CHAPTER 2

BIOLOGICAL BASIS OF SPEECH PRODUCTION AND ITS PARAMETERS

CHAPTER 2

BIOLOGICAL BASIS OF SPEECH PRODUCTION AND ITS PARAMETERS

2.1 PRODUCTIVE MECHANISM :

Sounds are vibrations with characteristics of frequency, intensity, and duration which produce certain sensations of audibility when impinging upon the ear.

2.1.1 AIR STREAM MECHANISM :

Within the vocal apparatus there are five cavities (Fig.2.1). These embrace the oral, nasal, pharyngeal, pulmonic and esophageal (including the stomach) cavities. A back velar (or uvular) closure is considered to be within, but terminating, the oral cavity; analogously, a velic closure lies within the nasal cavity; glottal and esophageal closures are contained within the pharyngeal cavity [1].

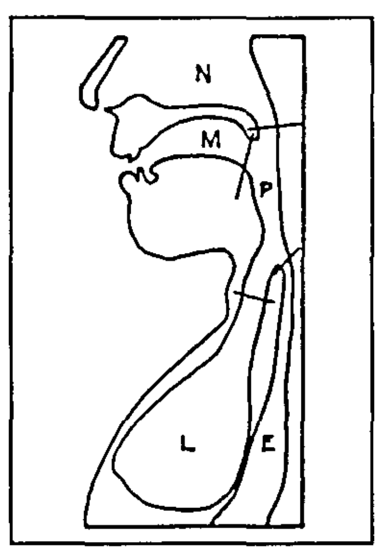

Fig. 2.1 Cavities

E– esophagus, L– lungs, M– mouth, N– nose, P– pharynx.

Any cavity or part of cavity which is completely shut off by itself from others by some closure, or any group of cavities or parts of cavities which are united by connecting passageways of air, forms an air chamber. Thus during the stop [p] there are three air chambers: the nasal, the oral-pharyngeal-pulmonic, the esophageal. There are three major air stream mechanisms—(a) Pulmonic, (b) Pharyngeal and (c) Oral air-stream mechanism [2,3].

(a) Pulmonic air stream mechanism—The lungs acting as an initiator may press lung air outward through the pharynx and mouth or nose (or some connected combination). This may be done in a rapid burst, as after a cough, where the pressure is obvious, or very slowly, as for a sustained frictionless continuant whose pressure is slight but which may be continued for about half a minute. During inspiration the enlarging of this chamber may like-wise be rapid, or sustained, the time being limited by the capacity of the lungs to continue expansion. More sound types can be produced with lung air than with any other mechanism.

(b) Pharyngeal air stream mechanism—The larynx with glottis closed may be thurst upward, compressing the air beyond it. If an egress is provided through the mouth or nose pharynx air will rush out. When there is no egress, a pharyngeal capped air stream results; these are glottalized stops (Fig.2.2). Sounds with pharynx air include the same general types as with lungs air, subject to certain changes and limitations [4].

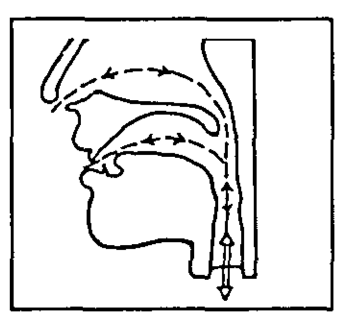

Fig. 2.2 Pharyngeal air mechanism

(c) Oral air stream mechanism—The back part of tongue while touching the velum may be thrust forward, compressing any mouth air caught between it and some stricture farther forward, at the tongue tip or lips or the like; if a place of escape is provided, an egressive air stream is produced (Fig.2.3). The back part of the tongue may be moved farther back in the mouth while maintaining its velar closure and in this manner rarefy the air in the chamber and initiate an ingressive stream to the mouth; the sounds of kissing, or of clucking to horses, are produced in this way.

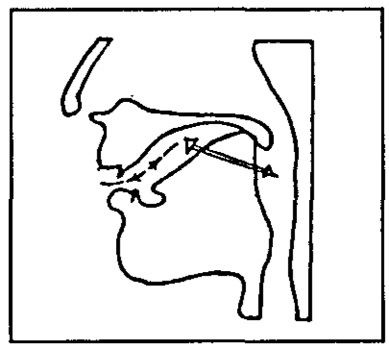

Fig. 2.3 Oral air mechanism

Like the pharyngeal air mechanism, the oral has small expansion-contraction powers. Sounds produced by it rarely lasts more than a fraction of a second, though their spirants can be continued weakly for a second or two.

