

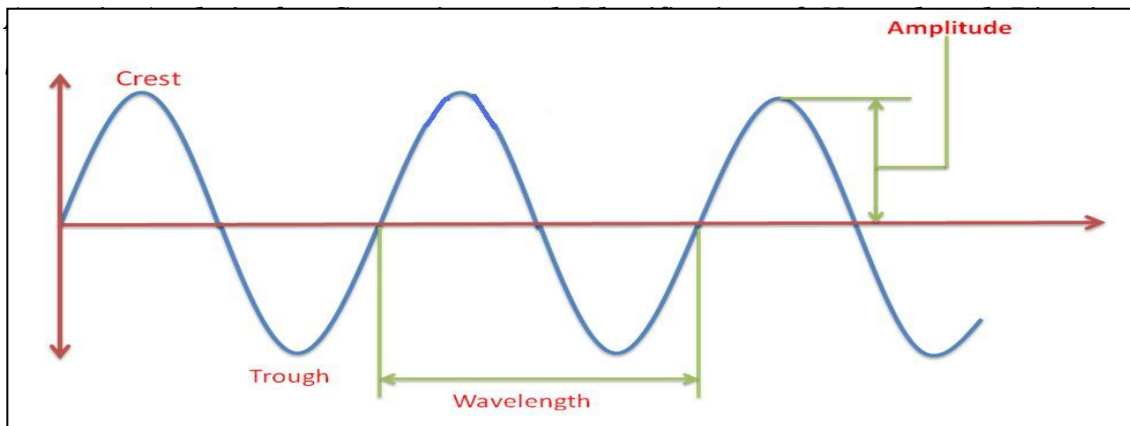
## **CHAPTER-2: MECHANISM OF SPEECH PRODUCTION AND LITERATURE REVIEW**

### ***2.1. Anatomy of speech production***

Speech is the vocal aspect of communication made by human beings. Human beings are able to express their ideas, feelings and emotions to the listener through speech generated using vocal articulators (NIDCD, June 2010). Development of speech is a constant process and requires a lot of practice. Communication is a string of events which allows the speaker to express their feelings and emotions and the listener to understand them. Speech communication is considered as a thought that is transformed into language for expression.

**Phil Rose & James R Robertson** (2002) mentioned that voices of individuals are complex in nature. They aid in providing vital information related to sex, emotional state or age of the speaker. Although evidence from DNA always remains headlines, DNA can't talk. It can't be recorded planning, carrying out or confessing to a crime. It can't be so apparently directly incriminating. Perhaps, it is these features that contribute to interest and importance of forensic speaker identification.

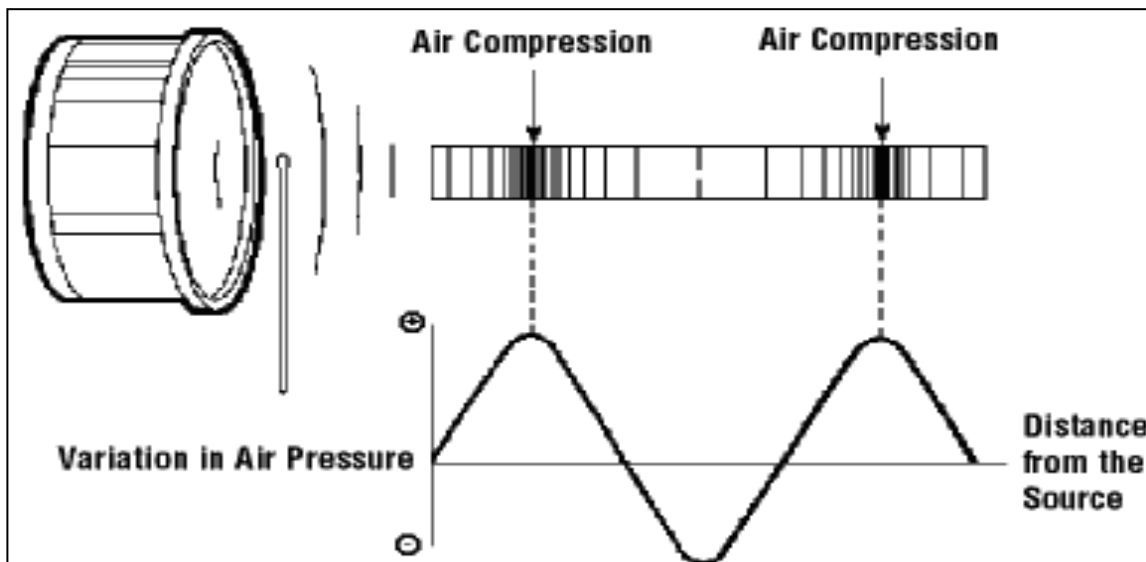
Speech signal is a multidimensional acoustic wave (as seen in figure 2.1.) which provides information about the words or message being spoken, speaker identity, language spoken, physical and mental health, race, age, sex, education level, religious orientation and background of an individual (Schetelig T. & Rabenstein R., May 1998).



