

Lecture Notes- Vocal anatomy

Sanjeev S Koni

Assistant Professor

Introduction-why speaker recognition?

Forensic speaker identification is the application of science to solve the problems related to identification of the unknown speaker in criminal investigation. Voice is defined as a series of words. Although DNA is considered as the most important evidence for criminal investigation, but the truth is that DNA can't talk. The voice of a person can be successfully used as a biometric feature as it is well accepted by the users and can be easily recorded using microphones and hardware of low costs. It can provide an unconventional and more secure means of permitting entry without any need of remembering a password, lock combination etc or the use of keys, magnetic card or any other fallible device which can be easily stolen. In the present era, widely available facilities of telephones, mobiles and tape recorders results in the misuse of the device and thus, making them an efficient tool in commission of criminal offences such as kidnapping, extortion, blackmail threats, obscene calls, anonymous calls, harassment calls, ransom calls, terrorist calls, match fixing etc. The criminals nowadays are more frequently misusing these modes of communication, believing that they will remain incognito, and nobody would recognize them. It is fortunately no longer true. The voice of an individual can successfully recognize him and pin the crime on him.

What is speech?

Speech is the vocal aspect of communication made by human beings. Human beings express their ideas, thoughts and feelings orally to one another through a series of complex movements that alter and mold the basic tone created by voice into specific, decodable sounds. Speech wave may be uniquely specified in terms of source and filter characteristics. Development of speech is a regular 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 considered as a thought that is transformed into language for expression.

Speech signal is a multidimensional acoustic wave (fig 1), which conveys the information about the words or message being spoken, identity of the speaker, language

spoken, pathology of speech, physical and mental health of the speaker. It also contains the features that may reveal geographical origin, ethnicity or race, age, sex, education level, religious orientation and background of an individual.

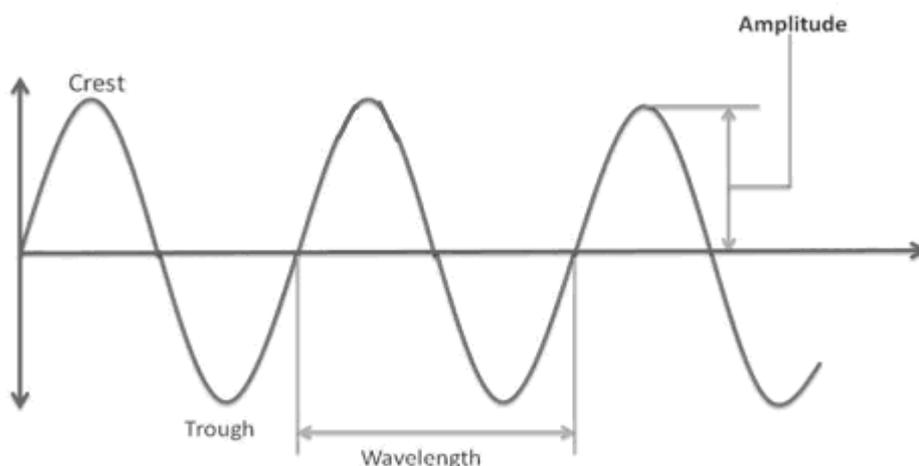


Fig 1: Propagation of sound wave

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.

Speech is a compelling biometric for several well known reasons and particularly because it is the only one available modality in a large set of situations.

Physiological arena of speech production

The mechanism of speech is a very complex one and to undertake analysis of any language it is important to understand the processes that go to make up the message that a speaker transmits and a listener receives. The primary source of sound production is disturbance in air and such disturbance in the speech sound is provided by movement of certain body organs such as chest muscles, vocal cords, lips, teeth, tongue etc (fig 2). This disturbance in the form of sound waves travels to the ear of the listener, who interprets the wave as sound.

The mechanism for generating the human voice is divided into three parts: the respiratory system (produces the adequate airflow or air pressure to vibrate the vocal folds), phonatory system (including vocal folds of larynx) and articulators (parts of vocal tract, above the larynx consisting of tongue, palate, cheek, lips etc, which articulates or filter the sound coming out from the larynx) (fig 3 & fig 4). The air from the external environment is drawn into the lungs during inhalation. It remains stored in the lungs for a short period and finally expelled from the lungs under pressure by the process of exhalation. The exhalation process involves the

transfer of air under pressure from lungs to larynx. The larynx contains of a set lip like structure called vocal folds that set the molecules of this breath stream into vibration. For sound to be produced, these molecules have to vibrate at a rate that falls within a particular range. The process by which molecules of air are set into vibration is known as phonation.

