive speech therapy session has been developed where devices are manipulated based on the user's pronunciation is correct. In the developed prototype, green LED was enabled when the pronunciation was correct and red LED was enabled when the pronunciation was wrong.The benefits of the system are: user friendly, feedback is interactive, employs simple concept and is accurate. As an extension we can design the system to support other devices instead of an LED. That devices can be activated only if the user pronunciation is correct.

REFERENCES

[1] Rohit Ranchal, Teresa Taber-Doughty, Yiren Guo, Keith Bain, Heather Martin, Paul Robinson J., et al., "Using Speech Recognition for Real-Time Captioning and Lecture Transcription in the Classroom", IEEE Transactions On Learning Technologies, Vol. 6, No. 4, October-December 2013, pp:299-311.

[2] Darryl Stewart, Rowan Seymour, Adrian Pass and Ji Ming,"Robust Audio-Visual Speech Recognition Under Noisy Audio-Video Conditions",IEEE Transactions On Cybernetics, Vol. 44, No. 2, February 2014,pp: 175 – 184.

[3] Kunxia Wang, Ning An,Bing Nan Li ,Yanyong Zhang and Lian Li, " Speech Emotion Recognition Using Fourier Parameters", IEEE Transactions On Affective Computing, Vol. 6, No. 1, January-March 2015,pp:69-75.

[4] Jesper Jensen and Zheng-Hua Tan," Minimum Mean-Square Error Estimation of Mel-Frequency Cepstral Features–A Theoretically Consistent Approach", IEEE/ACM Transactions On Audio, Speech, And Language Processing, Vol. 23, No. 1, January 2015,pp: 186 - 197.

[5] Zixing Zhang, Eduardo Coutinho, Jun Deng and Bjorn Schuller," Cooperative Learning and its Application to Emotion Recognition from Speech" , IEEE/ACM Transactions On Audio, Speech, And Language Processing, Vol. 23, No. 1, January 2015, pp:115-126.

[6] Aykut Çubukçu, Melih Kuncan, Kaplan Kaplan and Metin Ertunc H.," Development of a Voice-Controlled Home Automation Using Zigbee Module",2015 23nd Signal Processing and Communications Applications Conference (SIU), May 2015,pp. 1801- 1804.

[7] Gemma Hornero, David Conde, Marcos Quílez, Sergio Domingo, Maria Pena Rofriguez, Borja Romero, " A Wireless Augmentative and Alternative Communication System for People With Speech Disabilities", Vol: 3, IEEE Access ,2015, pp: 1288 – 1297.

[8] Juan Rafael Orozco-Arroyave, Elkyn Alexander Belalcazar-Bola˜nos, Juli´an David Arias-Londo˜no, Jesus Francisco Vargas-Bonilla, Sabine Skodda, Jan Rusz, "Characterization Methods for the Detection of Multiple Voice Disorders: Neurological, Functional, and Laryngeal

Diseases", IEEE Journal Of Biomedical And Health Informatics, Vol. 19, No. 6, November 2015, pp:1820-1828.

[9] Onur Erdem Korkmaz and Ayten Atasoy," Emotion Recognition from Speech Signal Using Mel-Frequency Cepstral Coefficients", 9th International Conference on Electrical and Electronics Engineering (ELECO), Nov 2015,pp:1254-1257.

[10] Cristina Tiple,Silviu Matu, Florina Veronica Dinescu, Rodica Muresan, Radu Soflau, Tudor Drugan, et.al," Voice-related quality of life results in laryngectomies with today's speech and expectations from the next generation of vocal assistive technologies", 5th IEEE International Conference on E-Health and Bioengineering, November 2015,pp:1-4.

[11] Kais Khaldi , Abdel-Ouahab Boudraa and Monia turki,"voiced/unvoiced speech classification-based adaptive filtering of decomposed empirical modes for speech enhancement", iet signal processing, vol.10, issue.1, January 2016, pp:69-80.