**Figure 2.1.: Representation of properties of sound wave (Courtesy "Waves". Mrbigler.com. Web.)**

Often, the listeners can effectively identify a voice of a speaker who is known to him. However, the case becomes complicated whenever a listener is provided with the voice of an unknown caller.

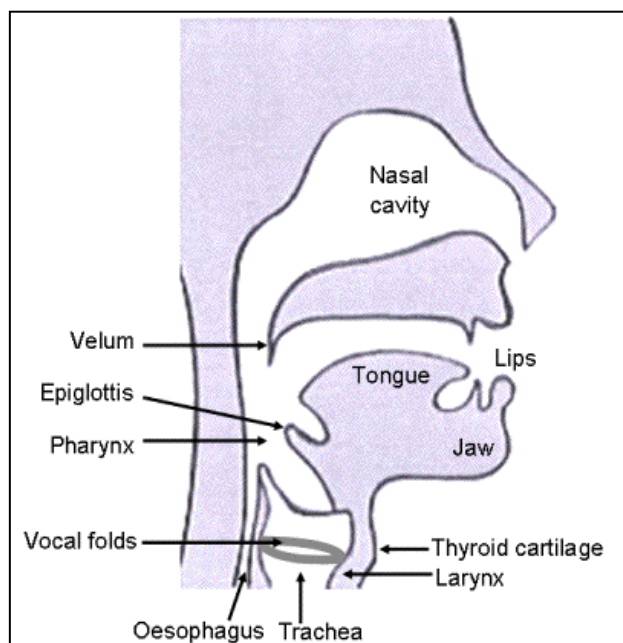
The mechanism of speech production is very complex and before conducting the analysis of any language it is important for every expert to understand the processes of production and perception of speech (D. B. Fry, 1979). Sound is the basic requirement for speech production, initiated by a simple disturbance of air molecules (as seen in figure 2.2.).



**Figure 2.2.: Generation of sound wave due to compression and rarefaction of air molecules (Courtesy Canadian Center for Occupational Health and Safety. Web.)**