The vibration pattern of molecules produced by phonation is complex. It contains a wide range of frequencies and has a buzzing sound. This buzz is moulded into speech sounds by vocal tract. The vocal tract consists of the pharynx (throat), oral cavity and nasal cavity. The configuration, or shape, of the vocal tract at a particular moment determines what speech sound will be produced. The configuration of the vocal tract can be changed by movement of several structures within it specifically, the tongue, lips, lower jaw and soft palate.

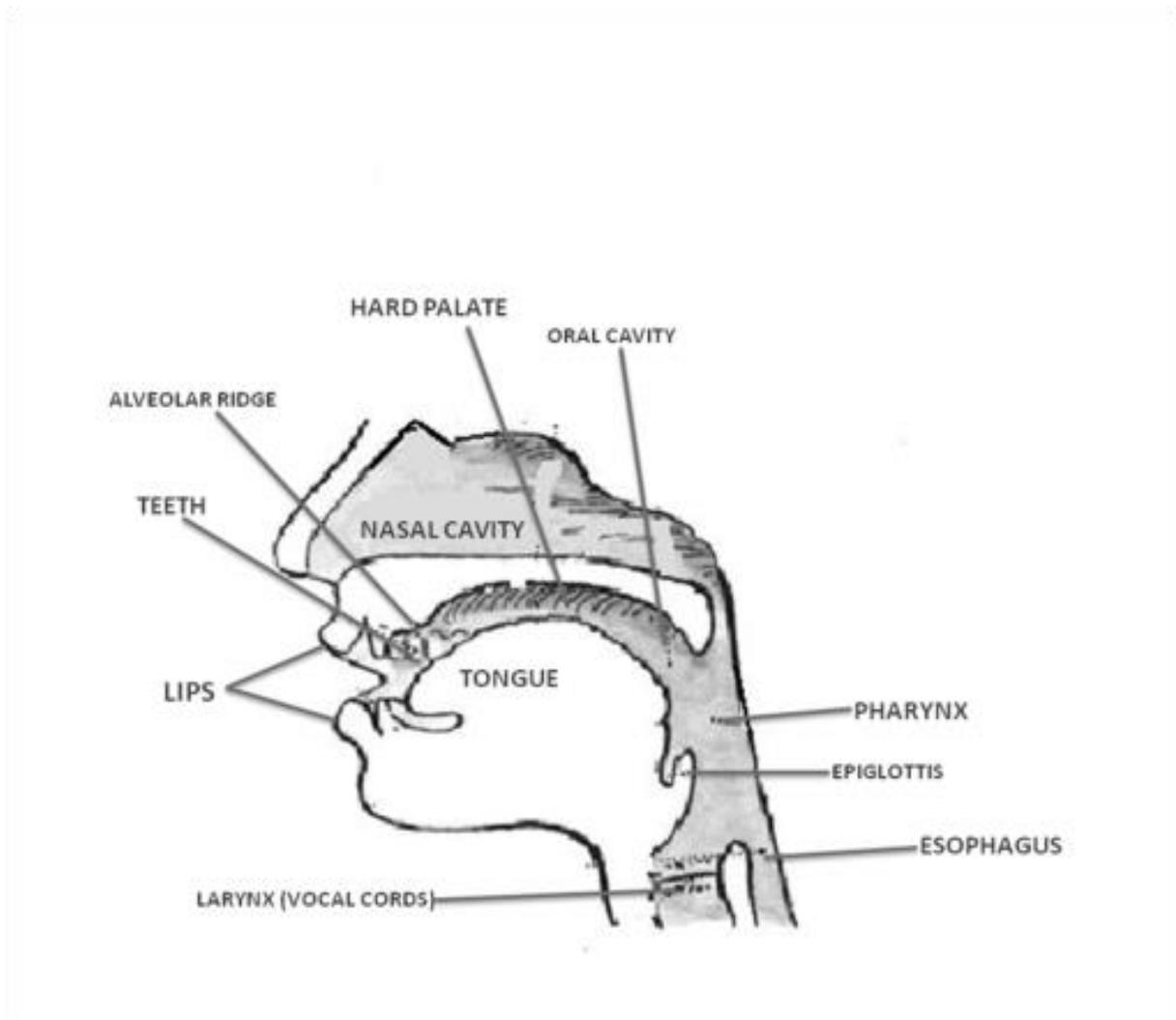


Fig 2: Diagrammatic representation of Vocal cavity

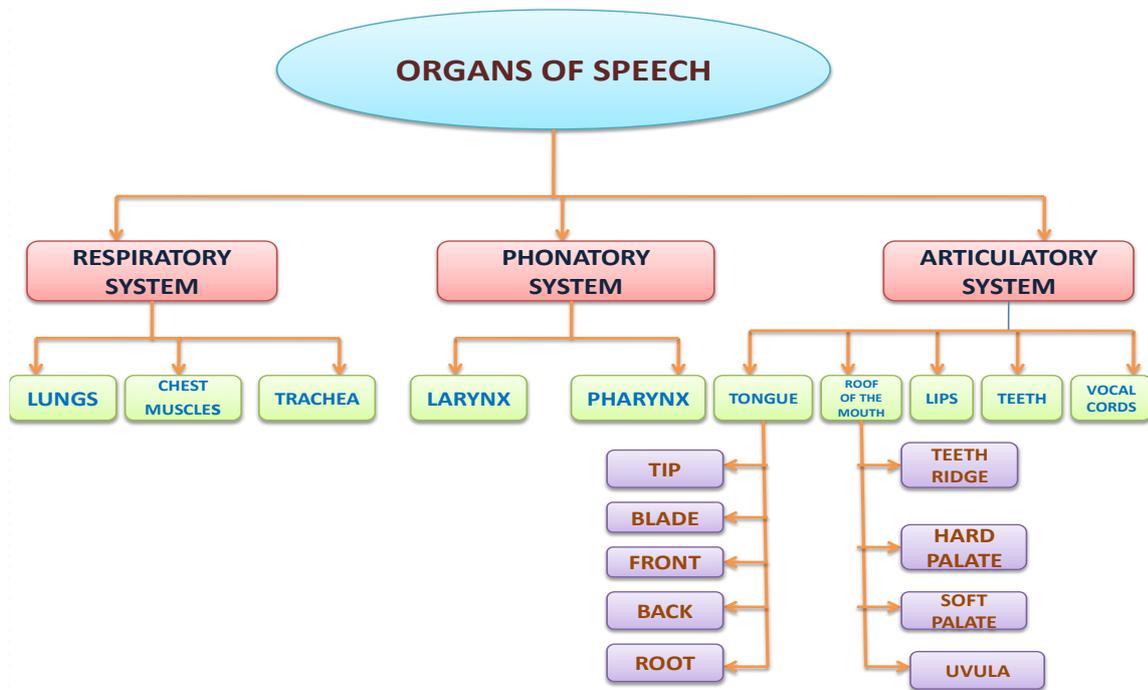


Fig 3: Organs of speech

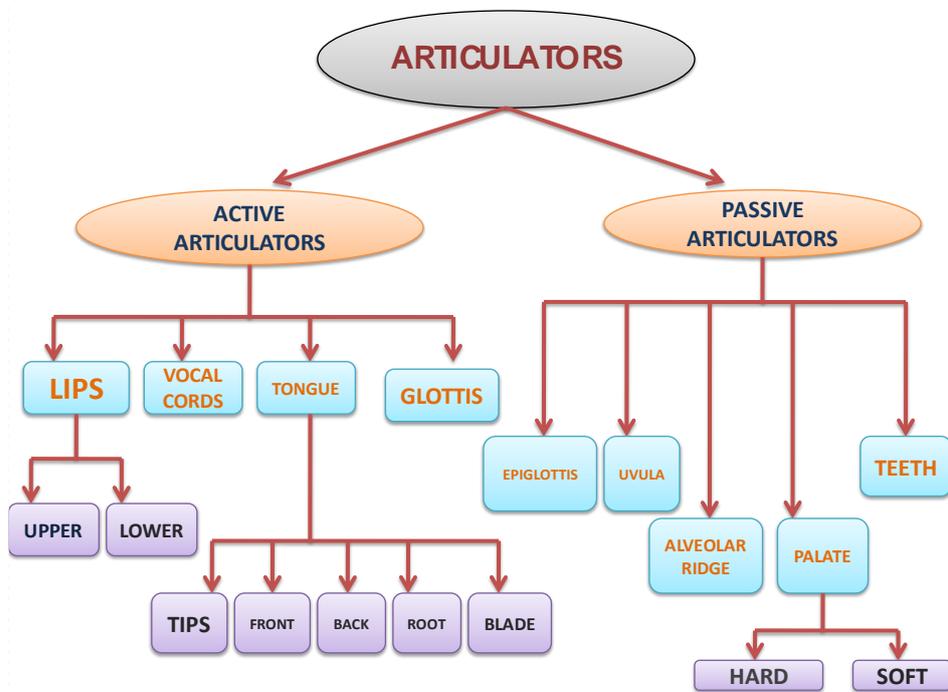


Fig 4: Types of articulators

Uniqueness in speech sound

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 are activated by brain. Therefore, voice quality is dependent on 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 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 quite small. This proves that the voice of all individuals is distinctive.