Speech For The Disabled

Venkatraman.S and T.V. Padmavathi, *Member, IAENG*

Abstract— **Pen-based interface systems which are used for on-line recognition technology have been widely developed. But almost all systems aren't taking into consideration for speech disabled persons. Engineers make things not only simple but also feasible. One of the important problems that the community faces is that people with disabilities are finding it hard to cope up the fast growing technology. I propose a device system which can be used to provide speech for the disabled persons to help people who are dumb or having difficulties to speak. The words that the person wants to say are entered as text. This text is converted into speech by the speech processing unit. The speech processing unit uses the concept of phonetics, and processes each phoneme. Phonetics is focused on sounds; it isn't necessarily concerned with how a particular letter sounds. Instead, what's important is how certain combinations of letters affect the sound. Hence by using phonetics it is possible to produce speech that is as good as actual speech. The frequently spoken words are stored in memory and can be easily retrieved by using hot keys. The system will be made portable in addition. The entire system will be housed inside a cell phone casing with charging facilities.**

Index Terms—**Phoneme- The smallest phonetic unit capable of conveying a distinct meaning, Phonetics- The system of sounds of a particular language,**

I. INTRODUCTION

Various kinds of speech recognition systems are available today but there are hardly a few which help the 'deaf-and dumb'. In the recent years, due to rapid industrialization there has been a rapid increase in the number of speech-disabled victims due to oral diseases, accidents, etc...

Whatever the cause may be, they are rendered helpless and unable to communicate with the outside world! Hence it's high time that we not only focus on our development, but also theirs! We require speech recognition software. For this, the required input for the speech recognition software can be either a written input in the form of text or voice input (which uses similarities in sounds).

Manuscript received October 21, 2008. This work is supported and sponsored by SSN college of engineering, India.

S.Venkatraman is an under-graduate student of SSN college of engineering (Phone number: 091-9843896257; e-mail: svenkatraman3@yahoo.com).

T.V. Padmavathi, Assistant professor in SSN college of Engineering, (e-mail : padmavathytv@ssn.edu.in)

The existing approaches to tackle this problem would be to:

A. Use a speech recorder IC that would record a specific set of words and they would be used to generate the required speech.

B. Use of Dsp processors and dsp algorithms to the reproduce sound.

Fig1. shows the increase in speech disabled persons as a percentage of the totally disabled persons. The picture is clear indicating the steady increase in the number of affected persons.

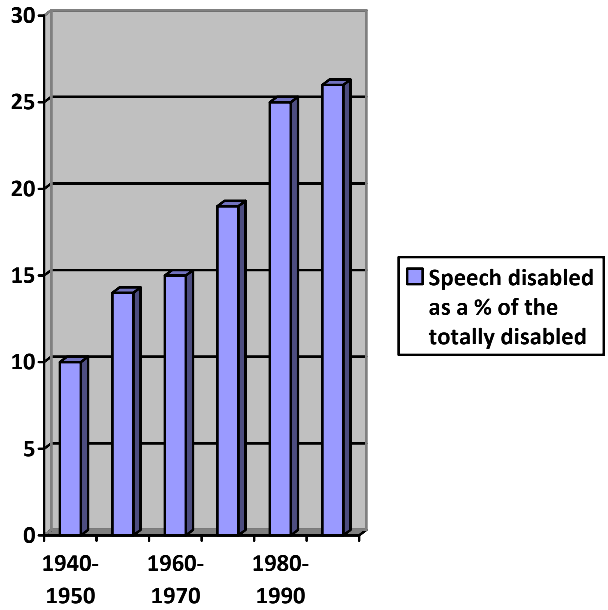


FIG 1. Graph indicating the increase in speech disabled people.

II. EXISTING TECHNIQUES AVAILABLE:

USE OF DIGITAL SIGNAL PROCESSING:

There are three main areas in Speech Processing- speech synthesis, Speech recognition, Speech coding. In speech synthesis, a machine is developed, which can accept a piece of English text and convert it to natural sounding speech.

