

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 Introduction

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

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

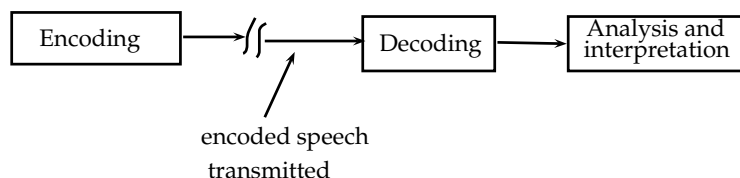


Figure 1.1: Phonetic Process.

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 generated by speaker and perceived by a 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 [Jurafsky]:

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 through these organs. The air flow gets obstructed, while passing through, for example, *larynx*, *vocal cord* and other organs like tongue, lips, and uvula, as shown in Fig. 1.2.

The Fig. 1.3 demonstrates the mechanical model of human vocal system. It shows that air is pumped by lungs through muscle power, which passes through pharyngeal cavity, then enters to one of the two destinations: oral cavity and nasal cavity, and some times divided into both. Finally the air is released through nose and mouth.

Some of the examples of sounds produced are given below. A letter enclosed with symbols // is taken as representation of sound of that letter, as the following example illustrates.

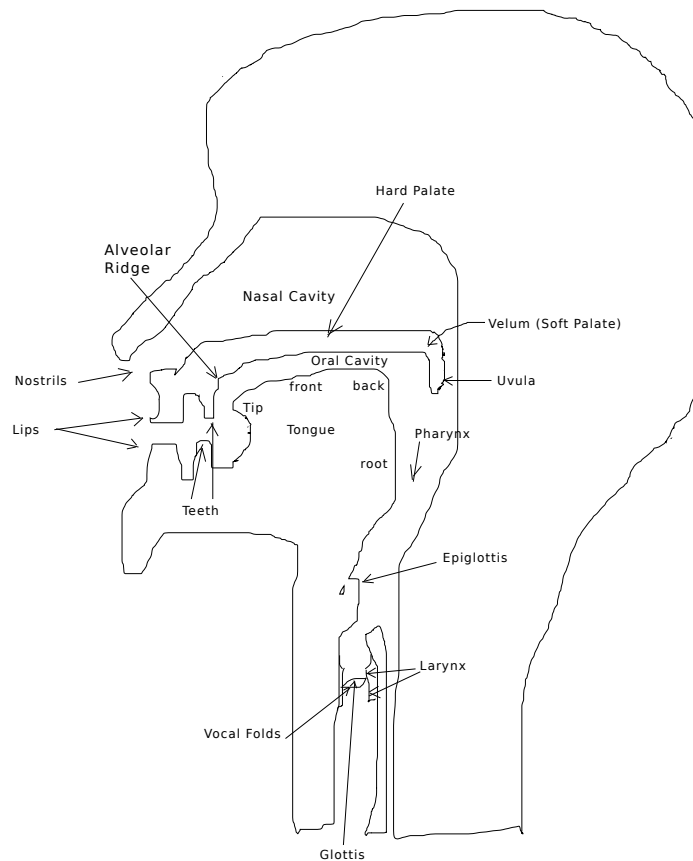


Figure 1.2: Human Phonetic Speech Organs

/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*

When the soft palate is closed, the air flows towards the nose, called *nasal sound*. At this time, air does not flow out of mouth. The nasal sound example is, the sound produced by pronouncing the letters *m*, *n* and it is represented by /m/, /n/, respectively.

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 of letter 'r'.

The consonants and vowels combined make *syllables*. A 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].

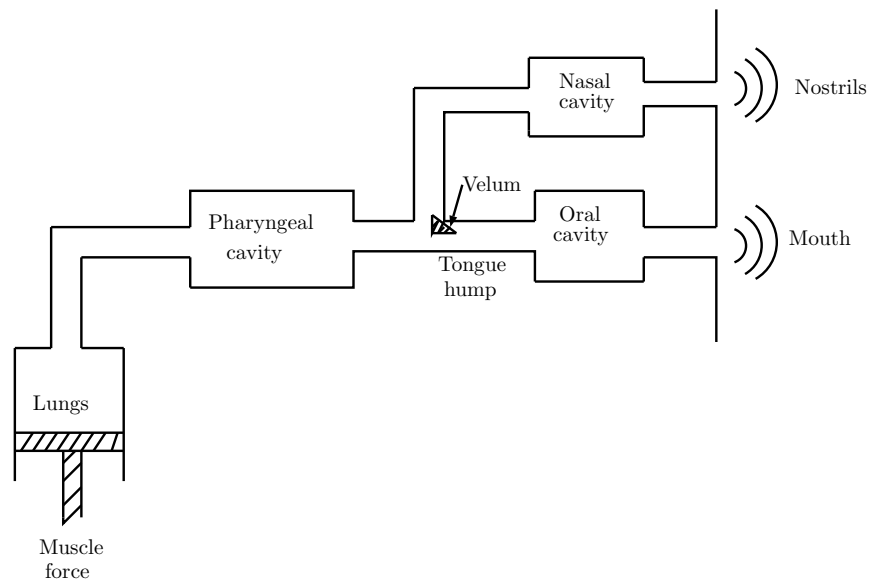


Figure 1.3: Human Phonetic Speech Organs' model

1.2 Speech sounds

The speech sounds are classified as *vowels* and *consonants*. The vowels are produced due to free-flow (i.e., unobstructed flow) of sound, while during the pronouncing of consonants, some kind of obstruction exists. However, there is no one-to-one correspondence between an alphabet, e.g. of English, and the corresponding sound. The same alphabet 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 letter *a* is of short sound, like *r* in 'father',
- in the word 'age', the *a* has sound like in word 'made',
- in the word 'all', the *a* has sound like in word 'ball'.

