

4 Phonetics

Speech is a very hierarchical and complex physical phenomenon, including issues related to cognition, language, physiology, hearing and acoustics. A research including features of these fields is referred to as *phonetics*. In a view of speech processing we need to focus on lower categories that concern, for example, following issues:

What kind of phonemes do exist? What is the fundamental frequency/what are the formants at a given time? How the speech should be coded? What phonemes can be found in the speech sample?

To answer these questions one must make clear of basic knowledge on speech signal and speech itself.

Most of the acoustic properties of speech stem from the properties of the speech production system which will now be studied and later on pursued to model.

Speech Organs

Nice figures related to this text can be found in

<http://www.opiskelijakirjasto.lib.helsinki.fi/fonterm/006.htm> (text in Finnish) or

<http://www.phon.ox.ac.uk/~jcoleman/phonation.htm> (english)

Acoustically thinking, speech consists of the variation of air pressure that has compressed air in *lungs* as a power source. When inhaling, the diaphragm and the intercostal muscles tighten causing the chest to expand, yielding low pressure in the lungs and air flowing in. In exhaling the muscles relax causing the chest to shrink and the air flows out because of high pressure. Speech is present almost exclusively in exhaling. What is good to keep in mind is that the speech organs (lungs, tongue, vocal cords etc.) have originally developed to enable other human activities like breathing and eating, and have only later adapted to speech production.

The *larynx* is a sophisticated organ, whose part is to change the air flow from lungs into periodic pressure fluctuation. The larynx comprises of the following parts: *the thyroid cartilage (adam's apple)*, *the vocal folds* and *the arytenoid cartilage*. The gap between vocal cords is called *the fissure of glottis* or *the glottis* and it comprises of *the vocal glottis* (between the vocal folds) and of *the cartilaginous glottis* (between the arytenoid cartilages). Homo sapiens can control the shape of the glottis in diversified positions by using the muscles in the larynx.

The *vocal tract* is a general term for the speech organs after larynx. It splits in parts: *the pharyngeal cavity*, *the nasal cavity* and *the oral cavity*.

The most important organs in the vocal tract are *the tongue*, *the soft palate*, *the lower jaw* and *the lips*. The tongue is the most important organ for speech production: its different postures determine the most of phonemes. The soft palate is a muscle that can separate the oral cavity from the nasal cavity.

Short English-Finnish vocabulary:

alveolar ridge hammasvalli

arytenoid cartilage kannurusto

bronchus keuhkoputki

cartilage rusto
cricoid cartilage rengasrusto
diaphragm pallea
epiglottis kurkunkansi
false vocal folds taskuhuulet
glottis äänirako
hyoid bone kieliluu
larynx kurkunpää
lungs keuhkot
nasal cavity nenäontelo
palate kitalaki
thyroid cartilage kilpirusto
tongue kieli
trachea henkitorvi
vocal folds äänihuulet
vocal tract ääntöväylä
oral pharynx, pharyngeal cavity nieluontelo
pharynx nielu
uvula kitakieleke
velum, soft palate kitapurje

Speech Production

Speech is produced when the air originating from the lungs passes the vocal folds and gets modulated in the vocal tract. The sound at the glottis can be interpreted as an excitation filtered by the vocal tract. The air flow as itself is soundless, and so the phonemes are produced by making changes to the air flow. The most important mechanisms are the vibration of the vocal cords causing periodic excitation, forming narrow passage in the vocal tract causing noisy excitation, and closing the vocal tract entirely in some place and then removing the closure, which causes an impulse-like excitation.

In the following we present the different types of excitation:

