Contents

Phonetics and Phonology: General Overview

The Phoneme: Its Nature and Use

The Phoneme and the Allophone

Articulatory Classification of Consonants

Articulatory Classification of Vowels

Intonation: elements and functions

O’Connor’s System for English Intonation: principles, units and structure

Stress: its types and phonetic correlates

Human Vocal Tract

Transcription: definition, types, principles and systems

The International Phonetic Alphabet: purpose, history and development

Speech Acoustics: definition, waveforms and spectrograms.
Phonetics and Phonology: General Overview

John Laver defines *phonetics* as “the scientific study of speech”.

Phonetics has four main aspects: the *articulatory aspect*, which deals with the organs of speech and their use in producing speech sounds; the *acoustic aspect*, which is the study of the physical transmission of speech sounds from the speaker to the listener, the *auditory aspect*, which is the study of the reception and perception of speech sounds by the listener [Wikipedia], and the *phonological aspect*, or *phonology*, which deals with the roles of speech sounds in constituting and differentiating language forms. The latter is often considered a separate field of studies.

The function of phonology is to relate the phonetic events of speech to grammatical units operating at the morphological, lexical, syntactic and semantic levels of language. Therefore, phonetics and phonology are linked and impossible without each other.

The distinction between phonetics and phonology goes back to Ferdinand de Saussure’s distinction between *speech* and *language* (*langue* /lɑ̃ɡ/ - the rules of the language as a system and *parole* /pəˈroʊl/ - the concrete realizations of *langue*). Indeed, Nikolai Trubetzkoy defines phonetics as “the study of sound pertaining to the act of speech” and phonology as “the study of sound pertaining to the system of language”.

The core part of phonology is the theory of the phoneme - a minimal sound unit capable of distinguishing language forms. The word was already used by Ferdinand de Saussure, but it was Jan Baudouin de Courtenay who first employed it in the sense we’re used to. Lev Shcherba gave a comprehensive definition of the phoneme, and the theory has been developing ever since.

Phonetics as an interdisciplinary study employs a vast array of methods borrowed from various fields such as physics (acoustics), psychology and physiology. A considerable part of phonetic research is based on experiments.
Phonetics has many practical applications, varying from teaching pronunciation of foreign languages, treating speech disorders and developing eloquence to automatic speech recognition and speech synthesis, which is very important for human - machine interactions.

**The Phoneme: Its Nature and Use**

A phoneme is the central element of phonology. There are almost as many definitions of the *phoneme* as there have been phoneticians and phonologists.

*Daniel Jones*, a famous British phonetician, became the first linguist in the western world to use the term phoneme in its current sense. Before giving his definition of the *phoneme*, it is a good idea to explain some key points of his approach.

According to him, several distinct sounds in a language have to be considered as if they were one for the orthographic, grammatical and semantic purposes. Thus, although the English k’s in *ki:p* (keep), *ko:l* (call) and *ku:l* (cool) are distinct sounds, it is necessary for practical linguistics purposes to treat them as if they were one and the same.

In fact, these elements of language commonly termed *sounds* are in reality small families of sounds. To such families the term *phoneme* is also applied. Each family consists of an important sound and of other related sounds which, so to speak, “represent” it in particular sequences or under particular conditions of length, stress or intonation.

The most important member of the phoneme may be termed as the “principal member of the phoneme”. This is the most common variant of a phoneme, the variant that is most likely to be pronounced in isolation. The other variants of the same phoneme may be called *sub-phonemic variants* or *allophones.*
Allophones can depend on the neighboring sounds (the phonetic context) and on the phonetic position (for example, in a stressed syllable).

To give a full definition of “phoneme” it is necessary to mention the term **phonetic context**. It means that a sound as the part of a particular sequence is influenced by its position in that sequence and by the neighbouring sounds, which are referred to as its phonetic context. *In the same phonetic context* means ‘when surrounded by the same sounds so being in the same conditions’.

Finally, the definition of the **phoneme** by Daniel Jones reads as follows:

A **phoneme** is *a family of sounds in a given language which are related in character and used in such a way that no one member ever occurs in a word in the same phonetic context as any other members*.

By the way, sometimes a phoneme consists of only one member or of sounds which are so nearly alike that they may be regarded as indistinguishable. For example, English consonant phoneme **f** comprises only one member. Harold Palmer has suggested the term **monophone** for such a phoneme.\(^1\)

An individual consonant or individual vowel in the phonology of a given language can occur in different locations (different syllable structures or different environmental contexts). For any given consonant or vowel, the range of possible locations is constrained by the phonology of the language. This range of the permitted locations constitutes the **phonological distribution**.\(^2\) For example, in English the phoneme **[h]** is found only in the beginning of the word, and the phoneme **[ŋ]** – only in the final position.

