

Lecture 1: Speech Phonetics

Lecturer: K.R. Chowdhary

: Professor of CS

Disclaimer: These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.

1.1 Phonetics

The human speech recognition process can be represented as shown in figure 1.1.

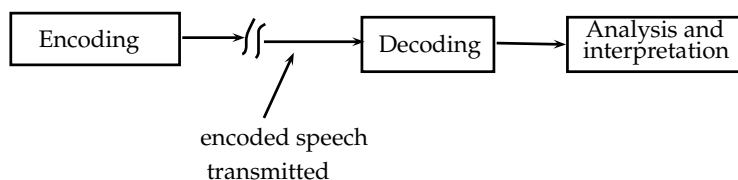


Figure 1.1: Phonetic Process.

The *encoding* is a mental and psychological process, and achieved by commands sent to vocal organ by the brain. Through scientific study, the details and mechanism of speech produced by vocal organs has been studied. It has been found out, as how the speech is perceived by human listener, and how the different sounds are combined to form syllables, words, and sentences. This study is called *phonetics* and corresponding knowledge is called *phonology*.

There are three branches of phonetics:

1. *Articulatory phonetics*: It deals with articulation, i.e., production of speech sounds.
2. *Acoustics phonetics*: It deals with study of sound waves, and various properties like amplitude, frequency, and their effects.
3. *Auditory Phonetics*: It is study of reception and comprehension of speech.

The speech is not the primary function of speech producing organs, since it gets produced due to air flow from lungs to / through these organs. The air flow gets obstructed, while passing through, for example, *larynx* and *vocal cord*, as shown in figure 1.2.

Some of the examples of sounds produced are given below. The letter enclosed with symbols // is representation of sound of that letter.

//p/, //t/ : vocal folds are wide open

//h/ : vocal folds are brought closer (called fricative sound)

//b/, //d/, //g/ : sound is produced due to vibration in *glottis*

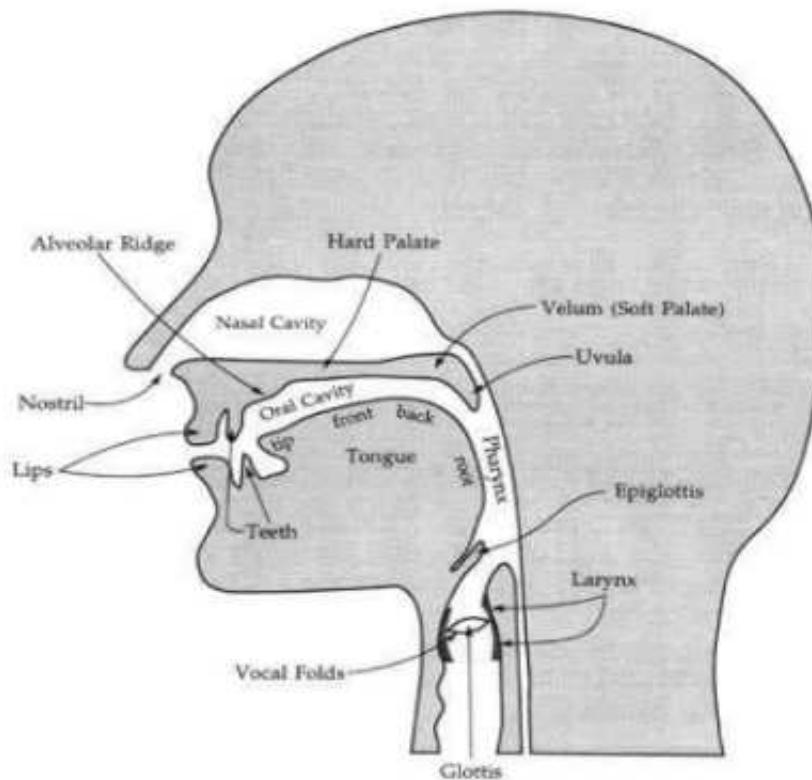


Figure 1.2: Human Phonetic Speech Organs.

When the soft palate is closed, the air flows towards the nose, called *nasal sound*. The sound produced by *m, n* due to this is represented by /m/, /n/.

Uvula is a soft hanging part at the end of mouth, it is visible when the mouth is wide open. It vibrates during the articulation of the sound 'r'.