## ***Acoustic Analysis for Comparison and Identification of Normal and Disguised Speech of Individuals***

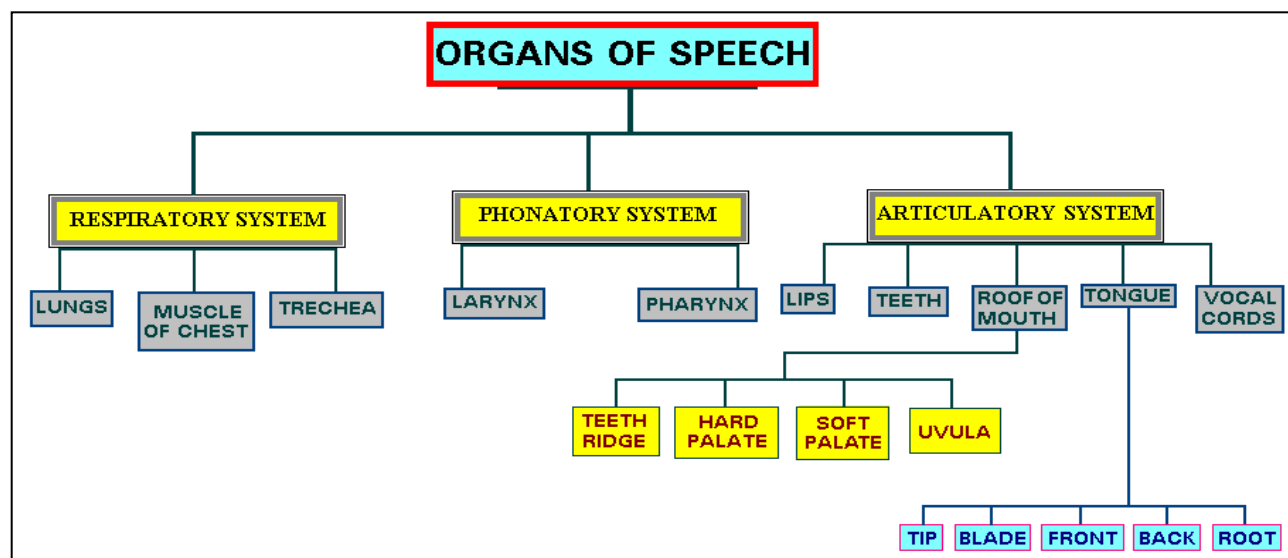
Such disturbance in the air molecules in the vocal passage is provided by movement of certain body organs such as chest muscles, vocal cords, lips, teeth, tongue, palates etc (as seen in figure 2.3.). This disturbance in the form of waves travels to the ear of the listener, who interprets the wave as sound.



**Figure 2.3.: Physiology of speech production (Courtesy "Voice Acoustics: An Introduction to the Science of Speech and Singing". Newt.phys.unsw.edu.au. Web.)**

Both human beings and animal produce sound symbols for communication purpose. The former can articulate the sound to produce and create language through which they communicate with one another. The animals produce vowel like sounds but cannot articulate and thus are unable to create a language of their own. The capability of human beings for articulation of the sound distinguishes them from the other species. The parts of the human body which are directly involved in the production of speech are usually termed as the organs of speech. There are three main organs of speech: Respiratory organs, phonatory organs and articulatory organs (as seen in figure 2.4.).

## *Acoustic Analysis for Comparison and Identification of Normal and Disguised Speech of Individuals*



**Figure 2.4.: Organs involved in the process of speech production**

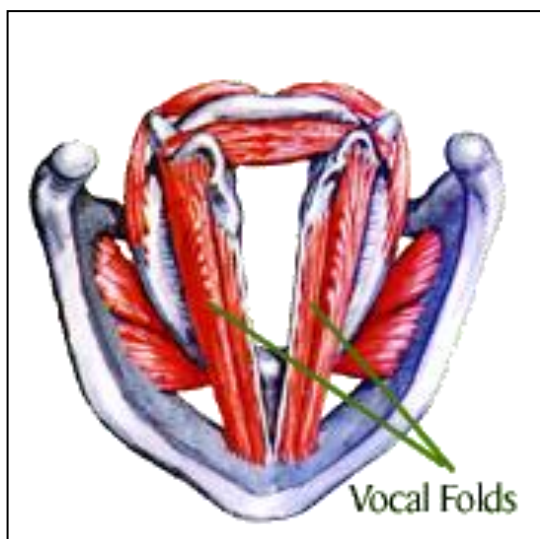
The most important function of the *lungs*, which is relevant to speech production, is respiration and it is responsible for the movement of air. Lungs are controlled by a set of muscles which make them expand and contract alternately so that the air from outside is drawn in and pushed out alternatively. When the air is pushed out of the lungs it passes through the *windpipe or trachea*, which has at its top the *larynx*.