There are several methods used for the speech analysis:

A. Short time Fourier analysis-

Here the Fourier transform is given as

$$Xn(e^{jwk}) = \sum_{-\infty}^{\infty} x(m)w(n-m)e^{-jwm}$$

Reconstructed signal is given as,

$$y(n) = \sum_{0}^{N-1} Xn(e^{jwk})e^{jwk}$$

ISBN: 978-988-17012-2-0

B. *Cepstral Analysis-*

Here let x(n) be a voiced speech signal, e(n) is the excitation function and h(n) is the vocal tract impulse response . The excitation function and the vocal tract impulse response are convoluted to produce speech signal,

x(n)=e(n)*h(n);

De-convolution is carried out by a general non-linear filtering method referred to as homographic filtering. In this method, the convolution operation is converted into addition giving the output known as complex cepstrum.

C. *Linear prediction analysis.*

- Powerful speech processing technique
- Here, sample values of speech can be approximated as a linear combination of past p samples
- An output of a program written using concepts of DSP is shown. It uses the algorithm of linear prediction analysis,

The object of linear prediction is to form a model of a linear time-invariant digital system through observation of input and output sequences. That is, to estimate a set of coefficients which will describe the behaviour of an LTI system when its design is not available to us and we cannot choose what input to present.

Fig2. is the output of the speech processing technique that was obtained through MATLAB.

III. Medical Treatments Available:

For the speech disabled people, it is possible for them to get their voice by surgery- Cochlear Implant surgery. However this surgery can be performed only to speech disabled children below the age of 6. Considering the risk of surgery, it also costs around 5-6 lakhs. So it's practically not possible for every person to get a surgery. So, we engineers have to innovate to help them reach the same level as us in our society..

IV. Proposed System for Speech Disabled Persons:

My proposal will help the people whom are unable to speak, or having difficulties in speaking. Whatever the person wants to say is written, or entered as text. This entered text is processed using a Speech processing unit. This speech processing unit uses phonetics. Hence by using phonetics it is possible to produce speech that is as good as actual speech. Phonetic alphabet may have 17 vowels and 24 consonants.

Phonetics is not dependant on how he letter is, and depends on how the letter would sound. Further, the frequently spoken words can be stored in memory (such as 'and', `also', `up to', etc...) and can be easily retrieved by using hotkeys.

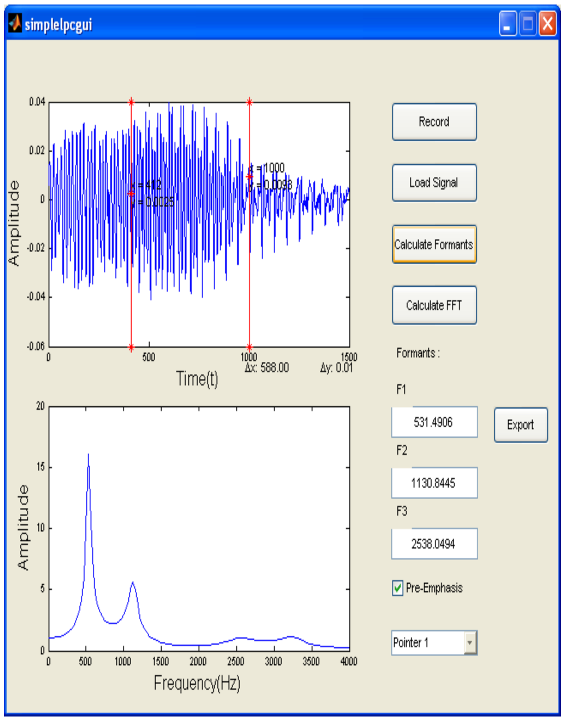


FIG 2.OUTPUT OF SPEECH SYNTHESIS USING MATLAB

V. Feature

A microcontroller shall be used (EMIC IC, Speak-jet, etc...) for this application, and of course, some means to get the input (using a keyboard or a touch screen). In addition to the cell phone peripherals, LCD and keyboard, the system will be made portable. The entire system will be housed inside a cell phone casing with charging facilities, thus enabling it to be almost as much as our day-to-day cell phones. The new system is composed of controller, speech IC, audio amplifier and the speakers. The concept is a phonetic based approach that provides good sound reproducing capabilities. This approach is simple and does not require high end processors. This concept can be extended to other regional language by creating phonetic alphabets for that language. Fig3(a) shows the AVR microcontroller, and fig3(b) and fig3(c) shows the speech chips used.

