Your Voice & How it Works

Anatomy: Parts of the Voice

- **Larynx (pronounced LAIR-inx, not LAHR-nix)**
  The larynx is the voice box. The vocal folds (also called vocal cords) are part of the larynx. The vocal folds vibrate to create the sound of the voice.

- **Pharynx (pronounced FAIR-inx)**
  The pharynx is the throat. It goes up from the larynx and divides into the laryngopharynx (just above the larynx), oropharynx (going into the mouth) and nasopharynx (going into the nose).

- **Trachea (pronounced TRAY-key-ah)**

  The trachea is your windpipe. It's the tube that connects your lungs to your throat. The larynx sits on the top of the trachea.

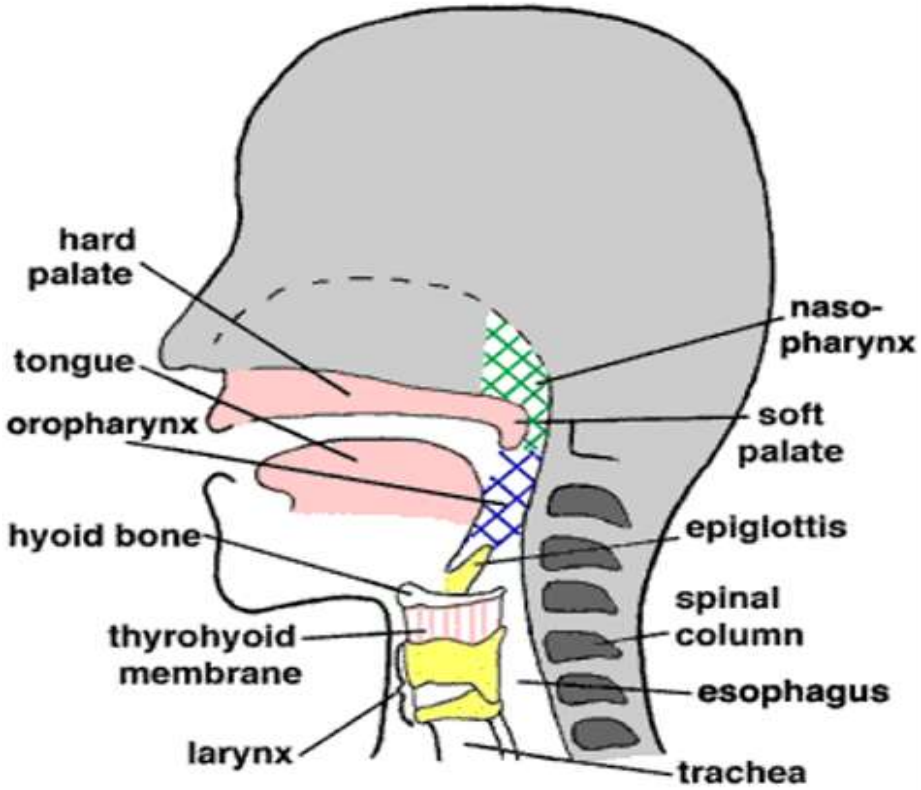


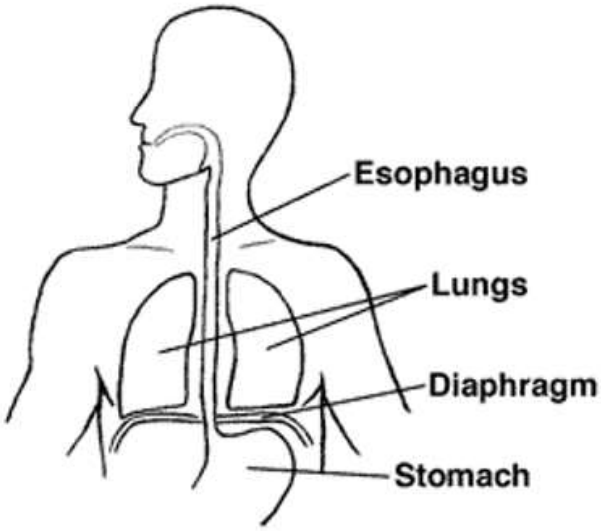
Figure 1: *Notice we didn't label the brain. Of course the brain is crucial for voice production*