A crucial feature of the members of a family, or, allophones, is that they, if used one instead of another, cannot change the meaning of a word. On the contrary, the phonemes can change the meaning, like in English words 'sit' and

---

1 D.Jones. The Phoneme: Its Nature And Use  
2 Laver. Principles of Phonetics
'sat'. These words can be called **minimal pairs**. This function - to distinguish words and morphemes - is regarded as the main function in English tradition. According to Russian phonologists, the main function of the phoneme is to constitute higher-level units such as morphemes /ˈmɔː(r)fiːm/ and words.

### The Phoneme and the Allophone

(information from the previous passage)

Phonemes are usually transcribed using symbols enclosed in slant brackets while allophones - in square brackets.

It is important to note that:

- A **phoneme** can be considered as psychological equivalent of a sound
- An **allophone** represents a variation of a phoneme associated with structural and environmental context

The nature of allophone is clearly seen from examples. Let us look at 4 British English words with a phoneme /p/: 

peat /pit/ => [pʰit], where the [pʰ] is aspirated, but where the lips are not rounded, because the lip position for the following vowel is unrounded;  

pot /pɒt/ => [pʰwɒt], where the [pʰ] is aspirated and lip-rounded;  

spot /spot/ => [spʰɒt], where the [pʰ] is lip-rounded but unaspirated;  

port /pot/ => [pʰwɒt], where the [pʰ] is aspirated and lip-rounded (a degree of lip-rounding is slightly more extreme than for pot, because the lip position for the following vowel has a smaller, more rounded opening in this case).

---

3 Laver. Principles of Phonetics
Thus, in these words, the pronunciations of /p/ show a strong degree of phonetic similarity (they all involve the lips closing completely, with a simultaneous absence of voicing). However, they also show a variety of phonetic differences from each other (differences of aspiration versus non-aspiration and differences in the presence and degree of lip-rounding).

**Articulatory Classification of Consonants**

As opposed to vowels, consonants involve a narrowing or 'stricture' at an identifiable place in the vocal tract. To classify a consonant a phonetician should answer four questions as those:

1) “Where does the obstruction occur or what **place of articulation** it is?”

Phonetic description of place of articulation concentrates on a section or 'slice' through the midline of the vocal tract, the mid-sagittal plane as it is known. The moving articulator is called the **active articulator**, and the stationary part of the vocal tract that it moves towards is called the **passive articulator**.

Although the vocal tract is a continuum, there are several contrastive areas by passive articulators:

- The upper lip (so the articulatory zone is labial)
- The upper teeth, either on the edge of the teeth or inner surface (dental)
- The alveolar ridge, the gum line just behind the teeth (alveolar)
- The back of the alveolar ridge (post-alveolar)
- The hard palate on the roof of the mouth (palatal)
- The soft palate further back on the roof of the mouth (velar)
- The uvula hanging down at the entrance to the throat (uvular)
- The throat itself, a.k.a. the pharynx (pharyngeal)
- The epiglottis at the entrance to the windpipe, above the voice box (epiglottal)

The regions are not strictly separated. For instance, in some sounds in many languages, the surface of the tongue contacts a relatively large area from the back of the upper teeth to the alveolar ridge, so the resulting sounds will be called **denti-alveolar**. The terms like pre-velar⁴ or post-velar⁵ may be used to specify more precisely where an articulation takes place.

Concerning an active articulator the following areas are known to be contrastive:

- The lower lip (labial)
- Various parts of the front of the tongue (coronal):
  - The tip of the tongue (apical)
  - The upper front surface of the tongue just behind the tip, called the blade of the tongue (laminal)
- The surface of the tongue under the tip (subapical)
- The body of the tongue (dorsal)
- The root of the tongue and the throat (pharyngeal)

Passive and active articulators can “interact” with each other in different ways⁶ - the position of active articulators can be neutral - for example, when the moving articulator is a lower lip and articulatory target region is an upper lip. The resulting sound will be bilabial like [b].

Yet a number of languages use the articulations where the active articulator is displaced from its neutral position. The resulting sounds are, for example, labiodental consonants when the lower lip produces an obstruction with the upper teeth. In some languages of India and China there are also retroflex sounds when

---

⁴ intermediate between palatal and velar  
⁵ between velar and uvular  
⁶ this passage is my interpretation of the table from The Sounds of the World's Languages (Ladefoged, Maddieson), p.15 + one sentence from Handbook of the IPA
the tip of the tongue is curled back from its normal position to a point behind the alveolar ridge (they can be described as sub-apical).

2) “What type of obstruction is involved or what manner of articulation it is? There are several ways or manners in which articulation can be accomplished. The articulators can completely close off the oral cavity for an instant or a relatively long period, they may narrow the space considerably or the articulators may simply modify the shape of the vocal tract by approaching each other.