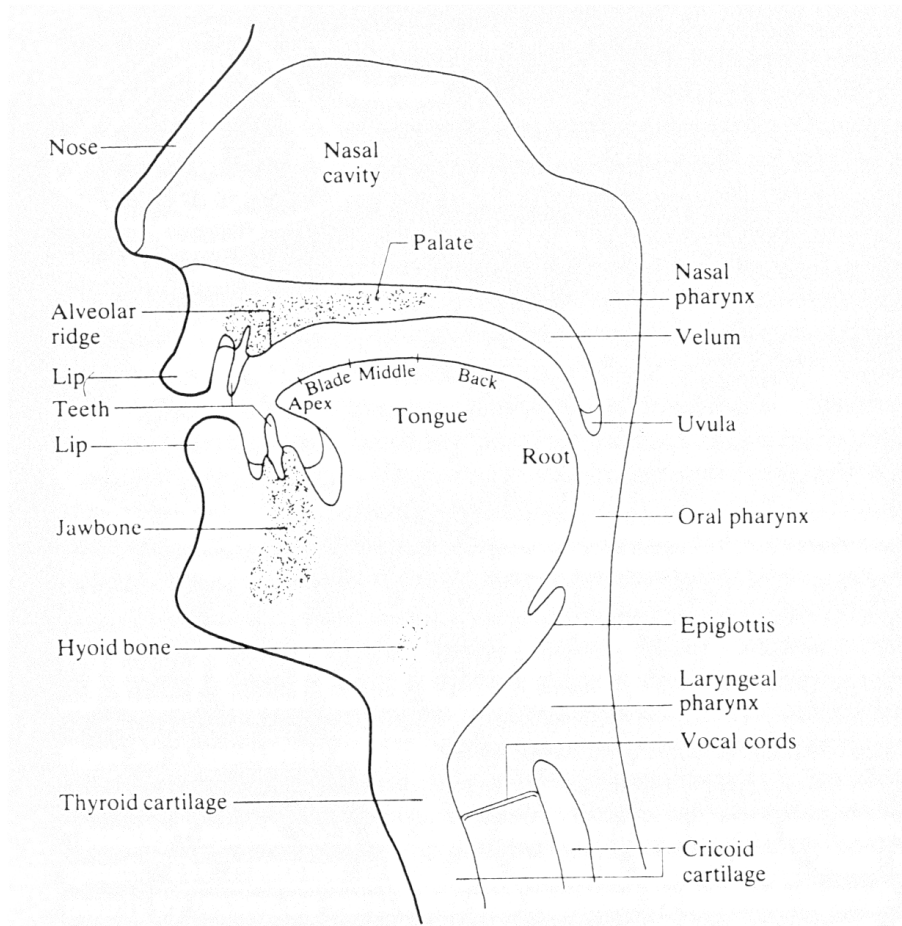


Figure 1: Principal organs of speech (T.W. Parsons, Voice and Speech Processing, McGraw-Hill, Inc., 1987. s. 63).

phonation the most important type of excitation caused by regular vibration of the vocal folds. The glottis opens and closes periodically causing chopping up of the air flow, typically on frequency of 50-500 Hz depending on the speaker. One open-close-sequence is called a period and its duration determines the fundamental frequency. Phonemes with vibrating vocal folds are called voiced sounds (e.g. vowels and resonant sounds).

breathing during normal breathing (no talk) the glottis is wide open.

unvoiced sounds the glottis is more or less open and the vocal folds are not vibrating.

whisper during whisper the vocal glottis is closed and the cartilaginous glottis is open causing friction (or friction noise).

The modulation modifies (filters) the excitation according to the posture of the speech organs. Each phoneme is related to a typical posture and shape of the vocal tract which can be modeled as an acoustic tube. The shape of the tube builds up certain resonancies, called formants, that can be seen in the spectrum of the sound. Formants are of fundamental importance in the modulation performed

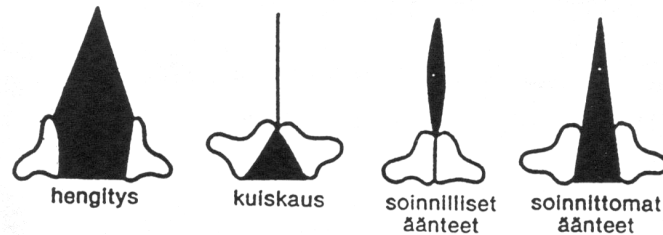


Figure 2: Position of vocal folds in different phonation types. From left to right: breathing, whisper, phonation, unvoiced.

by the vocal tract; the classification of all vowels and some of the consonants can be done by looking at the formants (frequency, strength and bandwidth).

The following sections describe the essentials of how the speech organs produce a certain phoneme and what are the acoustic properties of the sound due to the configuration of the speech organs. *Articulatory phonetics* covers the science of configurations of the speech organs to produce a particular phoneme. *Acoustic phonetics* is the science of the relations between the configurations of the speech organs and the acoustic wave. Acoustic phonetics will be discussed in Section 4.2.

4.1 Articulatory phonetics

One important goal of phonetics is to classify all phonemes existing in all languages. For this purpose *International phonetic alphabet* (IPA) was created in 1888. Most scientist are unanimous of the classification but for notational reasons (there are no symbols of IPA in typewriters) other phonetic alphabets are also used, for instance Arpabet.