Glottis is a passage between the two horizontal folds of elastic muscular tissues called the *vocal folds* (vocal cords), which, like a pair of lips, can close or open the glottis (i.e., the passage to the lungs through the windpipe). The main physiological function of the vocal folds is to close the lungs for its own protection at the time of eating or drinking so that no solid particles of food or liquid enter the lungs through the windpipe.

## ***Acoustic Analysis for Comparison and Identification of Normal and Disguised Speech of Individuals***

By virtue of their tissue structure, the vocal cords are capable of vibrating with different frequencies when the air passes through and this vibration is called ***voice***. After passing through the glottis and reaching the pharynx the outgoing ***air stream*** can escape either through the nasal cavity having its exit at the mouth. When the air escapes through the mouth, the nasal cavity can be closed by bringing the back part of the soft palate or the velic into close contact with the pharyngeal wall. Such a type of closure of the nasal passage is called ***velic closure***. When there is velic closure, the air can escape only from the mouth. The nasal cavity can also be kept open when the air passes through the mouth allowing part of the air to pass through the nose also. The oral passage can be closed at sometime so that the outgoing air is temporally shut up in the ***pharyngeal cavity***. In such cases, the air escapes through the nostrils creating the nasal consonants (W. Tecumseh Fitch, 2010).

During phonation, each cycle of vocal fold vibration is caused both by the sub glottal air pressure that built up to separate the folds and the Bernoulli effect which affirms that, as the air rushes through the glottis with great velocity, creates a region of low pressure against the inner sides of each fold bringing them together again (Ray D Kent & Charles Read, 2002). The whole process is made possible by the fact that the folds themselves are elastic (as seen in figure 2.5.). Their elasticity not only permits them to be blown open for each cycle, but the elastic recoil force (the force that restores any elastic body to its resting place) works along with the Bernoulli effect to close the folds for each cycle of vibration.



**Figure 2.5.: Basic structure of vocal cords (Courtesy Vocal Cord Paralysis. Slideshare.net. Web.)**

## ***Acoustic Analysis for Comparison and Identification of Normal and Disguised Speech of Individuals***

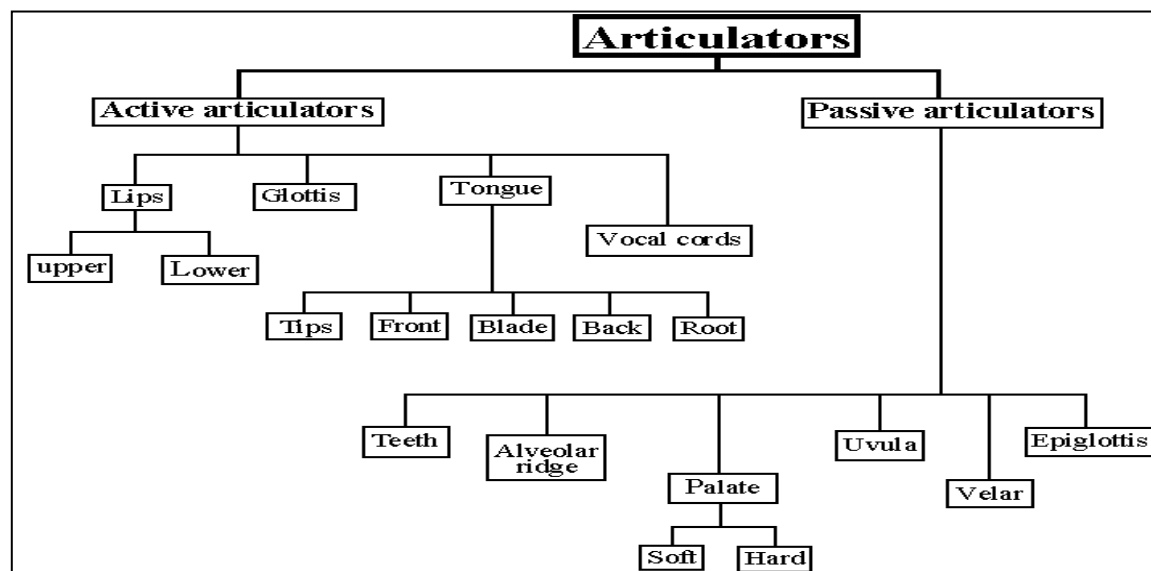
The vocal folds move in a fairly periodic way. During sustained vowels, for example, the folds open and close in a certain pattern of movement that repeats itself. This action produces a barrage of airbursts that set up an audible pressure wave (sound) at the glottis. The pressure wave of sound is also periodic; the pattern repeats itself.