If the airflow is blocked somewhere in the vocal tract the resulting sound will be called **occlusive**. There are two situations possible (not including different airstream mechanisms involved): 1) occlusion of the vocal tract stops all airflow, oral and nasal, then the resulting sound will be called **plosive** or **stop**, 2) occlusion of the vocal tract shifts the airflow to the nasal tract (so the velum - the soft part of the palate at the back of the mouth - is lowered). Consonants where this happens are called **nasals**.

If the articulators form a considerable narrowing, the airflow is turbulent, and this creates sound of a hissing kind known in phonetics as frication. Such a sound is called a **fricative**. If even less narrowing is made in the vocal tract, an **approximant** will result, in which the airflow is not turbulent and no frication is audible.

**Laterals** are sounds where air escapes not in the midline of the vocal tract but at the side.

**Trills** are sounds in which the air is repeatedly interrupted by an articulator vibrating in an airstream.

A very short contact, similar in duration to one cycle of the vibration of a trill, is called a **tap**.
3) “Do the vocal folds vibrate or not?” There we have a binary opposition depending on the state of the glottis: if the vocal cords are producing acoustic energy by vibrating as air passes between them, the resulting sounds will be called **voiced**, if they are not, the resulting sounds will be called **voiceless**. Voicing distinctions are actually more fine grained than implied by this two-way distinction but the nature of this opposition is closely related to the nature of language.

4) “What airstream mechanism is involved in a production of this sound?”

While some languages rely exclusively on air from the lungs for sound production, many languages additionally use one or both of two other 'airstream mechanisms' to produce some of their consonants. The first of these two mechanisms is called ‘glottalic’. It involves closing the glottis, and squeezing or expanding the air trapped between the glottis and a consonant stricture further forward in the vocal tract. There are two most common situations possible:

a) If the air is squeezed, and therefore flows outwards abruptly when a closure further forward is released, or briefly but continuously through a fricative stricture - the sound is known as an **ejective**

b) If instead the air between the glottis and a closure further forward is expanded, reducing its pressure, air flow into the mouth abruptly at the release of the forward closure, the resulting sound will be called **implosive**.

'Velaric' airstream sounds, usually known as **clicks**, again involve creating an enclosed cavity in which the pressure of air can be changed, but this time the back closure is made not with the glottis but with the back of the tongue against the soft
palate, such that air is sucked into the mouth when the closure further forward is released.  

**Articulatory Classification of Vowels**

Vowels are sounds which occur at syllable centres and involve a less extreme narrowing of the vocal tract than consonants. Unlike consonants, they cannot easily be described in terms of a 'place of articulation'. Instead, they are classified in terms of an abstract 'vowel space', which is represented by the four-sided figure known as the 'Vowel Quadrilateral'. This space bears a relation, though not an exact one, to the position of the tongue in vowel production.

The reference points of the chart are **cardinal vowels** that represent the highest and the lowest positions of the tongue combined with the most front and back positions. The first vowel /i/ is rather like the vowel in English ‘heed’ or French ‘si’ (‘if’), the body of the tongue is displaced forwards and upwards in the mouth, towards the hard palate. Since the tongue is near the roof of the mouth this vowel is described as 'close', and since the highest point of the tongue is at the front of the area where vowel articulations are possible, it is described as ‘front’. Conversely, for the vowel labelled [a], which is rather like the vowel of Standard Southern British or General American English ‘palm’, the tongue body is displaced downwards and backwards, narrowing the pharynx. The most extreme version of this vowel is taken as a second fixed reference point. The space between the tongue and the roof of the mouth is as large as possible, so this vowel is described as 'open', and the tongue is near the back of the mouth, so it is described as 'back'.

---

7 The 'tut-tut' or 'tsk-tsk ' sound, used by many English speakers as an indication of disapproval, is produced in this way, but only in isolation and not as part of ordinary words. Some other languages use clicks as consonants
If the tongue body is raised as close as possible at the back of the mouth, just short of producing a velar consonant, and (as is common in languages) the lips are simultaneously rounded and protruded, the close back vowel [u] results, which is similar to the vowel of French ‘vous’ ('you'). And if a vowel is produced in which the highest point of the tongue is at the front of the mouth and the mouth is as open as possible, the result is [a]. This is rather like the quality of the vowel in ‘cat’ in contemporary Standard Southern British English. These two extreme vowels may also be regarded as fixed references.

This chart should be divided further: the vowels can be not only low and high, but also mid-low (/ɛ/, /ɔ/) and mid-high (/e/, /o/). There are now four defined vowel heights. The eight reference vowels are known as the 'primary cardinal vowels'. The description of the primary cardinal vowels outlined above differs slightly from that of the English phonetician Daniel Jones who first defined them, but is in accord with a widespread conception of them today. One can notice that [i e ɛ a ɔ] are unrounded, and [ɔ o u] have increasing rounding, this reflects a relationship commonly found in languages between vowel height, backness, and rounding. Lip activity is, however, independent of tongue position, and many languages exploit this in their vowel systems.