In a nutshell, air streams may be either egressive or ingressive. Exactly halfway between egressive or ingressive types lie the percussion mechanisms; if same part of the chamber wall of a percussion mechanism begin to move inward or outward, especially the lungs, glottal closure, or velar closure, the mechanism immediately becomes an active air type [5,6].

2.2 THE SENSORY BASIS OF SPEECH :

None of the sounds of human speech is produced by an organ biologically designed for a speech function; all are heard by an organ endubitably designed for the hearing of sounds. Language is founded upon the sense of hearing; there is no organ of speech in the biological sense. We shall begin with a description of the mechanism of breathing and proceed thence to trace the breath stream from its origin to its emergence.

2.2.1 LUNGS :

The biological purpose of respiration is to cleanse the blood of its accumulation of carbondioxide and to provide it with fresh oxygen from the outer air. This chemical transfer takes place in the lungs, which are large spongy bodies composed of many small vessicles or tubes, called alveoli. Each of these alveoli receives its supply of air form a larger tube, called a bronchiole, and the *bronchioles in turn join two* larger systems, coming together in the left and the right bronchus respectively.

The two bronchi, or bronchial tubes, unite at the base of the trachea, which is the sole passageway for the air from the throat to the lungs (Fig.2.4).

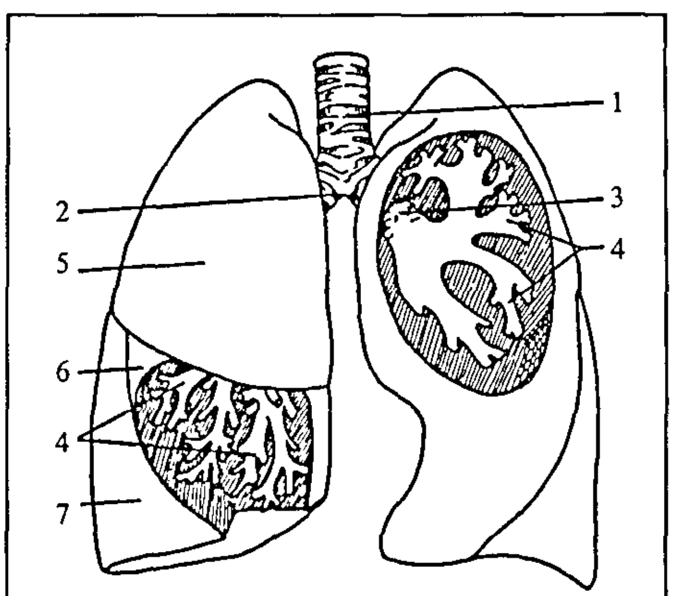

Fig. 2.4 Lungs, Bronchi, Bronchioles

1. Trachea, 2. Right bronchus, 3. Left bronchus, 4. Bronchioles 5. Upper lobe, 6. Middle lobe, 7. Lower lobe.

2.2.2 BREATHING MOVEMENTS :

The lungs are incapable of movement by their own energy although they possess a minimum of elasticity which causes them, if they have been squeezed together, to expand sufficiently to open the alveoli to the access of air. The inflation and deflation of the lungs is accomplished by alternately reducing the pressure upon their outer surfaces until air from the outside is induced to flow in, and increasing this pressure upon their outer surfaces until the air contained in the lungs is force to flow outward. Such pressure are called negative when they are less than that of the outer air and positive when they are greater than this. Nature accomplishes this function by enclosing each lung in an air-tight chamber, the pleura, and then producing the required negative and positive pressures by movements of the thorax. Perforation of the pleura makes the production of the required negative pressure impossible.