FIG 3 (a). AVR microcontroller

FIG 3 (b). Speech IC

FIG 3(c). Speech IC

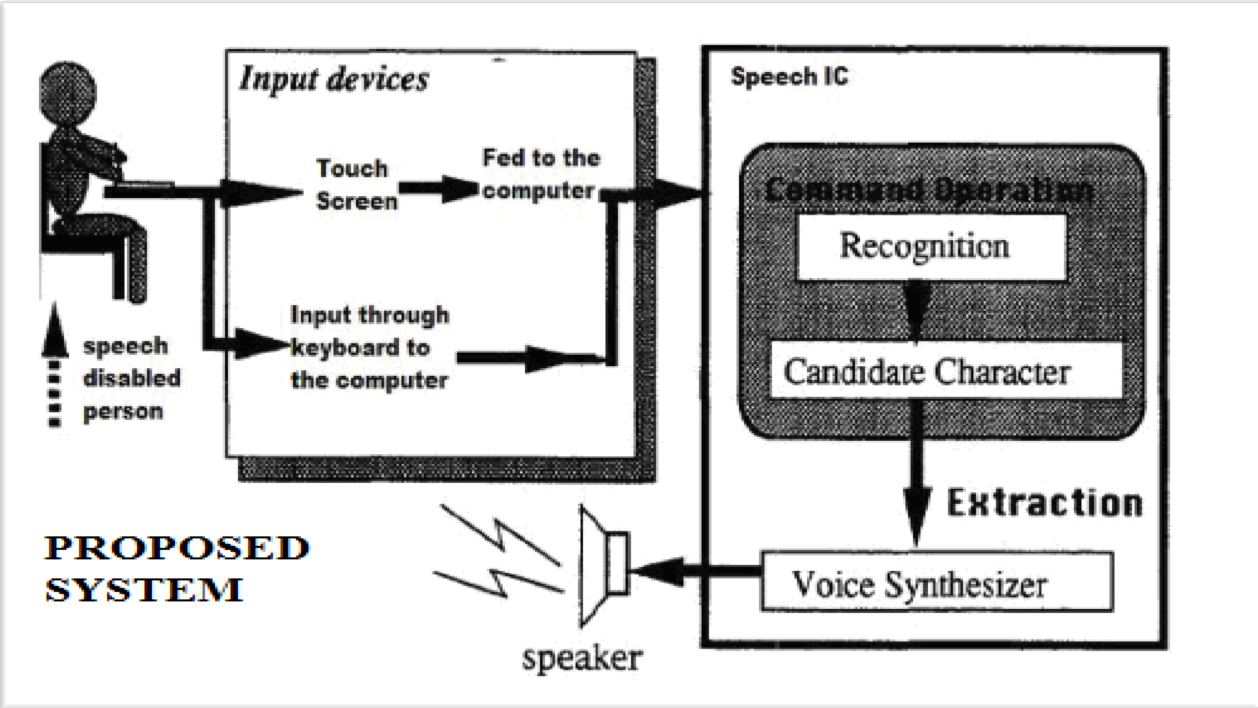


FIG 4. Working of the whole system

VI. Working:

The working of this model is quite simple.

- We use input, which is basically a written text, typed through keyboard, or we may even use a touch-screen to get the input data.
- It is then fed to the microcontroller through one of its ports and it is processed to produce the required speech.
- It is further given to voice synthesizer and it is given to an amplifier with required amplification

The output speech is obtained through a speaker and it is almost as good as normal speech.. This is explained in Fig4.