To reflect this, eight 'secondary cardinal vowels' are therefore defined which differ only in lip position from their primary counterparts.

In addition to the cardinal vowels already discussed, which lie on the outside edge of the quadrilateral, there are symbols for mid-central vowels (a typical example is schwa), and for vowels at a number of intermediate locations. If detailed phonetic description is required, most vowels in a language have to be placed in relation to a reference vowel. However, not all the vowel existing in the world’s languages can be fully characterised by these three parameters - height, backness, and lip rounding.

---

8 a short version from Handbook of the IPA
French is the example of a language which has nasalized vowels. The production of these vowels involves the lowering of the velum and the opening of the passage to the nasal cavity which results in the special quality of these vowels.

There is one more distinctive feature of vowels - voicing. In most languages vowels are all voiced, but there exist some languages with devoiced vowels, for example, Japanese.

Vowels can also vary in duration (for example, Finnish distinguishes long and short vowels) and tenseness (for example, English tense and lax vowels).

All the vowels described are referred to as monophthongs, because they show a stable medial phase of relatively unchanging quality. A different type of vowels, whose medial phase shows an audible change of quality, is a diphthong. A diphthong is a vowel which combines 2 vowel elements - a nucleus and a glide - in one syllable, so that the tongue moves from one position to another. A much rarer type is a triphthong where an audible change of quality in the medial phase moves first towards one target and then undergoes a change of target in mid course.

All in all, the larger the vowel system is - the more complicated features it will employ for distinguishing vowels. In the smallest systems the vowels differ only in height, whereas the big ones can use a variety of features.

**Intonation: elements and functions**

Intonation is a suprasegmental feature which is superimposed on every pronounced sequence of speech segments. Despite its crucial role in forming an utterance, it is a phenomenon which raises discussions about its nature, its functions and its units.

---

9 definitions from Laver. Principles of Phonetics

10 Anya Mamuchina
First of all, it should be mentioned that in Russian tradition the intonation is defined by five parameters: pitch (or melody), intensity, tempo, pauses, and timbre /ˈtæmbə/. In English tradition intonation is associated with pitch only.

The acoustic correlate of pitch is the fundamental frequency, that is the first frequency in the spectrum of the speech signal. In terms of articulation pitch is the frequency of the vocal cords’ vibration.¹¹

The melody of a speaker's voice on any given occasion is not a matter of the absolute pitch values, in other words, pitch is relative. The perception of pitch depends on the speaker's expected pitch range (for example, the lowest pitches in the voice of a small child may be considerably higher than the highest pitches in the voice of a large man) and on the following and preceding pitch values in an utterance.

Pitch movements occur within the space called pitch range. There are several types of it. One is the organic range of the speaker's pitch, which is the maximum range of which the speaker's voice is physically capable. The second kind of pitch range is the speaker's current paralinguistic range that is used for momentary signalling particular attitudinal information (such as, in any given utterance, surprise, anger, sorrow, impatience, concern etc.). The third type is the linguistic range, which is the range within which the phonologically relevant pitch of the speaker's voice habitually varies in paralinguistically unmarked, attitudinally neutral conversation.

The fourth and for our purposes the most important type of pitch-range to be distinguished is the phonological pitch-span, which is the local range within which the speaker organizes relative values of pitch for prosodic purposes within the whole or part of a particular utterance.

The pitch range is characterized by two parameters. The first is the baseline, which is the series of pitch-values that is perceived as forming the floor of the current pitch-span. The second is the plateau, or topline, which is the series of

¹¹Anya Mamuchina
pitch-values that is perceived as forming the ceiling of the current pitch-span. The interval between the baseline and the topline defines the *width of the pitch-span*. One of the characteristic pitch-phenomena in many languages is the downward slope of both the baseline and the topline, with a progressive narrowing of the pitch-span within the linguistic range. This effect is usually called *declination*. The opposite phenomenon - the upward movement of the baseline and topline - is referred to as *inclination*.

Normally pitch represents a sequence of tone rises and falls. Monotonous pronunciation without pitch movements is highly unnatural and may even be a symptom of mental disease. However, *not all the movements are relevant*. If we draw the intonation contour of a given utterance, we will see relatively large movements - these changes are perceived by the speaker and form the melody, whereas relatively small movements are imperceptible and are referred to as *microperturbation*.

There are two ways to analyse a pitch contour of a given utterance, first - to consider the whole contour (like Delattre did), second - to divide it into smaller units (like O'Connor and Arnold, who described the structure of the contour as a combination of the head, nucleus and tail.