IPA-classification can be found in

<http://www.arts.gla.ac.uk/IPA/fullchart.html>

Phonemes can be considered on the *phonetical* basis, meaning that the examination is not based on any particular language, but the description of the phonemes is as accurate as possible based on their articulation (posture of speech organs). Another approach is the *phonological* one, in which one examines the phonemes of a particular language, especially the classes of phonemes which will be interpreted to be same in the language in question. For instance, [k] and [p] are different sounds in Finnish because the word *kala* (fish) is different than *pala* (piece). Meanwhile, the phonemes [s] (in the word “pass”) and [sh] (in the word “shot”) do not change the meaning of any word in Finnish, so they are interpreted as the same sound in Finnish. In Russian, for instance, they are different sounds. We will briefly return to the phonological approach after the description of the phonemes.

Description of Phonemes

Speech organs can create numerous distinct phonemes, but only a part of them will be applied in speech. A usefull grouping is given below.

Vowels The phonemes of all know languages can be divided into vowels and consonants. Vowels are voiced phonemes during which the vocal tract is open. However, in some languages, it may be useful to fine-tune this definition. For example, in Finnish, vowels can be defined as sounds ’in which

the voice has an unrestrained way out through the middle part of the mouth' (by this definition, we get rid of the nasals [n] and [m], and the lateral [l]).

The vowels can be further classified based on the following properties:

- the posture of the tongue
- the roundness of the lips
- how nasal is the sound

The position of the narrowest spot caused by the tongue is particularly crucial for the sound. This can be presented as *the vowel diagram*, where the effect of the highest position of the tongue has been illustrated.

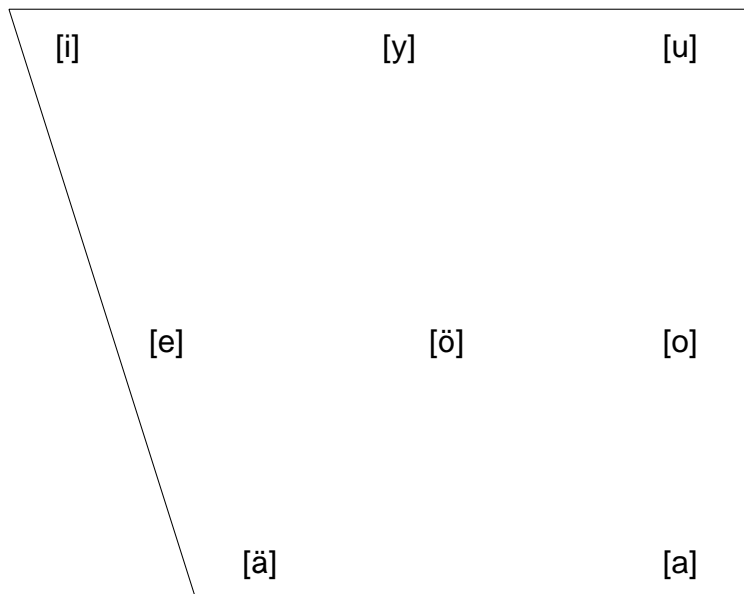


Figure 3: Vowel diagram representing the highest position of the tongue in different vowels. The simplified picture represents the mouth of a persons looking at left.

The *roundness of the lips* has the following effect on the terminology: A phoneme can be called labial if the lips are round, and illabial otherwise. For instance [i] (in) and [y] (news) in Finnish differ from each other mostly based on the roundness of the lips.

Nasality is related to whether the soft palate is down or up. When it is down (i.e., open), the air flows to the nasal cavity, and a nasal sound is produced. When the soft palate is up, an oral sound is produced.

The vowels in Finnish are [a],[e],[i],[o],[u],[y],[ä] ja [ö]. In addition, there are *diphthongs* that consist of two different, successive vowels that occur in the same syllable, for example in the word bay. Especially English language also contains *triphthongs* that contain three vowels instead of two, for example in fire and tower

Consonants For consonants it is typical that the air flow is somehow hindered in the mouth. The consonants can be classified based on:

- the way of pronunciation
- timbre
- the place of pronunciation

The way of pronunciation describes how free the air flow is in pronouncing the consonant. The consonants having free air flow are called *resonants* and those which don't have free flow are called *obstruents*.

Resonants can be further divided into:

- *semi-vowel* (aka *glides*). These are like vowels but the tongue or the lips form a narrower passage than in the pronunciation of vowels. In Finnish, [j] (yes) and [v] (via), and in English, [w] (water), are semi-vowels.
- *nasals*. The air flows through the nose only. In Finnish [n], [m], [ng] (doing).
- *liquid consonants*. The air flows from the mouth in a different manner than in vowels. This class is further divided into *laterals* where the air flows over the edges of the tongue ([l]) and *tremulants* where the air flow is discontinuous ([r]).