Some other nearby organs:

- **Esophagus**
  The esophagus is your food pipe. It's just behind the larynx and trachea. Your pharynx carries both air and food/water. The air goes through the larynx and trachea, and food and water go into your esophagus.
- **Spinal column**

  The spinal column is behind the esophagus. You can feel it by pressing the back of your neck.
- **Diaphragm**

  The diaphragm is underneath the lungs, inside the rib cage. It's shaped like a dome. The diaphragm is your main muscle for controlling respiration (breathing).



Mechanisms of Vocal Production

The **Physiological Basis of Vocalizing There are three main physiological components of vocal production:** the subglottal system, the larynx, and the supralaryngeal vocal tract (Lieberman & Blumstein, 1988). Figure 2 illustrates the three main physiological components of vocal production, showing the approximate locations of the vocal organs.

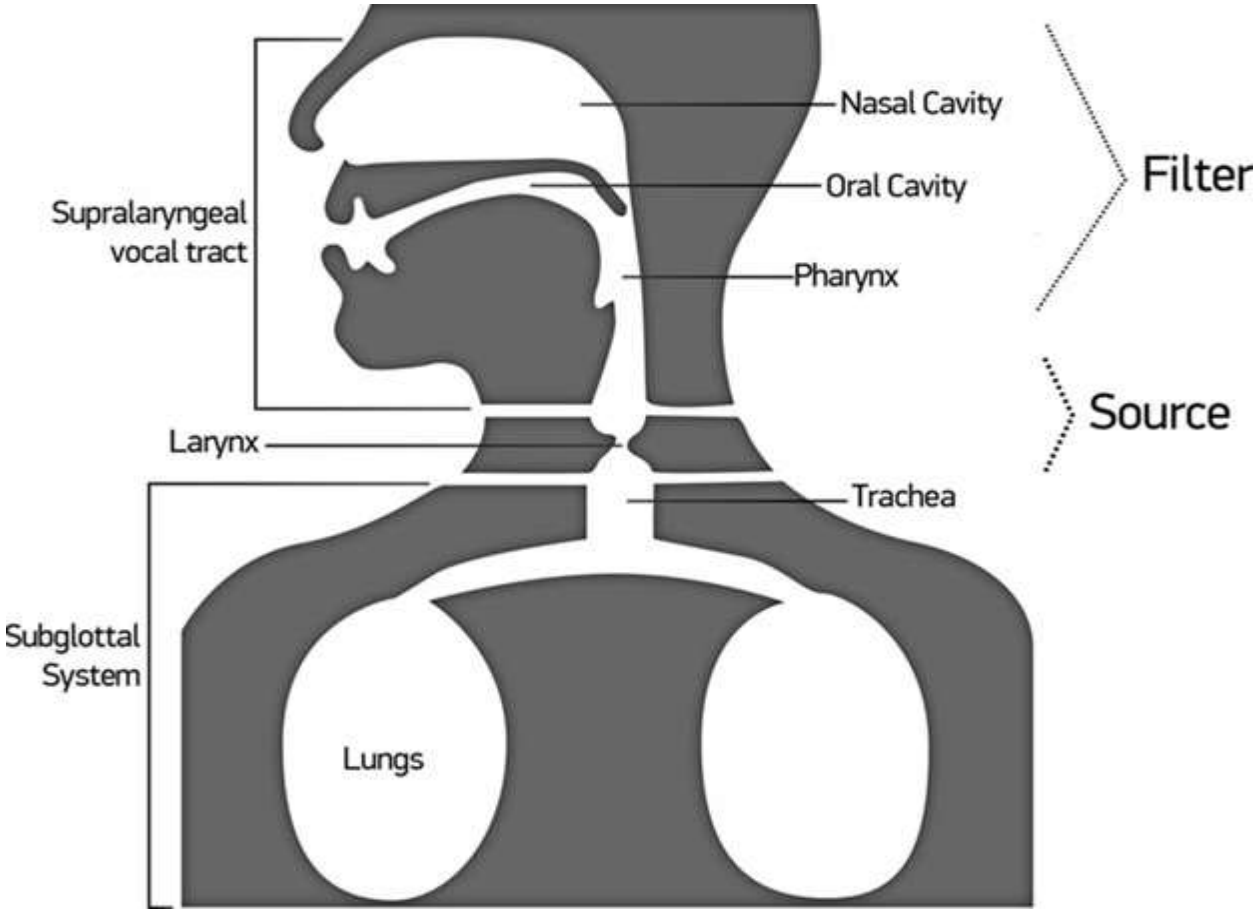


Figure 2. Illustration of the physiological basis of vocal production. *Left*: The three main physiological components of voice production: **the subglottal system, the larynx, and the supralaryngeal vocal tract**

The subglottal system generates airflow that supplies the energy necessary for vocal production. The trachea (windpipe) is a single tube branching into two airways, each leading into a lung in the subglottal system. Lungs are mammals' main respiratory organs, and their volume can be increased or decreased during vocal production by contracting the expiratory or inspiratory muscles.

The larynx converts the airflow into sound through the vibration of vocal folds. The pressure of the air coming from the lungs blows the vocal folds apart and sucks them together. Cyclic closure of the vocal folds relies on a negative pressure inside the glottis