VII. Algorithm:

The algorithm that is used is basically the Divide and Conquer algorithm. Here, the entered text is segmented and processed by the speech IC, and all the segments are finally compounded together to reveal the typed message. Each symbol is processed at a segment and each phoneme is compunded with the other to form a complete signal , which is then given to a speaker.

Here the segmentation is done at the zero crossing, as shown in the fig 5(a).

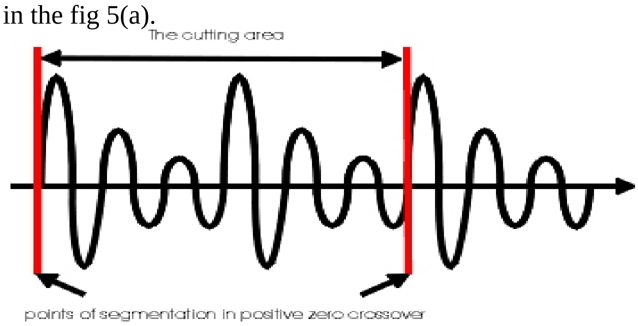


FIG 5 (a). Segmantation at zero crossing

If it is vowel, segmentation is done before the pinch mark at the zero crossing, as shown in fig5(b).

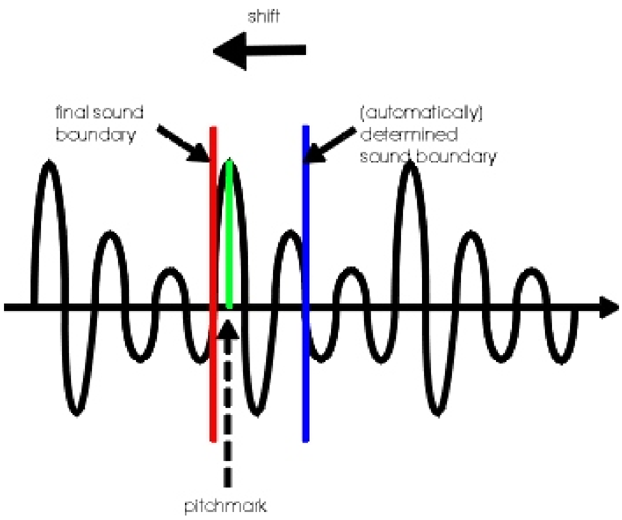


FIG 5(b). Segmentation for a vowel

After segmentation, these waves for each phoneme, are compunded together properly. Improper segmentation results in noise and discontinuity as shown in fig 5(c).

