CHAPTER-V

SUMMARY AND CONCLUSIONS

There are numerous ways of communication. Two people may communicate with each other through speech, gestures or graphical symbols. Man's most natural way of communication is through speech. Though writing seems to be an important means of communication and written words, appear to be more efficient means of transmitting intelligence, the amount of intelligence exchanged by speech is beyond compare.

The study of speech has gained importance in the field of man to machine communications and speech synthesis was the driving force behind initial attempts to process signal digitally. Because of the fast advances in microelectronics technology, it has found numerous applications. More efficient communication system can be developed with better knowledge of speech mechanism. More than any other branch of information technology, work on speech system seems to pull together a wide and diverse range of disciplines. The application potential of speech synthesis techniques has increased further because of availability of speech processors. The areas where speech synthesis techniques have been employed are computers, medicines, automotives, educational aids, music, etc.

The need of speech synthesis is felt in quite a few fields in electronics in order to generate artificial speech. However, the speech processors have limited vocabulary, so they are inefficient to produce some of the words. With the advent of speech analysis/synthesis, it would be possible to make more vocabulary. Hence, speech analysis/synthesis techniques are preferred. Speech processors can also be used for the limited vocabulary such as emergency signals.

In this work, it was proposed to update the vocabulary, with speech processor. SPO-256 speech processor was selected. With the help of allophone addresses it was possible to update the vocabulary. The speech data for the allophone set is contained within the internal 16k ROM of the SPO 256 AL2. This particular application (allophone set) requires only six addresses pins to address all the 59 allophones plus five pauses a total of 64 locations.

There are two modes available for loading an address into the chip. Mode was used for loading an address. Two interfacing pins are available for quick loading of addresses. They are LRQ and SBY. The programmable peripheral 8255 was interfaced with microprocessor and speech processor, i.e., the updating of vocabulary was implemented on microprocessor based system.

The intel 8255 A I/O device was designed for use with intel microprocessor 8085. It was used in mode 'O' with ports A and C as O/P and port B as I/P. The main line program was written at the address 6000 H. The loop up table was written at the address 6500H.

TBA 810, 7 watt audio power amplifier was used. It operates over a wide range of supply voltages with very low harmonic and cross over distribution. It has integral wing tab heat sinks.

In this work, 200 words update vocabulary was implemented for ex. the words 'one', 'a', 'e', 'o', 'u' have phoneme codes as follows:

Sound	Phoneme code
'One'	02, 23, 16, 1A, 0B, 02
'a'	02, 13, 03
'e'	02, 14, 02, 02, 13, 03
'i'	02, 0C, 14, 03
'o'	02, 35, 02
'u'	02, 16, 35, 02
Ampere	08, 07, 07, 10, 02, 09, 13, 0F, 02

Sound	Phoneme code
Graph	06, 02, 22, 27, 18, 28, 03
Home	04, 3G, 17, 10, D3
Jeep	04, 0A, 0C, 0C, 02, 0G, 03

Three steps involved in updating the vocabulary:

1. Initialize the microprocessor HL pair.
2. Initialize the programmable peripheral 8255, i.e., ports.
3. Initialize the counter for particular sound.

Speech Waveform Analysis Method:

In this work, speech waveform analysis has been carried out. Actually, it is possible to carry out the speech waveform analysis by various methods. Such as time domain analysis/synthesis, frequency domain analysis/synthesis, formant analysis/synthesis, curve fitting method, single cycle analysis. It was proposed to implement the speech waveform analysis by single cycle analysis method.

From single cycle analysis method the conclusions are that it leads to evaluation of period and formant frequency as well as the power and the energy content in a waveform. The spikes superimposed on the formant frequency and their variations in time, form very important information while synthesizing.

In this work, amplitude Vs time graphs of various sounds such as 'one', 'e', 'o', and 'u' were plotted. From this graphs, we came to the conclusion that there are no spikes superimposed on the waveforms of vowels e, o, u, while there are spikes superimposed on the formant frequencies of the waveform of 'one' and 'on'.