**Richard L Klevans and Robert O Rodman** (1997) said that a person's mental ability to control his vocal tract muscles during utterance is learned during his childhood. These habits affect the range of sound that may be effectively produced by an individual. The range of sounds is the subset of the set of possible sounds that an individual could create with his or her personal vocal tract. It is not easy for an individual to change voluntarily these physical characteristics.

Like all sound sources that vibrate in a complex periodic fashion, the vocal folds generate a harmonic series, consisting of a fundamental frequency and many whole-number multiple of that fundamental frequency (harmonics). The fundamental frequency is the number of glottal openings/closing per second.

**Articulators:** Articulation is a process resulting in the production of speech sounds. It consists of a series of movements by a set of organs of speech called the *articulators* (as seen in figure 2.6.). The articulators that move during the process of articulation are called active articulators. Organs of speech which remain relatively motionless are called passive articulators. The points at which the articulator are moving towards or coming in to contact with certain other organ are the *place of articulation*. The type or the nature of movement made by the articulator is called the *manner of articulation*.

## *Acoustic Analysis for Comparison and Identification of Normal and Disguised Speech of Individuals*



**Figure 2.6.: Classification of articulators used for speech production**

Most of the articulators are attached to the movable lower jaw and as such lie on the lower side or the floor of the mouth. The points of articulation or most of them are attached to the immovable upper jaw and so lie on the roof of the mouth. Therefore, nearly all the articulatory description of a speech sound; therefore has to take into consideration the articulator, the point of articulation and the manner of articulation (Laurel J. Brinton, 2000).

While describing the articulators and the points of articulation, it would be convenient to take the points of articulation first. The main points of articulation are the upper lip, the upper teeth, the alveolar ridge, the hard palate and the soft palate, which is also called the velum (Henry Rogers, 2000). The upper lip and the upper teeth are easily identifiable parts in the mouth. The alveolar ridge is the rugged, uneven and elevated part just behind the upper teeth. The hard palate is the hard bony structure with a membranous covering which immediately follows the alveolar ridge. The hard palate is immediately followed by the soft palate or the velum. It is like a soft muscular sheet attached to the hard palate at one end, and ending in a pendulum like soft muscular

## ***Acoustic Analysis for Comparison and Identification of Normal and Disguised Speech of Individuals***

projection at the other which is called the Uvula (Philipp Strazny, 2005). Besides the above points of articulation, the pharyngeal wall also may be considered as a point of articulation.

The two most important articulators are the lower lip and tongue. The tongue owing to its mobility is the most versatile of articulators. The surface of the tongue is relatively large and the different points of tongue are capable of moving towards different places or point of articulation. It may be conveniently divided into different parts, viz, the front, the center, the blade, the back and the root of the tongue. When the tongue is at rest behind the lower teeth then the part of the tongue, which lies below to the hard palate towards incisor teeth, is called the front of the tongue (Philipp Strazny, 2005). The part, which faces the soft palate, is called the back and the region where the front and back meet is known as center. The whole upper surface of the tongue i.e. the part lying below the hard and soft palate is called by some scholars as the dorsum.

### ***2.2 Individuality of human voice***

When we speak our vocal cords are stretched and relaxed by muscles that are attached to them in the throat. However, speech is coordinated efforts of lungs, larynx, vocal cords, tongue, lips, mouth and facial muscles, all activated by brain. Therefore, voice quality is dependent on a number of anatomical features such as dimension of oral, pharynx, nasal cavity, shape and size of tongue and lips, position of teeth, tissue density, elasticity of these organs and density of vocal cords.