Following are more some 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 one 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.

Due to above mentioned reasons, which shows that no uniform rule is adopted for pronouncing the letters of any language, and that is the reason of complexity of speech recognition.

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

The vowels in English language are classified into two categories: 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 vocal cords mouth is modified by different organs of speech. All the consonants are produced this way. When speech sound are produced, the vocal folds (cords) either vibrate or not.

- Vibrate: The voiced consonants are produced and have initial sound, like in *bat*,
- Non-vibrate: This produces voiceless consonants, i.e., having no initial vibratory sound at vocal cord), like in *pat*.

Note that the vibration is produced at vocal cord.

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*.

The consonants are therefore classified based on: 1. place of articulation, and 2. manner of articulation.

Place of articulation These are points at which air flow can be modified while pronouncing the consonants. For example, in the sounds /t/, /d/, /s/, /n/, the tip of the blade (i.e. tongue) move towards the teeth ridge step-by-step which each letter is pronounced. 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 useful in 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 phonetic transcription is divided into two classes;

Narrow transcription These phonetic transcriptions 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, in p, k, t, respectively. The [] brackets stand for narrow transcription.

Broad Transcription These *phonetic transcriptions* 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/ (name of phone) is a basic unit of language phonology, which is combined with other phonemes to form meaningful units of words, 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. 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.

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.

Different speech sounds, that are realization of the same phoneme call called *allophones*. The allphone variation is usually conditional, such that certain phoneme is realized as a certain allphone in a particular phonological environment.

A phoneme is a sound or a group of different sounds perceived to have same function by speaker of language or dialect in question. For example, English phoneme /k/ which occurs in words as *cat*, *kit*, *skit*. The phoneme 'k' in *kit* has aspirated sound but 'k' in *skill* has not aspirated sound¹ k is represented as [k^h] and unaspirated as [k]. This shows that [k] and [k^h] are allophones for the phoneme /k/.

¹In linguistics, aspirated means, the puff of air accompanying the release of a plosive consonant. Also, a sound produced by such puff is called aspirated.

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]

1.5 Syllable structures

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

Note that, syllable is a unit of pronunciation having one vowel sound with or without surrounding consonants, and it forms a word or part of a word. For example, "water" has two syllables, and "inferno" has three syllables. A syllable is a unit of organization of a sequence of speech sounds. Syllables are often considered as phonological building blocks of words. They can influence the rhythm of a language, its prosody, its poetic measure, and its stress patterns. The speech can usually be divided up into number of syllables, e.g., the word "ignite" is composed of two syllables: ig and nite.

The syllable writing began several hundred years before the first letters. The earliest recorded syllables are on tablets, which were written around 2800 BC in the Sumerain city of Ur.

The IPA is an alphabetic system of phonetic notation based primarily on Latin alphabets. it was devised by the International Phonetic Association in late 19th century as a standardized representation of the sound s of spoken language.

Counting 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. Thus, two syllables.

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

The other examples of syllables are:

Word	No. of syllables	Phoneme representation in IPA
<i>I</i>	1	/aɪ/
<i>eye</i>	1	/aɪ/
<i>owe</i>	1	/aʊ/

Word	Phoneme	Structure
<i>be</i>	/bi:/	CV
<i>ill</i>	/il/	VC
<i>bill</i>	/bil/	CVC
<i>tie</i>	/tai/	CV
<i>egg</i>	/eg/	/VC/
<i>beg</i>	/beg/	CVC
<i>no</i>	/naʊ/	CV
<i>on</i>	/an/	VC
<i>out</i>	/aʊt/	VC
<i>range</i>	/reɪndʒ/	CVVC
<i>snow</i>	/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.

Review Questions

1. What are the articulators?
2. How are the front vowels different from back vowels?
3. What is the difference between vibrating and non-vibrating consonants?
4. What are the alveolar sounds? Which letters produce these sounds?
5. Give examples of words, where the letters *a*, *s*, *d* are pronounced differently.
6. Define the following terms: Phone, Phoneme, Syllable, Phonetics, Phonology.
7. 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.
8. What is transcription? Explain.

Exercises

1. What are the syllables for the following words?

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

2. Explain the process of articulation sound generated by pronouncing of letters /d/ and /s/.
3. What do you understand by phonetic alphabets? Give any 10 examples of phonetic alphabets. What is significance or application of these alphabets?
4. Give the classification of Phonetics, and explain each type of it in brief.
5. Give algorithms to count the syllables of a word.
6. Describe the sound of following alphabets, as which sound organs produce sound, and in what order.
a, b, c, d, w, x, y, z.
7. Draw a diagram of phonetic speech organs of human, and explain the function of each.
8. In any language, the number of consonants are far more in count than the number of vowels. Can you give some classification of these consonants for English language.
9. What are the challenges of speech recognition? Explain in brief.
10. 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.
11. Give list of 5 different phone sequences which our vocal cords produce and that are not part of English language.

References

- [1] D. Jurafsky and J. Martin, *Speech and Language Processing, 3rd Ed.*, Pearson India, isbn: 3257227892, Nov. 2005.