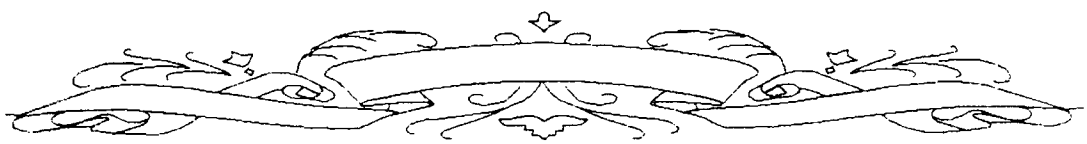

CHAPTER-I

INTRODUCTION TO SPEECH SYNTHESIS

1.1 HISTORY OF SPEECH SYNTHESIS:

Speech synthesis was the driving force behind initial attempts to process signals digitally.

In 'sixties, Computer Simulations were carried out in the medical research, oceanography, radar and problem in speech[1]. The area was scattered, however, and some of the existing techniques were not clearly comprehended. Charles M. Rader and Bernard Gold[2] of Massachusetts Institute of Technology (MIT) were amongst the first to realize the importance of providing unified and in depth coverage of this topic. They wrote a book in 1969 on digital signal processing. The book provides fundamental material useful to design digital hardware and general purpose digital structures to solve signal processing problems. In 'sixties, however, the DSP hardware used discrete components and consequently, because of the high cost of volume, its application could only be justified for very specialized requirements.

Oppenheim and Schafer[3] and Rabnier and Gold[4] worked hard to clarify, expand and organize the fundamentals and tools of digital signal processing. The field of DSP has grown considerably during the past decade with increased utilization in the advanced applications[5]. At present every major semiconductor manufacturer offers general as well as dedicated processors and devices. The DSP processors now available, have many advantages over general purpose microprocessors. In specific disciplines involving a lot of analyses, e.g., seismic processing, imaging, animation etc., DSP techniques have offered vast and complex computational capabilities thereby leading to compact designs and to production of efficient real time signal processing systems. The on-going microelectronics revolution with substantive supporting software, availability of DSP processors and interface chips have added further impetus to this growth.

1.2 APPLICATIONS OF SPEECH SYNTHESIS:

Speech synthesis has vide applications in almost all fields. Some of the areas where speech synthesis is extensively used are listed below:

A. Artificial intelligence in computer;

B. Computer network;

C. Robotics;

D. In medicine field computers can talk to the blind;

E. Verbal warnings and directions when some emergency condition exists;

F. Telecommunications;

G. Appliances;

H. Computer peripherals;

I. Automotive;

J. Personal computers;

K. Toys/Games;

L. Educational aids;

M. Electronic musical instruments.

1.3 ORIENTATION OF THE WORK:

Digital signal processing is a relatively new field which began with speech synthesis in early 'sixties'. The field of speech synthesis has grown fast and has shown progress in the past years due to advances in microelectronics technology. Speech synthesis finds applications in almost all fields due to high precision, stability, wide vocabulary and simple hardware of the system. The speech processors lead to compact design and production of real time systems. The speech processor chip SPO 256 is very popular and powerful processor chip. Now-a-days the fields where speech synthesis is used are computer networks,

artificial intelligence, medicine, telecommunication, educational aids etc. Thus, any scientific or technological problem requiring speech generation can only be tackled by speech synthesis.

The need of speech synthesis is felt in quite a few fields in electronics in order to generate the speech. With the help of speech processor it can be achieved. Speech processor SPO 256 is preferred because it is able to synthesize speech or complex sounds, using its stored program. It has many good features such as natural speech, wide operating voltage, word, phrase and sentence library. It has extra ROM which is expandable to 491 K, simple interface to most microcomputers or microprocessors and it also supports Linear predictive coding synthesis, Formant synthesis and Allophone synthesis. It has 16K ROM which stores both data and instructions.

In this work it was proposed to test speech generated by speech processor. SPO 256 speech processor was selected, further speech waveform analysis work was carried out.

In this work, single cycle analysis of some speech sounds of 'one', 'on', 'e', 'o', 'u' has been carried out. It was leading to evaluate the period, formant frequency, the power and the energy. Ultimately, the spikes superimposed on the formant frequency and their variation in time form very important information while synthesizing. It was also proposed to evolve a model in order to synthesis the sounds.