If, following the Russian tradition, we take the broad meaning of intonation, we should also discuss intensity, pauses, tempo and timbre.

The *intensity*, whose changes are caused by the increase or decrease of respiratory activity, is perceived as *loudness*. However, the perception of loudness depends on the fundamental frequency (that is why decibel scale, based on this relation, is used for measuring the intensity). Like pitch, the intensity decreases towards the end of the utterance. This process is referred to as *dynamic declination*.

---

12 Laver. Principles of Phonetics - simplified
13 Anna Mamuchina, ‘microperturbation’ from Laver
Pauses can be divided into perceived pause, and an acoustic pause or a silent pause: listeners tend to miss some acoustic pauses at unexpected locations.

A further distinction has to be made between filled pauses and unfilled pauses. Filled pauses usually occur in spontaneous speech; e.g. hesitation phenomena caused by planning problems or as discourse markers in dialogues. Typically, fillers are either /ə/, /ɛ/ or /ɛːm/. In contrast to filled pauses, unfilled pauses usually occur at grammatically motivated locations and usually consist of silence or breathing. Of course, all combinations of breathing, fillers and silence can be found.

Definitions of the tempo of speech need to distinguish between the speed with which an individual utterance is produced, versus the rate at which an overall multi-utterance speaking-turn is performed. Articulation rate describes the tempo of articulating an utterance, excluding any silent pauses, but including non-linguistic speech material such as filled pauses and prolongations of syllables. Speaking rate refers to the overall tempo of performance not only of all the utterances in the particular speaking turn, including any filled pauses and prolongations of syllables within the utterances, but also the durations of any silent pauses between the utterances making up the speaking-turn. The length of a particular syllable or a word (compared with the overall rate) can also convey prosodic information. For example, the last syllable in the utterance is usually lengthened.

Timbre is a parameter the least studied. Not all the researches consider it a means of intonation. It corresponds to the spectrum; however, it is still not entirely clear which features found in the spectra are relevant for the intonation. The timbre is often described impressionistically as, for example, wiry voice quality.

The major functions of intonation in speech are the following:
1) divides the continuous speech into units such as phrases or smaller ones -
tone groups or syntagmas in Russian tradition. The main means for this are
pauses.
2) shows not only the borders between the units, but also the type of connection
between them.
3) forms and distinguishes different types of utterances such as questions or
declaratives. The crucial component here is melody. The intonation serves as
a means to mark a word or a unit, to give it more prominence in order to
show its importance. The intensity and duration serve for this purpose.
4) expresses emotions, though not all the researchers agree that this is a
linguistic function and belongs here. Timbre is believed to convey most
information about emotions.¹⁴

O’Connor’s System for English Intonation: principles, units and
structure

Intonation is a suprasegmental feature which is superimposed on every
pronounced sequence of speech segments.¹⁵

Many linguists and phoneticians described English Intonation applying their
specific approaches and introducing new terms.

Thus, **Henry Sweet** suggested symbols for the tone movements: rise, fall,
rise-fall and fall-rise. **Harold Palmer** introduced terms “head” and “tale” and
**Roger Kingdon** added a new term “pre-head”.¹⁶

The most popular description of English Intonation ‘absorbing’ these systems
is the system by O’Connor and Arnold. Their book “Intonation of Colloquial

---

¹⁴ Anna Mamuchina  
¹⁵ Anna Mamuchina  
¹⁶ from the notes of Intonation course
English” was published in 1961 and was reprinted several times. This book includes a series of drills for practice and can be very useful for studying and practicing English intonation.

They divided an intonation contour (*tune* as they call it) for four patterns:

1. The *pre-head* consists all the syllables which precede the stressed syllable of the first accented word. It can be *low* or *high*.

2. The *head* begins with the stressed syllable of the first accented word and ends with the syllable immediately preceding the nucleus. It can be *low, high, falling* and *rising*.

3. The *nucleus* - the stressed syllable of (usually) the last accented word, the landmark of the highest importance.

Nuclear tones can be simple, complex (Fall Rise), compound (Rise + Fall + Rise). There are 7 principal nuclear tones: Low Fall, High Fall, Rise-Fall, Low Rise, High Rise, Fall-Rise, Mid-Level.