However, each person possesses unique voice quality, arising out of individual variations in vocal mechanism. The uniqueness of the voice lies in the fact that every person experimentally develops an individual and unique process of learning to speak and the chances that any two individuals will have same shape and size of vocal tract, vocal cavity and will control their articulators in same manner is extremely remote. This proves that the voice of all individuals is distinctive.



## ***Acoustic Analysis for Comparison and Identification of Normal and Disguised Speech of Individuals***

**Kersta L. G.** (1962) said that the experimental data encourages me to believe that unique identification from voiceprint can be made. It is my opinion, however, that identifiable uniqueness does exist in each voice and that masking, disguising or distorting the voice will not defeat identification if the speech is intelligible.

### ***2.3. Phonetics of speech & its branches***

Phonetics is the study of the production and perception of speech sounds. It is the study of the speech sounds that occur in all languages and gives the physical description of sounds in languages. Phonation refers to the function of the larynx and the different uses of vocal fold vibrations in speech production. It is a process by which molecules of air are set into vibration, through vibrating vocal cords.

Phonetics can be classified into three main categories:

- A. *Acoustic phonetics*: deals with the physical production and transmission of speech sound.
- B. *Auditory phonetics*: deals with the perception of speech sound.
- C. *Articulatory phonetics*: deals with the production of speech sound.

### ***2.4. Modes of phonation***

- A. Voiced sound: in such type of speech, the vocal cords vibrate and thus produce sound waves. These vibrations occur along most of the length of the glottis, and their frequency is determined by the tension in the vocal folds (Kenneth N. Stevens, 1998). All vowels and diphthongs together with some consonants like b, d, g, m, n, v, l, j, r produces voice sounds.
- B. Unvoiced sound: an unvoiced sound is characterized by the absence of its phonation. In such cases the vocal folds remain separated and the glottis is held open at all times. The opening lets the airflow passes through without creating any vibrations, but still accelerates the air by being narrower than the trachea (Michael Dobrovolsky & Francis Katamba, online

## ***Acoustic Analysis for Comparison and Identification of Normal and Disguised Speech of Individuals***

<http://catalogue.pearsoned.co.uk/>). Unvoiced sounds will not have a fundamental frequency which enables us to control pitch. Unvoiced sounds include plosive and fricative consonants like p, t, f, s, h etc.

- C. Whisper – such type of sound is created by keeping an opening between the arytenoids while the vocal folds are kept together. This changes the properties of voiced sounds, but leaves the unvoiced sounds unaffected.
- D. Creaking – these are low frequency sounds, produced when the glottis is closed, but also has settings in the vocal folds that only allow a short fragment of them to vibrate.
- E. Falsetto - is similar to voiced sounds in which vibrations occur along the whole length of the vocal folds, but attains characteristically high frequencies by shaping the vocal folds so that they get a thin edge.
- F. Murmur - is combination of voiced sounds without full closure of the glottis. Its breathy characteristic is caused by the constant airflow allowed through (Rodney Ball, Online <http://users.ecs.soton.ac.uk/kpd/book.pdf>).

### ***2.5. Classification of phonemes***

Any language is composed of a number of elemental sounds, called phonemes that combine to produce the different words of that language. English is formed by approximately 44 phonemes. Phonemes are classified into vowels, glides, semivowels and consonants.

A vowel is defined as a continuant sound (it can be produced in isolation without changing the position of articulators), voiced (using the glottis as a primary source of sounds), with no friction (noise) of air against the vocal tract passage. There are approximately 16 vowels in English. These English vowels are classified according to the position of the greatest arching point of the tongue for producing them (as seen in figure 2.7.). Considering the vertical plane, vowels are produced in a high, middle, or

*Acoustic Analysis for Comparison and Identification of Normal and Disguised Speech of Individuals*

low position of that arching point of the tongue from hard palate down (Michael Dobrovolsky & Francis Katamba, online <http://catalogue.pearsoned.co.uk/>).