The basic rate of vocal fold vibration is called fundamental frequency (f), which is the primary determinant of perceived pitch. **Voice intensity is the energy through the unit area carried by a soundwave, which increases as the amplitude of the soundwave increases. Intensity and amplitude are perceived as loudness.** Voice quality is the auditory coloring of the voice, which is determined by how the vocal folds vibrate

Acoustic parameter	Perceptual correlate	Definition
f_0 (Fundamental Frequency)	Pitch	The basic rate of vocal fold vibration
$F_1, F_2, \ldots$ (Formant Frequencies)	Vowel sound	Frequency peaks in the spectrum with high degree of energy
Voice Intensity	Loudness	Energy through the unit area carried by a soundwave

The Role of the Nervous System

The nervous system is responsible for coordinating the activities of the respiratory, laryngeal, and articulatory systems as they work together to produce complex speech sounds. Some aspects are governed by the central nervous system, which controls conscious processes. For example, you consciously move your lips in ways that produce different vowel and consonant sounds.

At the same time, the autonomic nervous system is also at play and controls unconscious body functions, like heartbeat and digestion. It also affects physiological tremor, which refers to the oscillations of muscles associated with the respiratory, laryngeal, and articulatory systems. The movements of all these muscles influence the qualities of the sound waves that are produced when a person speaks.

Anatomy of speech production

Speech signal is a multidimensional acoustic wave (as seen in figure 1.) which provides information about the words or message being spoken, speaker identity, language spoken, physical and mental health etc.

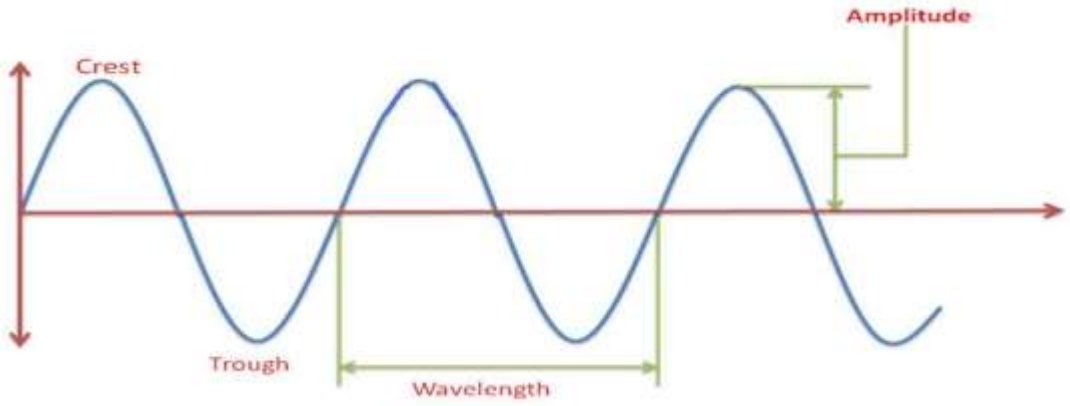


Figure 1.: Representation of properties of sound wave

Sound is the basic requirement for speech production, initiated by a simple disturbance of air molecules

(as seen in figure 2.).