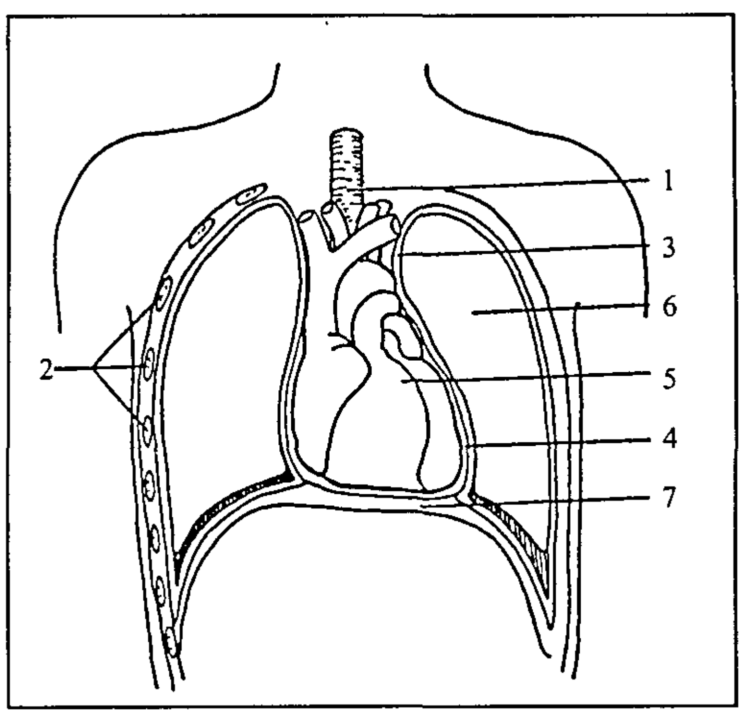

Fig. 2.5 Thorax, Rib Cage, Diaphragm

1. Trachea, 2. Section of ribs, 3. Outline Pleura, 4. Outline of tericardium, 5. Heart, 6. Lungs, 7. Diaphragm

The thoracic cavity (Fig.2.5) in which the lungs lie is bounded above and laterally by rib cage. The lower boundary is a large sheet of muscular and fibrous tissue, shaped like a tall dome with the concave side downward. This sheet is the diaphragm. It is attached in front to the sternum, or breastbone, and to the cartilages of the lower six ribs on either side. In the rear it is attached to the lumbar vertebrae. The diaphragm is the principal muscle of respiration. When its muscular fibers contract, the whole vault of the diaphragm is drawn downward until the pressure of its under surface against the viscera, supported by walls of the abdomen, prevents further downward movement. This point is soon reached. Then with its centre resting on the crowded viscera, the diaphragm, as it contracts further, lifts the front and side structures to which its margins are attached – the lower ribs and the sternum. These movements enlarge the thoracic cavity and thereby reduce the pressure per unit area upon the outer surfaces of the lungs. If the access of outer air is free as far as the trachea, the pressure on the inner surfaces of the lungs will be that of the outer air. Whenever, therefore, the movements of the diaphragm and rib cage so reduce the pressure on the outer surfaces of the lungs that this is lower than the pressure on their inner surfaces, air will flow into the lungs and inspiration will take place [7].

When the pressure in the thoracic chamber is reduced, the air flowing into the lungs will force these structures to expand into the thoracic cavity until the pressure on the outer surfaces of the lungs is equal to that on the inner surfaces. When the muscles of the diaphragm are relaxed, the weight of the raised portions of the thoracic walls and the elastic push of the viscera and abdominal walls force the diaphragm upward. This reduces the size of the thoracic cavity and increases the pressure per unit area on the outer surfaces of the lungs. The result is that the air is then forced out of the lungs until the pressure on the inside surfaces is equal to that on the outside surfaces. If for any reason the movements of the diaphragm are restricted, the required negative pressure in the pleurae can be produced by lifting the walls of the thorax. Indeed these walls are usually lifted to some extent in conjuction with the depression of the diaphragm in all but very quiet breathing. Two types of breathing are usually recognized; the abdominal type, in which the diaphragm is the prime mover; and the thoracic type, in which the movements of the diaphragm are minimized and the movements of the upper rib cage are made more extensive [8, 9].

2.3 THE MECHANISMS OF SPEAKING :

All the muscles, bones, cartilages and organs used in speaking have other functions which are biologically more important than producing the voice. The tongue, for example, even though a vital part of the speaking mechanism, is more important in eating. The vocal folds protect our lungs from irritants in the atmosphere and help to regulate the air flow (Fig. 2.6).