1.4 THEORY OF SPEECH SYNTHESIS:

The electronic speech synthesizer is a direct analog of the human vocal tract. The vocal system can be broken down into three main regions; lungs, larynx and vocal cavity. The sound is made by vocal cards. Vocal cards are made up of skin layers. The vibrating action of vocal cards generates several

resonant frequencies. Hence, different sounds are made by changing the shape of your vocal cavity, i.e. your throat, tongue, teeth and lips. The pipe model of human vocal tract is shown in Fig. 1.1

The sound wave resonate at odd quarter wavelength frequency in such a pipe. So we get the frequency;

$$\frac{1}{4} \times \frac{C}{L} \quad , \quad \frac{3}{4} \times \frac{C}{L} \quad , \quad \frac{5}{4} \times \frac{C}{L} \qquad \text{and so on}$$

i.e.

500 Hz, 1500 Hz, 2500 Hz;

Where,

C = Speed of sound = 34000 cm/sec and
L = 17 cm.

The resonant frequencies are called as formats. There are several harmonic frequencies also, which forms format bands.

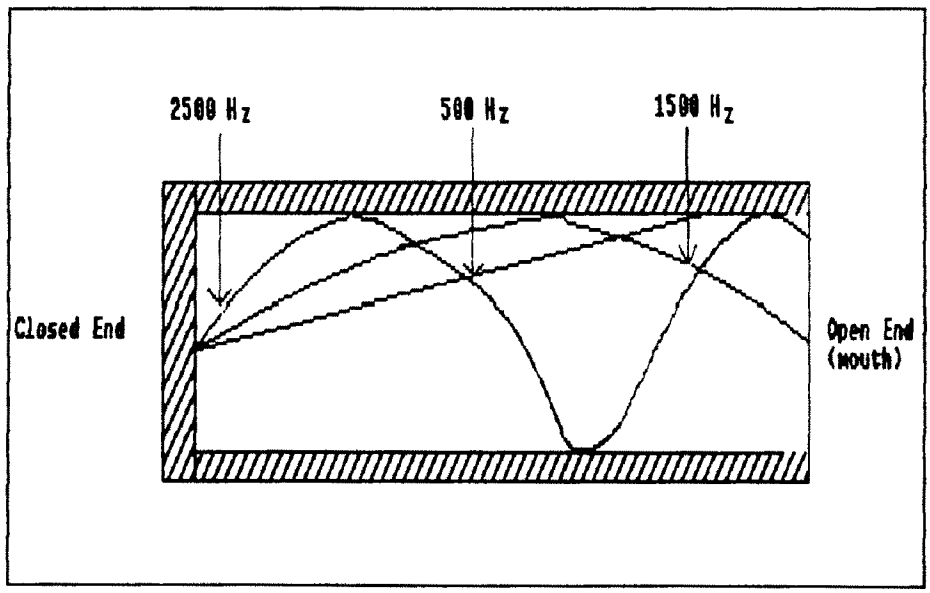

Fig. 1.1 Pipe model of human vocal tract

Two types of sounds are produced; one is voiced sound and other one is unvoiced sound. Voiced sounds are produced due to vibrating action of vocal cards. These are nothing but vowels, i.e. a, e, i, o, u. Unvoiced sounds are produced due to air turbulence, these are consonants. There are different types of voiced and unvoiced sounds depending upon the action of vocal cavity e.g. pure voiced sounds, nasal voiced sounds.

The sound of speech are called as phonemes. These are nothing but various types of voiced and unvoiced sounds. They are fundamental sounds of English language. By stringing phonemes together speech is created. Such type of 40 phonemes are there. The given phoneme may be pronounced by many different ways, so sound variations occurs and these sound variations are called as allophones. Each phoneme has several allophones associated with it. Approximately 128 sound variations are produced due to 40 phonemes.

REFERENCES

1. Ludman, L.C. "Fundamentals of Digital Signal Processing", Harper & Row, Publishers, New York, 1986.

2. Gold, B., and C. Radar. "Digital Processing of Signals, McGraw-Hill, New York, 1986.

3. Oppenhiem, A.V. and R.W. Schafer. "Digital Signal Processing", Prentice -Hall, Englewood Cliffs, NJ, 1975.

4. Rabiner, L.R. and B. Gold. "Theory of Application of Digital Signal Processing", Prentice-Hall, Englewood Cliff, NJ 1975.

5. Bateman, A. and N. Yates. "Digital Signal Processing", Pitman Computer System Series, 1988.