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.

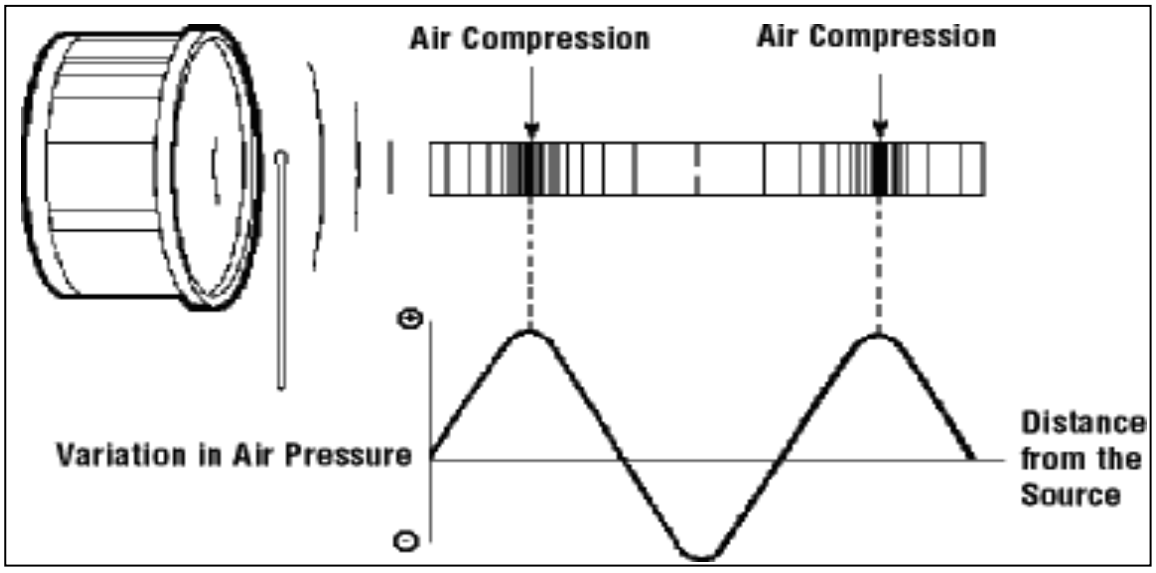


Figure 2: Generation of sound wave due to compression and rarefaction of air molecules