Vowel	Type	Example
I	High, Front, Unrounded	R <u>i</u> d, B <u>i</u> t
E	Mid, Front, Unrounded	R <u>e</u> d, B <u>e</u> t
æ	Low, Front, Unrounded	R <u>a</u> d, <u>a</u> t
ə	Mid, Central, Unrounded	Teach <u>e</u> r
ʌ	Low, central, Unrounded	C <u>u</u> t, B <u>u</u> s
ʊ	High, Back, Rounded	P <u>u</u> t, B <u>oo</u> k
ɔ	Mid, Back, Rounded	C <u>au</u> ght, <u>o</u> n
ʌ	Low, Central, Unrounded	C <u>a</u> r, F <u>a</u> r
a	Low, Back, Unrounded	C <u>o</u> t, P <u>o</u> t
u	High, Back, Rounded	B <u>oo</u> t, Sh <u>oe</u>
o	Mid, Back, Rounded	B <u>oa</u> t, C <u>oa</u> t

**Figure 2.7.: Classification of vowels of English language**

With reference to tongue height though it is possible to differentiate many levels; generally four vertical levels are recognized as reference points. They are ‘high’ and ‘low’ forming the two extremes as the highest and the lowest possible levels and

## ***Acoustic Analysis for Comparison and Identification of Normal and Disguised Speech of Individuals***

intervening middle level being distinguished as ‘higher mid’ and ‘lower mid’. Alternative terms used for these are (1) close (2) half-close (3) half-open and (4) open. The vowels thus produced are being called (1) high or close vowels (2) higher mid or half close vowels (3) lower mid or half open vowels and (4) low or open vowels (Laurel J. Brinton, 2000).

With reference to tongue position the distinction made is of (1) front (2) central and (3) back. The vowels thus produced are being called front, central and back vowels respectively (Rodney Ball, Online <http://users.ecs.soton.ac.uk/kpd/book.pdf>).

A semivowel has the continuant and a voiced characteristic of a vowel, but some degree of friction or noise accompanies its production.

Glides can be defined as vowel –like sounds, differing from the vowels in a lack of the continuant characteristic: glides are produced rather during a fast dynamic change of the articulators; the first sounds of the word “you a” and the word “what” provide examples of glides

Consonants are phonemes other than vowels or glides. Consonants always include friction and they can be classified in to voiced and voiceless and fricative, plosive, affricative or nasal according to the manner of production.

Consonants can be classified on the basis of the manner of articulation as stop, nasals, fricatives, trills, flaps, laterals, affricates, continuants, etc (Manisha Kulshreshtha & Ramkumar Mathur, 2012).

### ***2.6. Place of articulation:***

On the basis of the points of articulation, consonants can be classified as labials, dentals, alveolar, palatals and velars etc (Mannell, R., 2009).

Labials: labials are produced when an articulator comes into contact with the upper lip as the point of articulation. Usually the lower lip acts the articulator and the resultant sound is called bilabial sound.

Dentals: dentals are produced when an articulator comes into contact with the upper teeth as the point of articulation. The tongue acts as the articulator coming into

## ***Acoustic Analysis for Comparison and Identification of Normal and Disguised Speech of Individuals***

contact with the upper teeth. The resultant sound is called a dental sound. Instead of touching the upper teeth alone the tongue may come between the upper and lower teeth. Then the sound produced is called an interdental sound. Sometimes the lower lip may act as the articulator coming into contact with the upper teeth, then the sound produced is called a labio-dental sound.

Alveolar: An Alveolar sound is produced when an articulator, rather tip or the blade of the tongue, comes into contact with the alveolar ridge. The sound thus produced is called an alveolar sound.

Palatals: A palatal sound is produced when an articulator, usually the front of the tongue, comes into contact with the hard palate as the point of articulation.

Velars: A velar sound is produced when an articulator usually the back of the tongue, comes into contact with the soft palate or velum as the point of articulation.