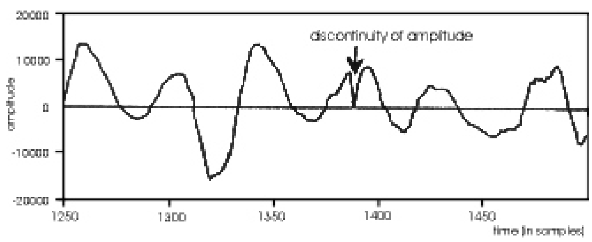
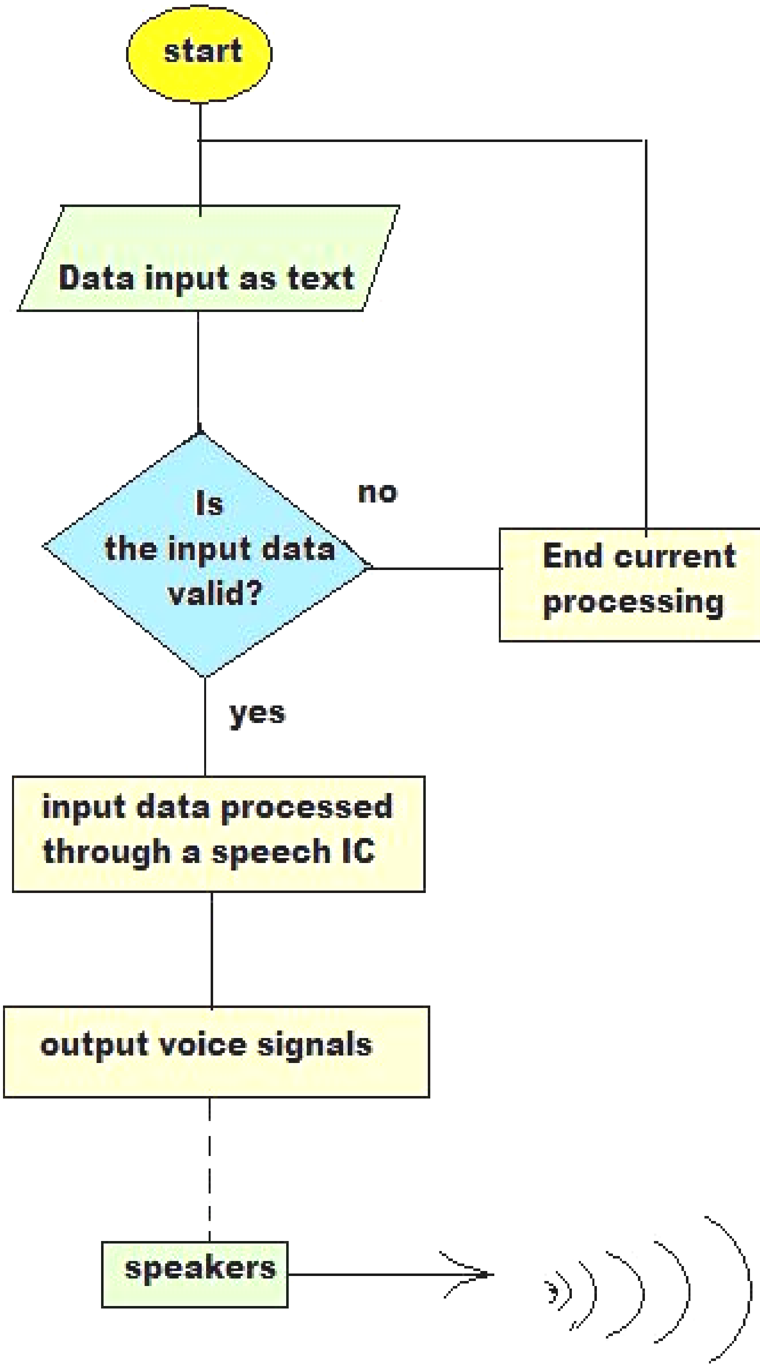


FIG 5(c). Noise

WORKING FLOWCHART:



Consider the word “protocol” - /pro/ /to/ /col/
This is further taken as, /p/ /r/ /o/ and so on. Now /p/ is segmented and it forms a waveform. Similarly, others, such as /r/ /o/ /t/,.. and all these segments are compounded to form a singal wave which is the message signal and it can be obtained at the output. The basic sequence of the segmentation operation is shown in fig 6.