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 3.).**

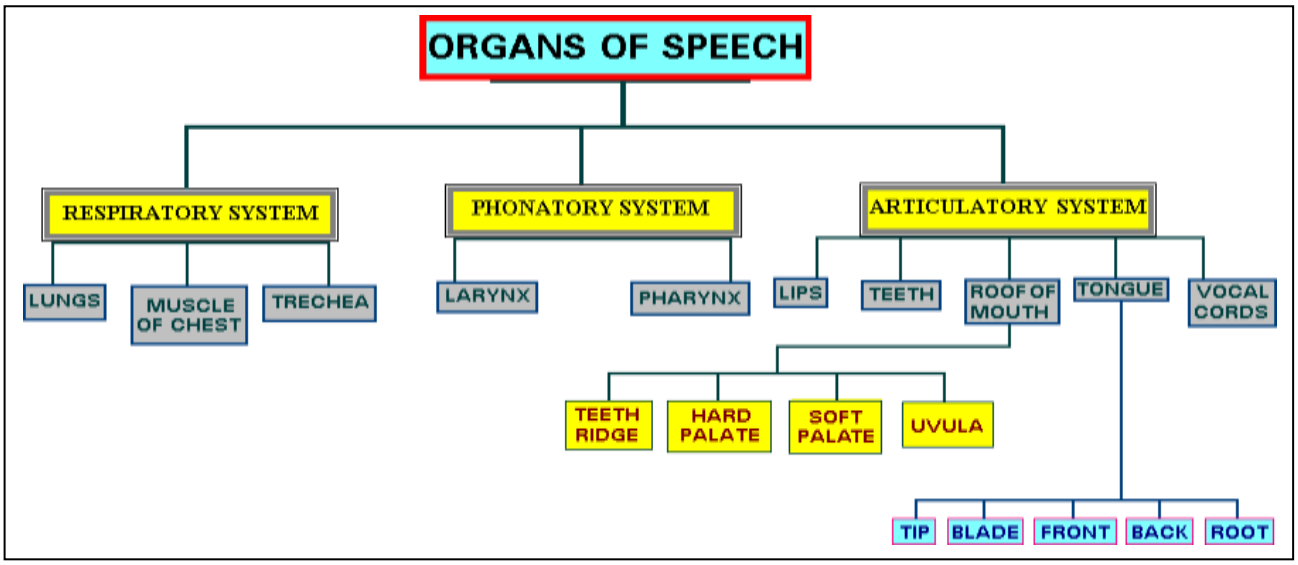


Figure 3.: Organs involved in the process of speech production

Dictionary meaning of Phoneme, Syllable and Phone:

Phoneme:

one of the smallest units of speech that make one word different from another word:

Syllable:

a word or part of a word which contains one vowel sound

a **phone** (from the Greek fōnḗ) is a distinct speech sound.

Categorization of speech sounds:

In Previous section we described the anatomy of speech production, the vocal folds and vocal tract being the two primary components, and described the mechanism of speech production, i.e., how we generate sounds with our speech anatomy and physiology. We saw that a sound source can be created with either the vocal folds or with a constriction in the vocal tract, and, based on the various sound sources, we proposed a general categorization of speech sounds.

1. **The nature of the source**: periodic, noisy, or impulsive, and combinations of the three;
2. **The shape of the vocal tract**. The shape is described primarily with respect to the place of the tongue hump along the oral tract and the degree of the constriction of the hump, sometimes referred to as the place and manner-of-articulation, respectively. The shape of the vocal tract is also determined by possible connection to the nasal passage by way of the velum;

3. **The time-domain waveform which gives the pressure change with time at the lips output;**
4. **The time-varying spectral characteristics revealed through the spectrogram.**

With these four speech descriptors, we embark on a brief study of the classification of speech sounds. We focus on the English language, but from time to time discuss characteristics of other languages.

Elements of a Language

A fundamental distinctive unit of a language is the phoneme; the phoneme is distinctive in the sense that it is a speech sound class that differentiates words of a language. For example, the words "cat," "bat," and "hat" consist of three speech sounds, the first of which gives each word its distinctive meaning, being from different phoneme classes. We saw earlier, and we will discuss further below, that many sounds provide this distinctive meaning, and such sounds represent a particular phoneme. To emphasize the distinction between the concept of a phoneme and sounds that convey a phoneme, the speech scientist uses the term phone to mean a particular instantiation of a phoneme.

Different languages contain different phoneme sets. **Syllables contain one or more phonemes, while words are formed with one or more syllables, concatenated to form phrases and sentences.**

There are various ways to study speech sounds that make up phoneme classes; the use of the above first two descriptors in this study is sometimes referred to as **articulatory phonetics**, while using the last two is referred to as **acoustic phonetics**.