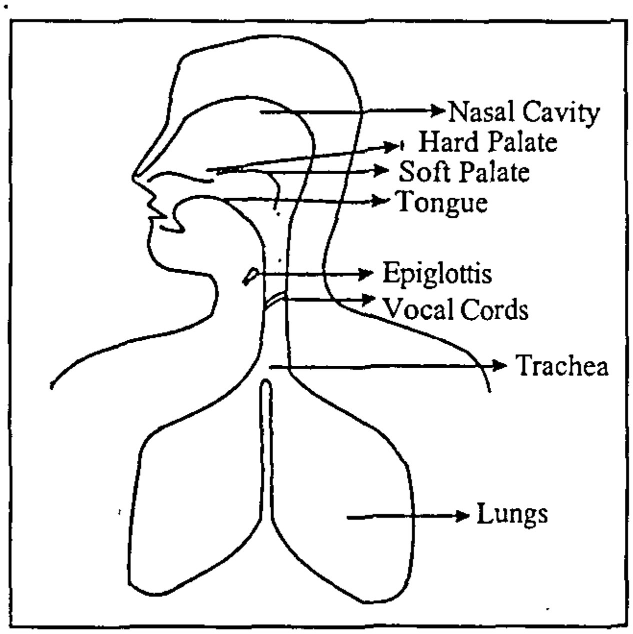

Fig. 2.6 Organs of speech

The very fact that speaking is a secondary function of these organs makes doubly important a program of vocal training, for although we were able at birth to breath, we had to learn to speak. In the process, many of us did not learn to speak well. Let us therefore forget for the present their primary biological functions and consider them together as a single mechanism – the instrument of speech. This instrument may be divided into two major parts : (a) The voice producing mechanism, including the motor, the vibrator, and the resonators, (b) the articulatory mechanism, including the tongue, teeth, lips, jaw and the hard and soft palates [10,11].

2.3.1 THE VOICE PRODUCING MECHANISM :

(i) The Motor—The motor part of the speech mechanism is essentially a pump for compressing air. It consists of (i) the lungs, which contain spaces for the air, (ii) the bronchial tubes, which converge into the windpipe or trachea, out of which the compressed air is released, (iii) the ribs, and other bones, cartilages, and tissues which serve to hold the motor in place and give leverage for the application of power, (iv) the muscles, which alternately expand and contract the area occupied by the lungs, thus first allowing air to enter and then compressing it for expulsion.

The human air pump works in two ways : certain muscles draw the ribs down and in when we exhale, so as to squeeze the lungs after the fashion of a bellows, while others – the strong abdominal muscles, squeeze in below to exert pressure up against the bottom of the lungs after the manner of a piston. This double action is also exerted when we inhale : one set of muscles pulls the ribs up and out to expand the horizontal space, while the diaphragm – a layer of muscles and flat tendon tissue – expands the vertical space by lowering the floor of the chest cavity; this two way expansion creates a suction, so that air rushes into the lungs. Thus, both inhaling and exhaling involve two coordinated actions; moving the ribbed walls of the chest and raising and lowering its floor.

(ii) The Vibrator—The air compressed in the lungs during exhalation is directed through the trachea into the larynx, which contains the main vibrating unit. The larynx is situated at the upper end of the trachea and is attached above and below by muscles which shift it up and down. The larynx itself consists of a group of small cartilages joined so that they can move as if on joints like the bones of the arm. The position of these cartilages can be changed by a number of small muslces which are delicately intertwined within the larynx, stretched between the cartilages, are the vocal folds. The folds are the tendonous inner or facing edges of two muscles. When sound is to be produced, they come together until there is only a tiny slit between them.

The compressed air from the lungs, pushing against and between the vocal folds, causes a vibration which results in sound. The pitch of this sound, its highness or lowness on the scale depends on the muscles which control the tension and length of the folds. The position of the larynx as a whole is adjusted to a proper relation with the air cavities above by the action of the larger outside muscles which hold it in place. The action of these two sets of muscles, particularly the small internal ones, is largely automatic, they can not be controlled individually. But we can operate these laryngeal muscles as a group to control pitch.

(iii)The Resonator—The sound produced in the larynx by the vibration of the vocal folds is thin and weak. It is resonated by a group of air chambers in the head and throat [12]. The principal resonators of the human voice are the upper part (or vestibute) of the larynx, the throat (pharynx), the nasal cavities and the mouth (Fig. 2.7).

