Abstract—Pen-based interface systems, which are used for character recognition technology, have been widely developed and used for creating a much more comfortable user interface experience than using a keyboard. 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. It is highly essential that a cost effective product is made available to the society to help speech-disabled people. In this project, we propose a device system, which can be used to provide speech for the disabled persons to help people with difficulties in speaking. The words that the person wishes to convey 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 by housing the entire system 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 is hardly a few which help the ‘speech-disabled’. 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, and so many other causes. 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).

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.

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

\[ X_n(e^{j\omega}) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)e^{-j\omega m} \]

Reconstructed signal is given as,

\[ y(n) = \sum_{k=0}^{N-1} X_k(e^{j\omega})e^{j\omega} \]

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.

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 dependent 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.

V. FEATURE

AVR/8051 microcontroller is being used for this application along with a Speech-IC (EMIC IC, Speak-jet, etc.) for this application, and keyboard or touch screen is used as means of input for this device. 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. Also the parts have to be chosen very carefully so that the components are optimized in their power requirements. The concept is a phonetic based approach that provides good sound reproducing capabilities. Also, in our design, we used an Ebox, housing a windows OS to simulate the working of this design and the Ebox also turns out to be really cheap. 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.

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..
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. We had worked on the module both through manual soldering with EMIC IC and with an Ebox module for Speech Synthesis. The Ebox excelled far better in terms of quality, but the size of the module had increased a little.

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 compounded with the other to form a complete signal, which is then given to a speaker.

If it is vowel, segmentation is done before the pinch mark at the zero crossing.

Here the segmentation is done at the zero crossing.
WORKING FLOWCHART:

1. Start
2. Data input as text
3. Is the input data valid? (If no, go to End current processing, if yes, proceed)
4. Input data processed through a speech IC
5. Output voice signals
6. Speakers
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 /k/ /o/ /t/... and all these segments are compounded to form a signal wave which is the message signal and it can be obtained at the output.

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. For the module in Ebox, Visual Basic is used for programming in C++. 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 was an initial problem we faced, but with the use of the Ebox for Speech Synthesis, the quality was much better than the soldered module that we used. 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!

X. BUDGET ESTIMATES

(Everything in rupees)

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Component</th>
<th>Quantity</th>
<th>Rate</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>AVR controller</td>
<td>1</td>
<td>4</td>
<td>250</td>
</tr>
<tr>
<td>2.</td>
<td>LCD, keypad, speaker, battery</td>
<td>1 each</td>
<td>10</td>
<td>1200</td>
</tr>
<tr>
<td>3.</td>
<td>EEPROM</td>
<td>1</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>4.</td>
<td>Speech IC</td>
<td>1</td>
<td>24</td>
<td>1200</td>
</tr>
<tr>
<td>5.</td>
<td>Pcb and discrete</td>
<td>1</td>
<td>10</td>
<td>500</td>
</tr>
<tr>
<td>6.</td>
<td>Touch screen</td>
<td>1</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

Total cost comes to 60$ (USD).
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 module provides a low cost product that would help the speech disabled live in the world like others. They will no longer be restricted by that inability to express their emotions in words as this project provides a cheap, portable and efficient solution.

XIII. REFERENCES:


[5]. http://users.ece.gatech.edu/~chl/ngasr03/ngasr03-jsh.pdf