One broad **phoneme classification for English is in terms of vowels, consonants, diphthongs, affricates, and semi-vowels.** Figure 3.17 shows this classification, along with various subgroups, where each phoneme symbol is written within slashes according to both the International Phonetic Alphabet and an orthographic (alphabetic spelling) representation

Phonemes arise from a combination of vocal fold and vocal tract articulatory features. Articulatory features, corresponding to the first two descriptors above, include the vocal fold state, i.e., whether the vocal folds are vibrating or open; the tongue position and height, i.e., whether it is in the front, central, or back along the palate and whether its constriction is partial or complete; and the velum state, i.e., whether a sound is nasal or not. It has been hypothesized that the first step in the production of a phone is to conceive in the brain the set of articulatory features that correspond to a phoneme. A particular set of speech muscles is responsible for "activating" each feature with certain relative timing. It is these features that we may store in our brain for the representation of a phoneme. In English, the combinations of features are such to give **40 phonemes**

Human languages use many different sounds and in order to compare them linguists must be able to describe sounds in a way that is language independent. Speech sounds can be described in a number of ways. Most commonly speech sounds are referred to by the mouth movements needed to produce them. **Consonants and vowels are two gross categories that phoneticians define by the movements in a speech sound. More fine-grained descriptors are parameters such as place of articulation.**

What is a consonant sound?

•A consonant sound is a speech sound that is produced by the partial or complete obstruction of air by the lips, teeth, tongue or throat. The Collins Dictionary defines a consonant sound as "a sound such as 'p', 'f', 'n', or 't' which you pronounce by stopping the air flowing freely through your mouth".

Phonemes arise from a combination of vocal fold and vocal tract articulatory features. Articulatory features, corresponding to the first two descriptors above, include the vocal fold state, i.e., **whether the vocal folds are vibrating or open; the tongue position and height, i.e., whether it is in the front, central, or back along the palate and whether its constriction is partial or complete; and the velum state, i.e., whether a sound is nasal or not**

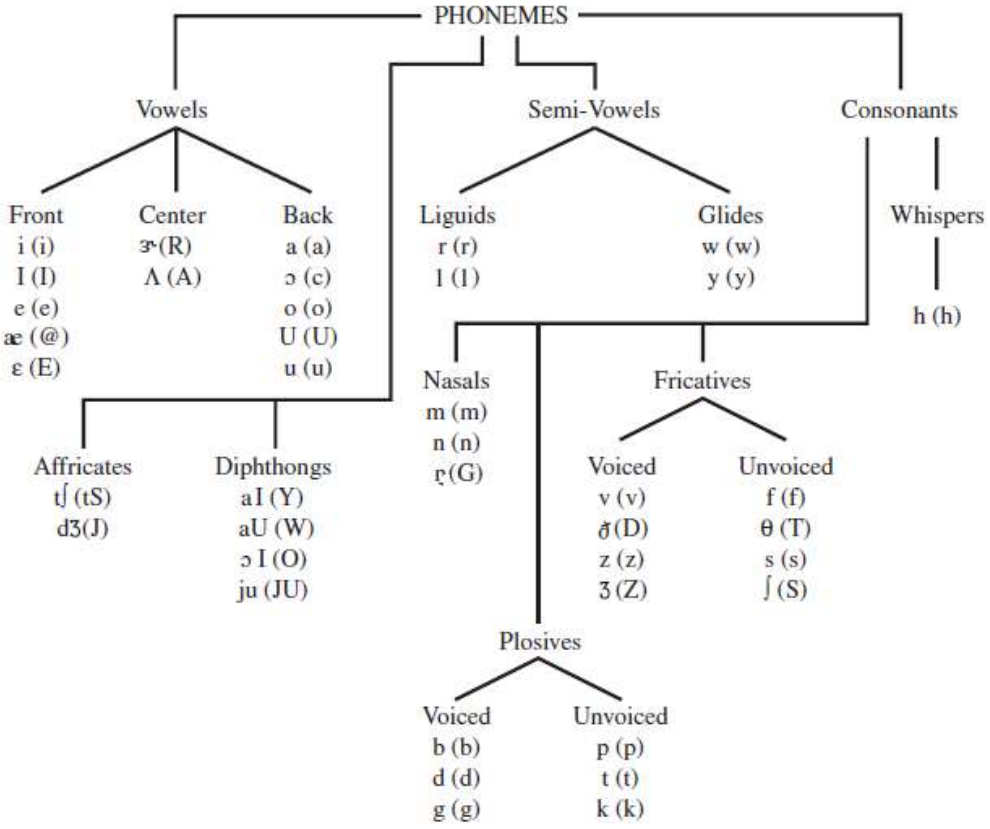


Figure 3.17 Phonemes in American English [6],[32]. Orthographic symbols are given in parentheses to the left of the International Phonetic Alphabet symbols.