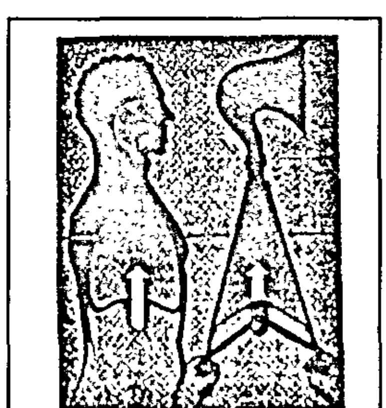

Fig. 2.7 The voice as a wind instrument

These resonators act much as do the resonating parts of a musical instrument; they amplify the sound and they modify its quality, making it rich and mellow or harsh, or whining. Moreover, changes in the size and shape of some of these chambers result in the different tone qualities that constitute the vowel sounds [Fig 2.8].

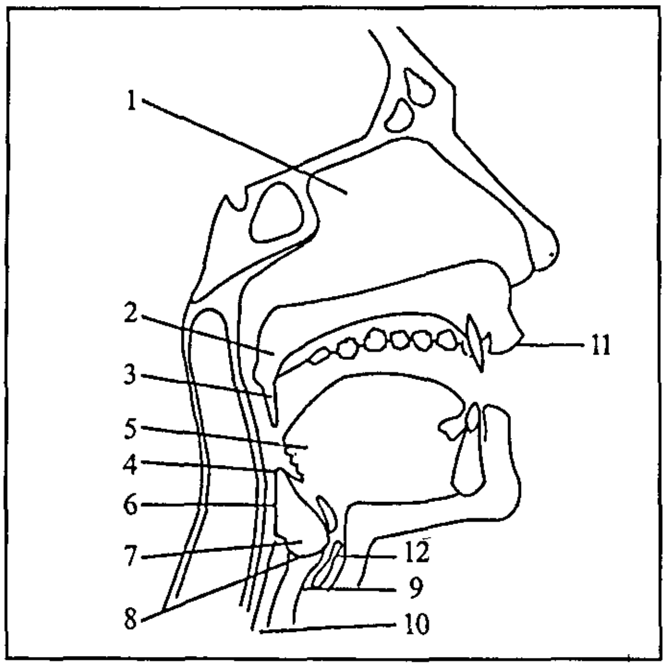

Fig. 2.8 Voice generation involves lungs pharynx mouth and nose.

1. Nasal cavity	4. Epiglottis	7. Larynx	10. Oesophagus
2. Soft palate	5. Tongue	8. Vocal fold	11. Lips
3. Oral pharynx	6. Pharynx	9. Trachea	12. Thyroid cartilage

2.3.2 THE ARTICULATORY MECHANISM :

The tongue, lips, teeth, jaw, and the hard and soft palates act as modifying agents in the production of speech sound. By moving them we modify the size and shape of the mouth, and therefore, the quality of the tone. Another important function of the modifiers is the formation of consonant sounds — the stops, hisses, and other interruptions in the steady flow of vowel sounds that serve to make words out of what would otherwise be mere vocal tones [13,14].

2.4 ACOUSTICAL PARAMETERS OF SPEECH SOUND :

There are three primary areas of acoustical parametric studies – Frequency, Time and Amplitude. The fourth one is Energy. Let us view these areas –

(a) Frequency – Frequency, as it relates to speech, is the number of sound waves per second produced by vibration of vocal folds. Speech sounds are made up of different combinations of frequencies. This is physical counterpart of pitch. The higher the frequency, the higher the scale of pitch.

(b) Amplitude – Amplitude is defined as the extent of displacement of sounding body. The amplitude refers to the extent to which vocal folds are displaced. This is the main physical counterpart of the perceived loudness of a sound, the greater the amplitude, the louder the sound, other things being equal. Intensity is related to amplitude in that it is the measure of strength of energy of a sound that produces the sense of loudness.

(c) Time – Time is defined as period for which a process exists. It refers to the duration of speech sound. The understanding of such temporal (time-related) requirements in the production and identification of speech sounds provides a greater understanding of phonetics of speech sounds.

(d) Energy – The sound produced in speech communication depends on energy with which the exhaled air is propelled from lungs to outside atmosphere [15,16].

The sound created by the interruption of the air-stream by the vibrating vocal folds is known as the voice source. The vibratory frequency of the vocal folds sets a fundamental frequency. A complex tone may be set by combining large number of higher harmonics or over tones. The amplitude of the harmonics decreases almost linearly with frequency – by about 12dB per octave.