Some of the notable features are high lighted below:

1. The vowels 'e', 'o', 'u' have no spikes superimposed on the waveforms.

2. The spikes superimposed on the formant frequency

Hence, it can be written as follows:

$$X(t) = a_0 + a_1 \cos 2\pi t + a_2 \cos 4\pi t + \dots\dots\dots a_n \cos 2\pi nt + b_1 \sin 2\pi t + b_2 \sin 4\pi T + \dots\dots\dots b_n \sin 2\pi nt$$

The steps involved are as follows:

1. First compute a_0

$$a_0 = AL/2$$

2. Compute a_n

$$a_n = \frac{At}{2\pi^2 n^2}\left[\frac{2\pi nt}{T} \sin \frac{2\pi nt}{T} + \cos \frac{2\pi nt}{T}\right]_0^T$$

3. Compute b_n

$$b_n = \frac{At}{2\pi^2 n^2}\left[\sin \frac{2\pi nt}{T} - \frac{2\pi nt}{T} \cos \frac{2\pi nt}{T}\right]_0^T$$

4. Sample the signal

5. $$fs = a_0 + a_1, \cos 2\pi t + a_2 \cos 4\pi t \dots\dots\dots + a_n \cos 2\pi nt + b_1 \sin 2\pi t + b_2 \sin 4\pi t + + b_n \sin 2\pi nt$$

The spike analysis leads to estimate maximum and minimum amplitude. As well as graphical representation shows that there are 2 types of spike variation. The results are highlighted below:

1. Graphical representation of spike leads to the conclusion that the spikes show variations. Some of the spikes vary with sinusoidal decay.
2. Frequencies of the waveforms of 'one' and 'on' are present at the initial 50 % duration of that sound.
3. The formant frequency of 'e' initially varies between 250 - 200 Hz and then varies between 200 - 150 Hz.
4. The formant frequency of 'o' varies very little between 200 - 175 Hz.
5. The formant frequency of 'u' drops from 250 to 200 Hz.

Thus, it can be concluded that as you go from e - o - u the variation of formant tends to decrease.

6. The variation in amplitude with time is almost smooth for'e'.
7. The variation in amplitude for 'o' shows sudden increase of amplitude at 0.11 sec.
8. At the same time, the amplitude variation of 'u' is very random.

Spike Analysis:

The total spread of 'one' and 'on' is 330 ms and 355 ms, respectively. The formant frequency varies between 200 Hz -125 Hz and 200 Hz - 143 Hz, respectively. In 'one' there are 38 spikes while in 'on' there are 35 spikes.

For spike analysis frequency domain analysis was used. Frequency analysis for a continuous time period signal was computed as follows:

∴ Fourier series represent in the form

$$X(t) = C_0 + 2 \sum_{k=1}^{\infty} | C_k | \cos (2\pi k f_0 t + \theta k)$$

Where,

C_0 is real valued and $X(t)$ is real

1. Some of the spikes decrease rapidly.

2. It is noted that the spikes having duration '1' m sec show rapid change in the amplitude and then the amplitude is constant, i.e. they have a mathematical equation e^{-at}.

3. The spikes having duration other than '1' m sec show sinusoidal decay in the waveform amplitude, i.e. they have a mathematicaL equation of the form $e^{-at} \sin \theta$.

In the work, it was also proposed to evolve a model in order to synthesis the sounds.

The calculated formant frequency of various sounds have no. of other frequency harmonies and other amplitude variations superimposed on it. The amplitude variation of formant frequency can be estimated as follows:

1. Fit the time amplitude graph in 2nd order or 1st order damped waveform with the help of step input.

2. From the waveform its damping ratio and further R, L, C can be calculated for the system.

3. As well as 1st order system can be generated with R, C values and can get the required waveform.

4. The harmonics present in the waveform of the speech are in the form of voltage spikes. These can be mixed in the waveform with the help of microprocessor and one differentiator FIR system.

By mixing amplitude variation and harmonics distortion we can get the required waveform of the speech in its electrical equivalent signal. By sending it through speaker we can get the required synthesized sound of the required formant frequency.