4. The *tail* - all the syllables following nucleus.

A total number of possible combinations for basic pitch patterns is, as they say, 105, even without considering compound tunes. However, it is not necessary to deal with such a big number of units for two reasons:

- Some of the patterns occur very much more frequently and with a much wider usefulness than others
- Some different patterns have very slight differences in meaning

Therefore, O’Connor and Arnold suggest 10 *tone groups* (unemphatic are realized with Low Pre-Head, emphatic with High Pre-head):

1. The Low Drop (High Head + Low Fall)
2. The High Drop (High Head + High Fall)
3. The Take-Off (Low Head + Low Rise)

4. The Low-Bounce (High Head + Low Rise or High Pre-head + Low Rise)

5. The Switchback (Falling Head + Fall-Rise)

6. The Long Jump (Rising Head + High Fall)

7. The High Bounce (High Head + High Rise)

8. The Jackknife (High Head + Rise-Fall)

9. The High Dive (High Head + High Fall + (Low Accents) + Low Rise)

10. The Terrace (High Head + Mid-level)

Each name of a tone group speak for itself, the point here is to help learners remember some features of each tone group. For example:

The Switchback: you are enjoying a ride on the switchback at a fair. It takes you down: Falling Head. Then up and down and up again: Fall-Rise.

The High Dive: you watch a swimmer make his dive from the high board. He plunges in: High Fall. He disappears for a second and then rises to the surface again: Low Rise).

After that, O’Connor and Arnold ascribe different meanings to various patterns, mentioning the speaker’s attitude to the situation. The authors show the effect of the tone groups in association with five main sentence types: statement, wh-questions, yes-no questions, command, interjection.

Finally, at the end of their book, there is a number of dialogues demonstrating the ten tone groups in action in everyday conversational speech.\(^\text{17}\)

\(^{17}\) some examples and meanings of the contours ca
Stress: its types and phonetic correlates

Where one of two more or less identical syllables is made more prominent than the other by an exaggeration of the value of one or more of the phonetic parameters of pitch, loudness, duration or quality, the more prominent syllable can be said to receive more stress. In this definition, phonetic stress is a gradient phenomenon, and the phonetic realization of any syllable can be said to show a greater or less degree of stress relative to the manifestation of some other syllable. Stress can also be regarded as a phonological property of the syllable, and a distinction can initially be drawn between stressed and unstressed syllables, using two degrees of phonological stress. The placement of phonological stress on a particular syllable within a word is a defining property of that word, and this can be referred to as word-stress or lexical stress [Laver].

As mentioned above, stress is defined by the following phonetic parameters: pitch, loudness, sound duration and vowel quality. This means that the stressed syllable can a) be different in pitch or bear some sort of pitch movement (sometimes in languages like Croatian or Swedish there are more than one pitch movements associated with lexical stress); b) be louder than the rest of the word; c) be longer or have a longer vowel; d) have a more prominently realized vowel than the rest of the word. This type of stress is complemented by the phenomenon of vowel reduction which can be either phonological (replacing one vowel phoneme by another) or phonetical (replacing one vowel allophone by another). Often these parameters are combined.

It is important to distinguish languages with stress associated with a certain pitch movements from tone languages, where every syllable has to have some tone movement.

Different languages employ lexical stress in different ways: some use fixed stress, which means that word stress is located on a certain syllable in the word (e.g. the last), these languages include Polish and French (however, some argue
that the latter has no stress at all). Other languages use **variable stress**. In this case stress can differentiate between the meanings of words comprised of the same phonemes. Such languages include Russian and English.

The main function of word stress is the **culminative** function, which means that is a characteristic property of a word. Basing on the number of stresses in a phrase, one can judge how many words there are in that phrase. Another function of word stress is the **contrastive** function, which is only found in languages with variable stress. This means that there are words in these languages that differ only in stress position. This is found, for example, in Russian and English.

**Human Vocal Tract**

**Human vocal tract** is the cavity in human beings where the speech sounds are produced and then filtered, so this is the system of organs that humans use to produce speech. However, the primary functions of these organs are not speaking but breathing and eating, and the vocal tract combines these functions. *Breathing during speech* is different from the normal breathing process, in the latter case the exhalation and the inhalation have equal durations, whereas in the former case the exhalation is longer.

The airstream used in speech - the source of energy for the speech sounds - is produced in the lowest area of the vocal tract - the **lungs**. The air then goes up through the *bronchi* /ˈbrɒŋkiə/, *trachea* /ˈtrækiə/, *larynx* /ˈlærɪŋks/ and *pharynx* /ˈfærɪŋks/ to the **oral** and **nasal cavities**. In the larynx the airstream meets the vocal folds - two bands of elastic muscle tissue. If the vocal folds come together and form a narrowing, the airstream going through the tensed folds (cords) causes their vibration, the result is phonation. This is how human voice is created. The mechanism described is used in the production of voiced consonants, resonants, and vowels. When the vocal folds are apart and lax, there is no voicing. We can observe this process in voiceless consonants and voiceless vowels (which are a rare
phenomenon but still existent). Vocal folds can also be closed, blocking the flow of air, and then opens suddenly - to produce a glottal stop.

Pharynx, which is above the larynx, is used as a contact area for producing some consonants such as /h/.