A rapidly varying air pressure in the vocal tract is analogous to an oscillation of proportional frequency. To the outside world, this is the sound meaningful or otherwise. The vibratory frequencies set the birth of the sound generated. In a meaningful or purposeful sound generation, the human being has full control over the pitch. In the case of speech the pitch control lasts less than one octave. On the other hand an activity like singing would have a pitch control over one octave. Professional singers extend the pitch control over two or more octaves. A plot of amplitude against frequency of a voice source is termed as a source spectrum.

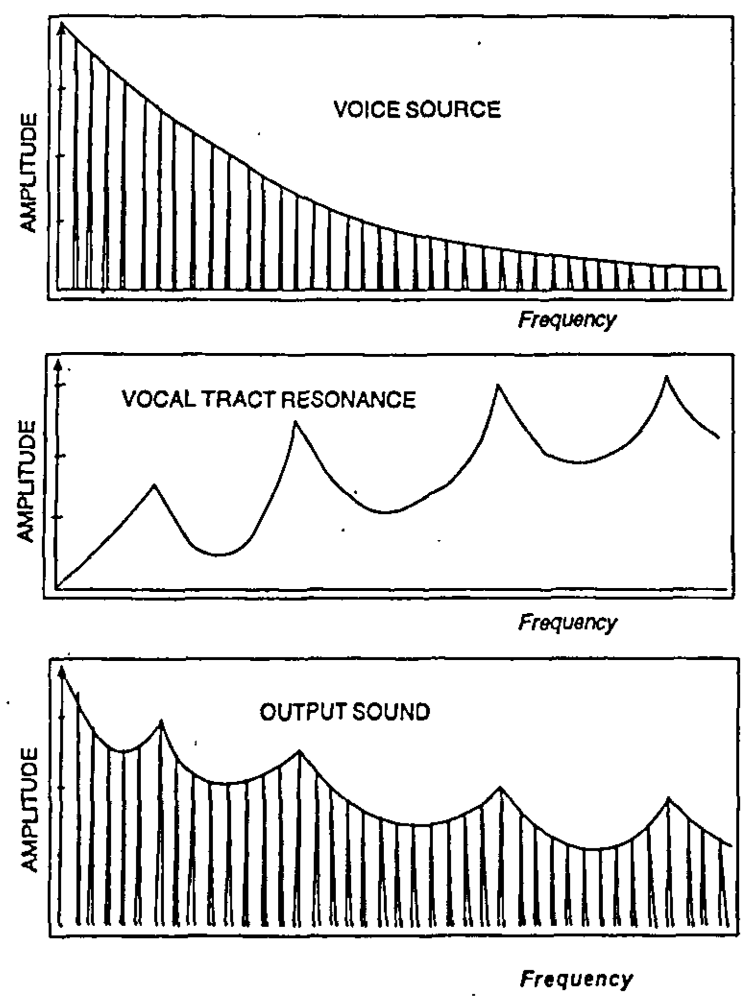

Fig. 2.9 Process of Voice generation.

The vocal tract is a resonator. As in the case of any resonator, its activity is extremely frequency-dependent. The resonant frequency of the acoustic resonator under-goes lesser attenuation than the nonresonant frequencies, transmitted through vocal tract. The sound generated by the vocal tract will have peculiar resonant frequencies rich in relative amplitude or greater relative loudness [17,18].

Consequently the resonant frequencies will have greater relative amplitude than the other frequencies. In other words, the radiated sound of a resonator becomes increasingly weaker if the frequency lies away from the characteristic resonant frequency of the vocal tract (Fig.2.9).

2.4.1 FORMANTS :

In human being, the vocal tract is seldom resonant with a single frequency. Usually, it exhibits four or more major resonances. These resonances are collectively called 'formants'. A graph of the frequency versus energy of a speech waveform exhibits peaks of formants in the energy spectrum in the vicinity of frequencies of the vocal tract resonances. The harmonics of the voice source entered into the vocal tract traverse with different degrees of attenuation. The amplitude at the lip opening is increased if the harmonics entering into the vocal tract lie closer to a formant frequency.

In the absence of formants, the spectrum envelope of the voice spectrum is smooth and devoid of any abrupt peaks. Formants create many peaks at the specified frequencies. The structure of the spectrum peaks, their amplitude and frequency vary for different sounds. It is this structural change in the envelope of the spectrum caused by the formants which makes discernible speech sounds.