Speech sounds, also known as phonemes, can be categorized based on various linguistic features. The primary categorizations include:

1. **Manner of Articulation**

    - **Stops (or Plosives):** Complete closure of the vocal tract, followed by a sudden release of air. Examples include /p/, /b/, /t/, /d/, /k/, /g/.
    - **Fricatives:** Partial closure of the vocal tract, causing friction. Examples include /f/, /v/, /s/, /z/, /ʃ/ (as in "sh"), /ʒ/ (as in "zh").
    - **Affricates:** Combination of a stop followed by a fricative. Examples include /tʃ/ (as in "ch") and /dʒ/ (as in "j").
    - **Nasals:** Airflow through the nasal cavity. Examples include /m/, /n/, /ŋ/ (as in "sing").
    - **Liquids:** Partial closure of the vocal tract, allowing air to flow around. Examples include /l/ and /ɹ/ (as in "r").
    - **Glides (or Semivowels):** Slight closure in the vocal tract. Examples include /j/ (as in "y") and /w/.
2. **Place of Articulation:**
    - **Bilabials:** Sounds produced with both lips. Examples include /p/, /b/, /m/.
    - **Labiodentals:** Sounds involving the lower lip and upper teeth. Examples include /f/, /v/.
    - **Alveolars:** Sounds produced at the alveolar ridge. Examples include /t/, /d/, /s/, /z/.
    - **Palatals:** Sounds involving the hard palate. Examples include /ʃ/ (as in "sh"), /ʒ/ (as in "zh").
    - **Velars:** Sounds produced with the back of the tongue against the soft palate. Examples include /k/, /g/.
    - **Glottals:** Sounds produced with the glottis. Examples include /h/.

3. **Voicing:**
    - **Voiced:** Sounds produced with vibration of the vocal cords. Examples include /b/, /d/, /g/.
    - **Voiceless:** Sounds produced without vibration of the vocal cords. Examples include /p/, /t/, /k/.

4. **Vowel Features**
    - Vowels are categorized based on tongue height, tongue backness, and lip rounding. Examples include /i/ (high front unrounded), /e/ (mid front unrounded), /u/ (high back rounded), /o/ (mid back rounded), /æ/ (low front unrounded), /ɑ/ (low back unrounded).

5. **Prosody**
    - Intonation, rhythm, and stress patterns contribute to the overall meaning and expression in speech. This includes features such as pitch, loudness, and duration.

These categorizations are fundamental in the study of phonetics and phonology, helping linguists analyze and describe the sounds of human languages.

Prosody refers to the rhythm, intonation, and expressive aspects of speech that go beyond the basic structure of individual words. *It encompasses the melodic and rhythmic features of speech, including* ***pitch, stress, tempo, and rhythm****. Prosody plays a crucial role in conveying meaning, emotion, and emphasis in spoken language.*

Key components of prosody include:

1. **Pitch:** The perceived frequency of a person's voice, often described as high or low. Pitch variations convey emotional nuances and can indicate the speaker's attitude or emphasis.

2. **Stress:** The emphasis placed on certain words or syllables within a sentence. Stress patterns help to convey the intended meaning and significance of specific elements in speech.

3. **Tempo:** The speed or rate at which someone speaks. Tempo variations can convey excitement, urgency, or calmness.

4. **Rhythm**:** The pattern of stressed and unstressed syllables, contributing to the overall flow and musicality of speech.

5. **Intonation**:** The variation in pitch within a sentence or utterance. Intonation patterns help convey the speaker's attitude, emotion, and the intended meaning of a statement.

Prosody is an essential aspect of effective communication, as it adds richness and nuance to spoken language. It allows speakers to express emotions, convey emphasis, and provide additional context beyond the literal meaning of words. In natural conversation, prosody helps listeners interpret the speaker's intentions and emotional state. Disorders or challenges related to prosody can impact communication and social interactions.