Despite the importance of the organs described, the vast variety of sounds existing in languages is produced higher in the vocal tract. There the airstream is affected by the action of several mobile organs - the **active articulators**. They are:

- **lips**, actively involved in the production of rounded vowels or bilabial consonants (/u, p/);
- **tongue** which is involved in the production of all vowels and most consonants (/n/, /e/);
- **uvula**, which is actively involved in some consonants such as the uvula trill /R/, but which also remains passive in others, for example, /q/ or /ʁ/;

The passive role - which means that the organ is involved in the production of a sound but does not move - is also performed by:

- **teeth** which are used for the pronunciation of labiodental, interdental, and dental consonants (/f, ɹ, d Russian/);
- **alveolar ridge** (the gum line just behind the teeth) involved in the production of alveolar consonants (/t/);
- **hard palate**, or the *roof of the mouth*, is a place of articulation for palatal consonants (/j/);
- **soft palate**, or *velum*, is a broad band of muscular tissue in the upper region of the mouth, the consonants produced here are referred to as velar consonants (/k/).

Another part of the vocal tract that should be mentioned here - the **nasal cavity**, which is not really an articulator but an area. The nasal cavity is the main non-oral resonance chamber. Unlike the oral cavity, the form of the nasal cavity can’t be changed. Typical consonants involving the nasal cavity are /n, m/.
All the organs described serve as a system for creating tone and noise and for modifying them by changing the form of the resonators.

**Transcription: definition, types, principles and systems**

Transcription is the systematic representation of language in written form. The source can either be utterances (speech or sign language) or preexisting text in another writing system.¹⁸

The first classificatory division of types of transcription depends on whether the motivation for constructing the transcription is primarily **phonological** or directly **phonetic**. **Phonologically motivated transcriptions** or **systematic transcriptions** include **phonemic** transcription if a connected text represents in terms of phonemes and **allophonic** transcription - in allophones. The main purpose of both phonemic and allophonic transcriptions is to provide their reader with information about phonological abstractions of the utterances concerned. In both cases, the transcriptions reflect generalizations about the typical habits of speakers of the accent in question, and are normally relevant to the speaker.

**Phonetically motivated transcriptions** are usually called **general phonetic transcriptions**, or **impressionistic transcriptions**. They pay no attention to the phonological value of the material transcribed and take into account all the phonetic properties of the speech. So if a transcription is made in circumstances where nothing can be assumed about the phonological system (like on the first stages of fieldwork, or when transcribing disordered speech), it is necessary to include all phonetic details because it is not clear which phonetic properties will turn out to be important.

One-symbol-per-phoneme transcription which is called (again) **phonemic** transcription can be also called a ‘**broad transcription**’. This transcription should be (but not necessarily) enclosed in **slanted lines** or **slashes (//)**. The object of

---

¹⁸ Wikipedia
attention here is the system of phonemic contrasts exploited by the language concerned, and the ways in which these are distributed over phonological and higher-level units of the language.

The transcription of both allophones and phones is conventionally enclosed in square brackets. They are called narrow transcriptions. General phonetic and allophonic transcriptions share the property of making more or less narrowly specific comment on phonetic details of pronunciation. Yet the focus of an allophonic transcription is the differences associated with structural and environmental context.

General phonetic transcription typically uses the stock of symbols available in the alphabet of the IPA. The acronym 'IPA' refers to the International Phonetic Association or International Phonetic Alphabet. This Alphabet consists of several tables: 1) the large table for pulmonic consonants where each column represents a place of articulation and each row reflect a manner of articulation, a voiced consonant is placed on the right; 2) the table for non-pulmonic consonants; 3) the vowel quadrilateral and 4) other symbols and diacritics to help a phonetician reflect the nuances of a sound realisation.

Other Alphabets have existed alongside the IPA, especially during the early period of the Association's existence.

The most known transcription systems in Russia are comparative linguistics transcription that uses some different symbols for some vowels and consonants and a cyrillic one which we know from school. Yet these two types of transcriptions are very hard to use making a narrow transcription.

---

19 the information from the next passage can be added
20 the if you have The IPA chart with you, a - otherwise
21 Valentina Kolesnikova
The International Phonetic Alphabet: purpose, history and development

The Association was founded in Paris in 1886 by a small group of language-teachers under the leadership of Paul Passy /pɔːl pæˈsiː/. Their primary aim was to make an easy to use notation for small children to help them acquire a pronunciation of foreign languages.

In less than 30 years the membership of the Association had grown from an initial 11 members in France to 1751 members in 40 countries.

The first alphabet that was employed and promulgated was a modification of the '1847 Alphabet' of Isaac Pitman and Alexander J. Ellis. Originally, the aim was to make available a set of phonetic symbols which would be given different articulatory values, if necessary, in different languages. The choice of symbols was dictated by the need to keep them as simple as possible, for the benefit of both teachers and schoolchildren. Since a large proportion of the membership was from Western Europe (or was linked in cultural and linguistic terms with Western Europe), it was inevitable that a roman base should be used for the Alphabet. Thus, to begin with, 'c' stood for the [∫] in the English word sheep, but the [∫] of French chat was represented by an 'x'.