The grouping of obstruents can also be refined:

- *stops* (aka *plosives*). In these obstruents the air flow is fully closed. In Finnish [p], [t], [k]. Also [b] (bank), [d] (duck), [g] (gate) can be considered as phonemes in Finnish although not everyone is using them; by the way, these are the voiced versions of [p] (aparatus), [t] (it), [k] (tackle).
- *fricatives*. The air flow is partially suppressed: [s] (sip), [h] (hit), and [f] (fit).

In addition, in Finnish all phonemes excluding [d], [g] and [f] can be doubled, for instance words *muta* (mud), *mutta* (but), *muuta* (else), *mutaa* (of mud) ja *muuttaa* (change,move) are all different words. The [ng]-phoneme is always long unless the following phoneme is a consonant (*kenkä* (shoe)). Moreover, [v] and [h] are seldom long except in “end-doublings” (*homevvaurio*).

In conclusion, the classification of phonemes in Finnish according to the way of pronunciation is as follows (long phonemes excluded):

- vowels: [a],[e],[i],[o],[u],[y],[ä],[ö]
- consonants
 - resonants
 - * semi-vowels: [j],[v]
 - * nasals: [n],[m],[ng]
 - * laterals: [l]
 - * tremulant: [r]
 - obstruents
 - * fricatives: [h],[s] (also [f])

* stops: [p],[t],[k] ([b],[d],[g])

Timbre characterises whether the consonant is voiced or non-voiced. All the consonants except [p],[t],[k],[h] ja [s] (and [f]) are voiced in Finnish.

The place of pronunciation tells where in the vocal tract is the most important narrow passage. For instance [p] and [t] are both unvoiced stops but [p] is placed in the lips and [t] behind the teeth. The different locations are:

bilabial between the lips

labiodental between the lower lip and the upper teeth

dental between the teeth

alveolar between the alveolar ridge and the tongue

palate-alveolar between the frontal palate and the tongue

palate between the palate and the tongue

velar between the soft palate and the tongue

uvular between the tip of the soft palate (uvula) and the tongue

pharyngeal between the back of pharynx and the tongue

Based on these properties all the consonants can be classified. For instance, [m] is a voiced bilabial nasal and [k] is an unvoiced palatal stop.

Phonology

The general abstract class of speech sounds is called *phoneme* whereas one single spoken realization of a phoneme is called a *phone*. Thus, all phones are in principle different. The phonemes belonging to the same class in some particular language, having still clear phonetical difference, are called *allophones*. For instance in Finnish all vowels can be pronounced in nasal fashion without changing the meaning of words, meaning that these are allophones in Finnish language. In other words, nasality is a phonetic feature in Finnish but for example in French, it is a phonological feature.

Joint Articulation

Although we hear phonemes as discrete objects, the successive phonemes are actually merging into each other. This is due to the relatively slow change in the posture of the speech organs, during which the waveform is smoothly changing its form. Moreover, it is rather common that when the phoneme has been pronounced 'clearly enough' (in order to the listener to understand), the organs start moving into the posture demanded by the next phoneme. The allophone used in a particular utterance usually depends on the surrounding phonemes, especially on the next phoneme.

Prosody

Prosody stands for the 'long-term' properties of speech, which contain mainly quantity, weight and intonation. Instead of a phoneme, the smallest unit in prosody is usually a *syllable*. There is no universal definition for syllable, but it can be defined in one language at a time. In Finnish language the main-rule is that the border of the syllable is located in front of every CV (consonant, vowel) group. (e.g. pu-heen-kä-sit-te-ly). Acoustically, the syllable borders can be roughly set in the positions where the acoustic wave has minima in power. In linguistics, a syllable is often a more useful unit than single phonemes.

Quantity means the length of the speech sounds. In some languages (e.g. Spanish) the length does not change the meaning of words. Meanwhile, in Finnish, the quantity is a distinguishing feature both in vowels (muta, muuta) and in consonants (muta, mutta). The quantity of a sound depends on several factors, such as 'the natural length' of the sound, the properties of the surrounding sounds, the position of the sound in the syllable and the weight of the sound.

The *weight*, as it says, means the weighting of some phoneme, usually using more power or altering the pitch. Thus, the weight is in practice created by altering the air pressure in the lungs and by adjusting the tension of the vocal folds. The term 'weight' may refer to either syllable weight or to word weight, although in Finnish the syllable weight is always on the first syllable.

Intonation refers to the changes of pitch during a longer period of time, for instance during a sentence. Pitch can alter the meanings of words in some languages (e.g. Chinese) but it is also applied to express punctuations in other languages, too. In English the pitch raises towards the end of an interrogative clause, whereas in Finnish the pitch in an interrogative clause is constantly somewhat higher than in declarative sentences. Intonation is mainly implemented by controlling the tension of the vocal folds.