Essentially, formants are frequency information specific and have the characteristics of a speech sound. They change their centre frequency and bandwidth during dynamic speech. (Since the frequency domain information changes much more slower than the time domam information, formant identification is an important tool in speech synthesis.)

The fundamental frequencies are decided by the shape of the vocal tract. It is easier to compare the vocal tract as a perfect cylinder closed at the glottis and open at the lips (Fig.2.10).

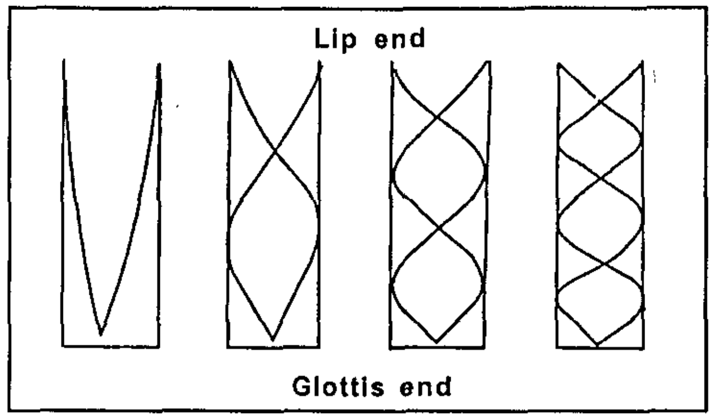

Fig. 2.10 Each formant can be associated with a standing wave.

For men the length of such a cylinder would be about 17 cm and for women about 13.2 cm. The resonant frequencies (f) can be defined by the relation :

$$f = \frac{nv}{4L}$$

where n is a positive odd integer, v the speed of sound in air and L the length of the vocal tract. Assuming the speed of sound in air as approximately 340 m/sec, the first four resonant frequencies for a male would be close to 500, 1500, 2500 and 3500 Hz. There may be little deviation in these frequencies in individuals, depending on the length of the vocal tract [10, 19].

2.4.2 MALE/FEMALE SOUND DIFFERENTIATION :

The most significant factor that is often mentioned regarding male and female voice is the difference in fundmanetal frequency. The fundamental frequency for females lies in the range of 150 to 300 Hz, while it is around 85 to 180 Hz for male voice (Fig.2.11).

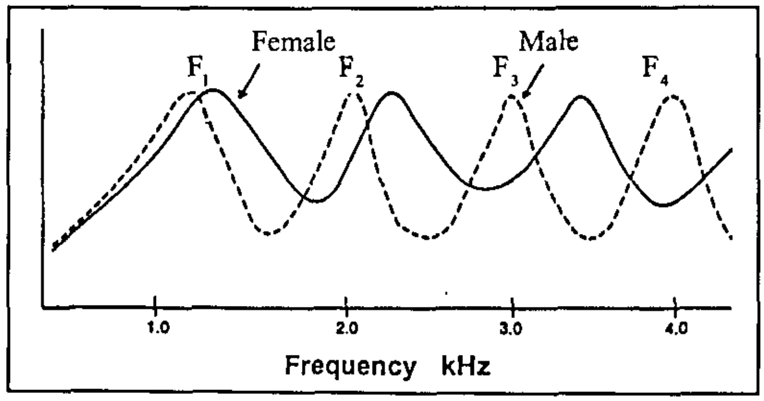

Fig. 2.11 Male voice has more formants

A female voice can be conveniently represented by four resonances while male voice is approximated by five resonances.

A better representation to iso-late male-female voice is to represent formants in a plot of first formant frequency versus second formant frequency, usually called $F_1 - F_2$ plane (Fig.2.12).