Consonants and vowels combined make *syllables*. So syllable is a vowel like sound together with some of the surrounding consonants. The IPA (International Phonetic Alphabets) symbol [.] is used to separate the syllables. For example *parsley*, and *catnip*, each have two syllables, as [par.sli] and [kaet.nip], respectively. The word *tarragon* has three, [tae.re.gan].

1.2 Speech sounds

The speech sounds are classified as vowels and consonants. The vowels are produced due to free-flow (unobstructed) of sound, while during the pronouncing of consonants some kind of obstruction exists. However, there is no one-to-one correspondence between the letter, e.g. of English, and the corresponding sound. The same letter in different words produce the different sounds, as in the following cases. For example, consider the letter *a*'s sound:

- In the word 'above', the *a* is of short sound, like *r* in 'father',

- in the word ‘age’, *a* has sound like in word ‘made’,
- in the word ‘all’, the *a* has sound like in word ‘ball’.

Following are more examples.

- The letter *d* in ‘damage’, ‘educate’, ‘picked’- has different sounds.
- The letter *s* in ‘sign’, ‘pleasure’, and ‘resign’ has different sounds.
- The following words have silence sound of some letter: knee, lamb, receipt, right, doubt, know, island, hour. In the preceding words, for letters *k*, *b*, *p*, *h*, *b*, *k*, *s*, *h*, respectively, are not pronounced.

Fortunately, the speech has *phonetic alphabets*, and each type of sound corresponds to some phonetic alphabet. And, these alphabets are finite – limited to say, about 100. These are universally accepted in all the countries, and speech processing systems. They are specified by *International Phonetics association* (IPA), called IPA alphabets.

1.2.1 Vowels

The sound of different vowels are:

- Vowel [i]: pronounced as [iy], highest point of tongue is toward to of mouth
- Vowel [a]: pronounced as [ae], tongue is at low front,
- Vowel [u]: pronounced as [uw], tongue at high back

When the tongue is raised towards front, the corresponding vowels are called *front vowels*, and when it is raised back, they are called *back vowels*.

1.2.2 Consonants

The *pulmonic* air stream, while passing from lungs through mouth is modified by different organs of speech. All the consonants are produced like this. When speech sound are produced, the vocal folds (cords) either vibrate or not.

- Vibrate: voiced consonants (have initial sound), like in *bat*,
- Non-vibrate: voiceless consonants (have no initial vibratory sound at vocal cord), like in *pat*.

The raising and lowering of *soft-palate* (*velum*) shuts down the nasal passage, for example the initial sound of word “king”. During this time the air passes only through oral passage. When velum is lowered (i.e. no closure) only the nasal sounds are produced, e.g., the final sound in the word *king*.

Place of articulation. These are points at which air flow can be modified. For example, in the sounds /t/, /d/, /s/, /n/, the tip of the blade (i.e. tongue) move towards the teeth ridge. These sounds are called *alveolar* sounds.

Manner of articulation. This refers to obstruction to airflow caused by narrowing/closing of articulators. It may be complete closure, approximate closure, or open approximate.

For example, [d] in *din*: first complete closure at teeth ridge, then sudden release of air. The other example is, [s] in *sing*, where front of tongue is raised towards alveolar ridge to form a narrow constriction for partial airflow. So in both these cases, the place of articulation is same, but the difference is manner in which articulation is done.

1.3 Transcription

IPA's phonetic alphabets helps us to recognize the human speech through universally acknowledged symbols which are very convenient way of showing where and how a particular sound is produced. The process of representing speech using IPA is called *transcription*.

Phonetic Transcription. It is kind of phonetic script or phonetic notation, which is visual representation of speech sounds (or phones). The most common type is phonetic transcription uses phonetic alphabet, such as IPA.

The transcription is divided into two classes;

1. *Narrow transcription:* Also called *phonetic transcription*, and refers the way speech is written in phonetic alphabets giving us many aspects of pronunciation possible. For example, the phonetic alphabets of *p*, *k*, *t* are represented as /p/, /k/, /t/ which are aspirated in the initial position. Many times, there is aspiration in the middle or at the end of a word.

The symbols [p^h], [k^h], [t^h] are used for representation of aspiration in the initial stage. The [] brackets stand for narrow transcription.