In 1888 a number of rules was set. The most important one was to use one sign for the same sound found in several languages. The resulting transcription should have been phonemic.

During the 1890s further changes were made to the Alphabet, mainly to provide a series of diacritics for transcriptions that were 'allophonic'.

Since its creation, the IPA has undergone a number of revisions. After revisions and expansions from the 1890s to the 1940s, the IPA remained primarily unchanged until the Kiel Convention in 1989. A minor revision took place in 1993 with the addition of four letters for mid central vowels and the removal of letters.
for voiceless implosives. The alphabet was last revised in 2005 with the addition of a letter for a labiodental flap.

From its earliest days the Association has tried to make explicit the principles which guide its work. The statement of these principles has been amended and updated from time to time; the current formulation was approved at the 1989 Convention of the Association:

1) The International Phonetic Association has a standard alphabet designed primarily to meet practical linguistic needs, such as:
   a) putting on record the phonetic or phonological structure of languages,
   b) providing learners of foreign languages with phonetic transcriptions to assist them in acquiring the pronunciation,
   c) working out roman orthographies for languages written in other systems
   d) working out orthographies for languages previously unwritten.

2) The IPA is intended to be a set of symbols for representing all the possible sounds of the world's languages. The representation of these sounds uses a set of phonetic categories which describe how each sound is made. (Like [p] = voiceless, bilabial, plosive). The sounds that are represented by the symbols are primarily those that serve to distinguish one word from another in a language.

3) The Association uses mainly roman letters or Greek letters that harmonize with the roman letters.

4) The construction and use of the IPA are guided by the following principles:
   a) two sounds occurring in a given language are employed for distinguishing one word from another, they should wherever possible be represented by two distinct symbols without diacritics.
b) two sounds with a small difference between them should be presented with a diacritic.

5) All phonetic transcriptions should be enclosed in square brackets [ ]. A transcription that notes only phonological contrasts may be enclosed in slanted lines or slashes //.

Speech Acoustics: definition, waveforms and spectrograms.

Acoustics is the branch of physics that deals with mechanical vibrations in gases, liquids and solid matter. A subdivision of it is speech acoustics, which studies speech sounds. Speech sounds, like any other sounds, are rapid fluctuations in air pressure. Speech sounds are produced when the air is made to move by a human’s vocal organs. When a person is speaking, acoustic energy is radiated from the vocal tract. This acoustic disturbance, consisting of pressure fluctuations, causes the listener's eardrum to move rapidly in and out – in when the pressure is positive, out when negative. Thus acoustic energy is transformed into mechanical energy at the eardrum. This mechanical energy, and the information it contains, go through several more transformations before arriving as patterns of neural energy at the listener's brain. The processing of the information in the listener's brain results in the percept of sound.

It is common to consider the air pressure in a speech wave as a function of time. The shape or form of such a function is called a waveform. These waveforms can be plotted on a graph. Sound waves are periodic – they repeat over certain periods of time. The most important parameters of a sound wave are its amplitude – maximum displacement from zero which reflects the highest air pressure involved, and frequency – the rate at which the wave repeats itself. These parameters, along with the speed of sound in the given matter (e.g. air), can be used to calculate other parameters, such as the period – the time for one complete
oscillation of the wave. The simplest waveform is represented by a sine function and is produced by a harmonic oscillation.

Complex waves can be analysed as sums of several simple sine waves. This is done through a mathematical transform called the Fourier transform. All the frequencies of these waves comprise the spectrum of the original signal. Since a speech signal changes over time, it is more informative to calculate a spectral slice – the spectrum of a small portion of the signal at a given time. Spectral slices across the whole signal can be organized into a (dynamic) spectrogram which reflects the changes occurring to the frequencies.

A simple oscillating object does not oscillate at a single frequency: every part of the object oscillates too at a higher frequency – an integer multiple of the original, fundamental frequency. These higher frequencies are called harmonics.

The central part of speech acoustics is the source-filter model of speech production developed mainly by Gunnar Fant and described in his book “The acoustic theory of speech production”. It describes speech production as a two stage process involving the generation of a sound source, with its own spectral structure, which is then shaped, or filtered, by the resonant properties of the vocal tract. There are three general categories of the source for speech sounds: periodic, noisy, and impulsive, although combinations of these sources are often present. Periodic waves are produced by the vocal cords. The harmonics of these waves are either amplified or filtered out depending on the configuration of the vocal tract. The amplified harmonics are called formants. This source is the only one for vowels and sonants, while the noisy and impulsive sources are chiefly used for fricatives and plosives.