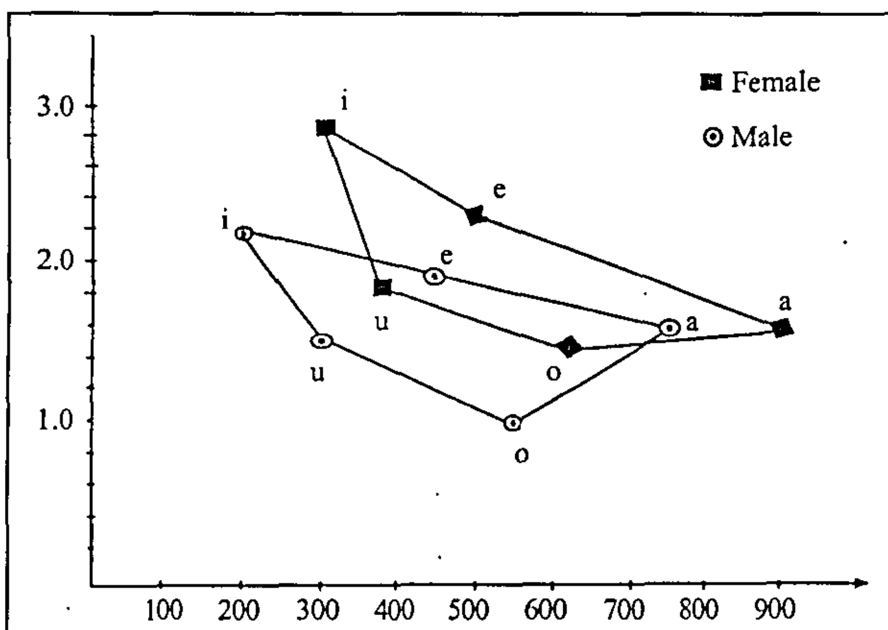

Fig. 2.12 $F_1 - F_2$ Plane vowel distribution

The energy source for the voice generator is obviously bigger for males due to the physical structure in command. This is bound to reflect in the speech energy too. The intensity of voiced speech generated by a female is about 6 dB less than the voice speech of a male in identical conditions.

Among the numerous anatomical differences, the size of the male and female vocal fold is more significant — shorter and symmetrical vocal folds in females as compared to the male version, together with sober intensity of the grottal waveform, tends to make female voice more breathier. Moreover, listeners' perception of female voice is hard printed in terms of intonation, stress, coarticulation and hence melodic cues [10, 19].

2.5 REFERENCES :

1. Pike, K.L. : Phonetics, Ann Arbor : The University of Michigan Press, London, 1955.
2. Zinkin, N.I. : Mechanism of Speech, The Hague, Mouton and Co., The Hague, Paris, 1968.
3. Louden, R., etal : Lung sounds, Am. Rev. Respir, Dis., Vol. 130, p-663, 1984.,
4. Wool bert, C.H. : The fundamental of speech (Logical and emotional), Harper and Brothers, New York, 1987.
5. Witting, C. : Physical and functional aspect of speech sounds, Uppsala Pub, Weisbaden, 1959.
6. Savithri, S.R. : Acoustics and psychological correlates of speech, J. Aco. Soc. I, Vol. XVII, p-41, 1989.
7. Heffner, R.M.S. : General Phonetics, The University of Wisconsin Press, 1952.
8. Mol, H. : Fundamental of phonetics, Mouton and Co., New York, 1963.
9. Hallel, M. : Acoustical properties of stop consonants, J. Aco. Soc. Am., Vol. 29, p-107, 1957.
10. Jacob, A.V. : Foundation of Voice Recognition and Synthesis, Part-I, Electronics for you, p-44, July, 1997.
11. Woof, J.J. : Efficient acoustic parameters for speakers recognition, J.Aco. Soc. Am., Vol. 51, p-2044, 1972.
12. Black, J.W. : Natural frequency, duration and intensity of vowels in readings, J. SHD., Vol. 14, p-216, 1949.
13. Monroe and Ehninger : Principles of Speech, Scott Foresman and Co., United states of America, 1964.

14. Cohen, A., etal : Analysis and automatic classification of breath sounds, IEEE Bio-Med. Engg., Vol. 31, p-585, 1984.

15. Zadgaonkar, A.S. : D. Litt. Thesis, Pt. Ravishankar Shukla University, Raipur (M.P.), 1992.

16. Singh, S. : Phonetic principles and practices, University Park Press, London, 1979.

17. Fredberg, T.J. : Acoustic determination of respiratory system properties, Annals Bio-Med. Engg., Vol. 9, p-463, 1981.

18. Bala S., etal : Fluctuations of Acoustic Intensity due to Bio-medical Oscillations of sound speech profile during a diurnal cycle, J. Aco., soc. I, Vol. XVII, p-215, 1989.

19. Coleman, R.O. : Male and female voice quality and its relationship to vowel formant frequencies, J.SHR., Vol. 14, p-565, 1971.