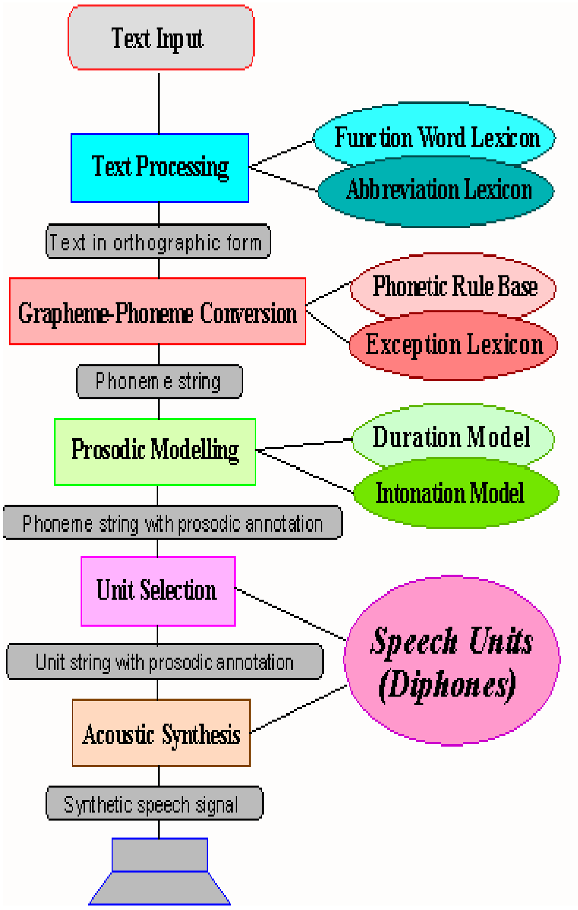


FIG 6. Flow chart

VIII. PROGRAMMING:

The programming is done in WINAVR compiler, an open source compiler for AVR controllers. The program would be written in C language for efficiency and portability. The challenging aspect is to make the system operate from a rechargeable battery and to optimize the system for power. This is most required as the system would be of not much use if the battery drains within a few hours.

IX. LIKELY PROBLEMS TO BE ENCOUNTERED:

The overall sound quality and the battery's power consumption. That is the sound quality may become distorted if the power is at a low value. But a very high powered rechargeable battery would make it less commodious and also more expensive and hence there is a need for optimum valued batteries too! Further problems that would be encountered would be through the PCB layout. The PCB layout needs to accommodate all the components and to fit it inside a cell phone. Bigger the PCB, bigger the cell phone casing has to be! Size is inversely related to comfort. For all practical reasons, it would be easier to select a reasonably big casing. Heat developed could also pose a threat!! Fig7. shows the image of the overall set.

X. BUDGET ESTIMATES

The Table 1 given below gives an idea about the total cost that would incur on the implementation of this proposal.
From this it is concluded that the total effective cost is pretty low and is affordable for the common people.

(Everything in rupees)

Table 1.

S.No.	Component	Quantity	Rate	Total
1.	AVR controller	1	250	250
2.	LCD, keypad, speaker, battery	1 each	1200	1200
3.	EEPROM	1	50	50
4.	Speech IC	1	1200	1200
5.	Pcb and discrete	1	500	500
6.	Touch screen	1	500	500

Total cost comes to 3800 (Indian rupees).

FIGURE 7 . The speech module

XI. ADVANTAGES:

- Highly portable
- Compact
- Small sized
- Total cost of the project is very low
- Good sound reproduction
- Can be recharged once total power has drained..

XII. CONCLUSION:

So far we have always seen technology rise up and it never stops to amaze us. The engineers help in building technology for the betterment of people. The above proposal provides a low cost product that would help the dumb to live in the world like others. They will no longer be restricted by that their inability to express in words as this project provides a cheap, portable and efficient solution.

XIII. REFERENCES:

[1]. B.S. Atal and S. L. Hanauer (1971), Speech analysis and synthesis by linear prediction of the peech wave. Journal of the Acoustical Society of America 50, 637-655

[2]. D.H. Klatt (1987), Review of text-to-speech conversion for English. Journal of the Acoustical Society of America 82, 737-793

[3]. Y. Sagisaka and H. Sato (1986) Composite phoneme units for the speech synthesis of Japanese. Speech Communication 5, 217-223

[4]. Schiel, F. (1999): Automatic Phonetic Transcription of Non-Prompted Speech Proc. of the ICPhS 1999. San Francisco, August 1999. pp. 607-610

[5]. 4. J.B. Carrol et al.,Word Frequency Book, American Heritage, New York, 1971

[6]. 5. J.S. Logen, B. G. Greene, and D.B. Pisoni, "Segmental Intelligibility of Synthetic Speech Produced by Rule," J. Acoustical Society of America, Vol. 86, No. 2, Aug. 1989, pp. 566-581.