2. *Broad Transcription:* These are also called *phonetic transcriptions*, and do not give details of the pronunciation but concentrates on how the word differs from other in pronunciation. For example, *pen*, and *ben*, have initial contrasts in initial sounds (differs), which cause the change in the meaning of the words. We call them as phonemes of the language concerned.

1.4 Phonemes

A phoneme /foʊni:m/ is a basic unit of language phonology, which is combined with other phonemes to form meaningful units, called *morphemes*. The phoneme can be described as “the smallest contrastive linguistic unit which may bring about a change of meaning.

There are hundreds of different sounds, called *phones*, that our vocal cord can produce. However, each language uses only a few of these. The study of sounds of a language is called *phonology*. Phonology is study of mental organization of sound system of any language. In this biggest units are *syllables*. Smaller units are segments, called *allophones*. A *Phonologist* studies these sounds in contrasts.

Phonetics is concerned with physical properties of speech sounds, e.g., how they are produced and received physically by human beings.

A *phoneme* is an abstract mental unit that represents speech sound (but it is not sound). The feature of phoneme is that they contrast with each other. For example, *pat* and *bat* are identical but contrast in sound gives them different meaning. Therefore, /p/ and /b/ are two different phonemes (it is convenient to denote

the contrasting sounds within slant lines to differentiate them from non-contrasting sounds). The pair here (called minimal pair) differ in only one sound segment.

1.5 Syllable structures

A *syllable* is combining sounds to form meaningful words or parts of words. Following are examples:

Word	No. of syllables	Syllables
<i>tin</i>	1	[tin]
<i>brother</i>	2	[broth.er]
<i>important</i>	3	[im.por.tant]
<i>computer</i>	3	[com.put.er]

How to count the Syllables. There are more than one algorithms for counting syllables in a word. The first is based on how many times you hear the vowel sound.

Algorithm : 1

1. Say the word
2. How many times you hear the sound of *a, e, i, o, u*.
3. The count is number of syllables in this word.

The other method is based on the concept that robot speaks the word.

Algorithm :2

1. Say the word as the robot says,
2. Pay attention to pauses you make,
3. Number of broken words is number of syllables.

Example for the word robot is: ro *pause* bot, 2 syllables.

We represent *V* (vowels) as sound preceded or followed with consonant *C* or cluster of consonants *CC*, *CCC*. Some examples of representation of phonemes for English words are:

Word	No. of syllables	Phoneme
<i>I</i>	1	/ai/
<i>eye</i>	1	/ai/
<i>owe</i>	1	/aʊ/

The other examples of syllables are:

be, phoneme is /bi:/ structure: CV

ill, phoneme: /il/, structure: VC

bill, /bil/, CVC

tie, /tai/, CV

egg, /eg/, /VC/

beg, /beg/, CVC

no, /naʊ/, CV

on, /an/, VC

out, /aʊt/, VC

range, /reindz/, CVVC

snow, /snaʊ/, CCV

1.6 Word and Sentence Stresses

Certain syllables consist of prominence which makes them stand out from the rest of the syllables in a word or sentence. It is stressed if it is uttered with greater muscular power. There are languages where each syllable receives equal prominence. The features which make a syllable stressed are; loudness, pitch, and vowel quality.

Exercises

1. What are the syllables for the following words?

Notebook, Pen, Pearson, Jodhpur, Rajasthan, Sanskrit, Hindi, Urdu, Junjabi.

2. Describe the sound of following alphabets, as which sound organs produce sound, and in what order.
a, b, c, d, w, x, y, z.
3. Draw a diagram of phonetic speech organs of mouth, and explain the function of each.
4. How can you judge whether some one pronouncing the alphabets are vowels or consonants? Imagine that you have never heard the name of the language to which these alphabets belong.
5. What is transcription? Explain.
6. Define the following terms: Phone, Phoneme, Syllable, Phonetic, Phonology.
7. What are the challenges of speech recognition? Explain in brief.
8. For the Hindi language, answer the following in brief:
 - (a) List the vowels, and consonants of Hindi language
 - (b) Are there some classification of these vowels?
 - (c) Give some examples of syllables of Hindi language.
 - (d) Are the number of vowels and consonants different in Hindi and Sanskrit? Justify.
9. What can be the major applications of Speech recognition?
10. What can be the major applications of TTS?

References

- [1] D. JURAFSKY AND J. MARTIN, "Speech and Language Processing," *Pearson India*, 2002, Chapter 1.