LIN 3201

Sounds of Human Language

Manual

By

Ratree Wayland

Program in Linguistics
University of Florida
Gainesville, FL

Introduction

There are approximately 7,000 languages in the world, and the sounds employed by these languages show both similarities and differences. Thus, an interesting question that one might ask is, "What factors affect the sounds a language can or cannot use?" First of all, we are constrained by what we can do with our tongue, our lips and other organs involved in the production of speech sounds. This factor may be referred to as the 'articulatory ease' factor. Secondly, we are constrained by what we can hear or what we can perceptually distinguish. This is the 'auditory distinctiveness' factor. Thus, no language in the world has sounds that are too difficult for native speakers to produce or to perceptually differentiate. To nonnative speakers, however, certain sounds may prove challenging to both produce and perceive. One of the goals of this course is to familiarize students with the various sounds employed in the world's languages. Students will learn how to describe, produce and perceptually distinguish these sounds.

Describing Speech Sounds

Phonetics is concerned with describing speech sounds that occur in the world's languages. Speech sounds can be described in at least two different ways. First we can describe them in terms of how they are made in the vocal tract (articulatory phonetics). As speech sounds leave the mouth, they cause disturbances in the surrounding air (sound waves). Thus, another way that we can describe speech sound is to analyze its acoustic sound wave (acoustic

phonetics). In this course, the emphasis will be on describing sounds in terms of how they are produced by the vocal organs. Additionally, we will refer to speech sounds that we describe using the symbols of the International Phonetic Alphabet.

The International Phonetic Alphabet (IPA)

An IPA symbol is a shorthand way of representing elements of a sound (i.e, how it is made) just as CO_2 or H_2O are ways of representing the combination of one atom of carbon and two atoms of oxygen to form carbondioxide or the combination of two atoms of hydrogen and one atom of oxygen to form water. For example, a speech sound made when the lips are blown apart by a strong build up of air are represented by the IPA symbol [p] or [b] depending on whether or not the vocal folds are simultaneously pulsating. The vocal folds are vibrating or pulsating in [b], but not in [p]. Thus, the IPA symbols are like chemical symbols in that they represent the articulatory components of sounds. Put another way, each IPA symbol is a little bundle or package containing the information on how that speech sound is made.

THE INTERNATIONAL PHONETIC ALPHABET (revised to 1993, updated 1996)

CONSONANTS (PULMONIC)

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	ɱ m			n		ɳ	ɲ	ŋ	N		
Trill	B			r					R		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

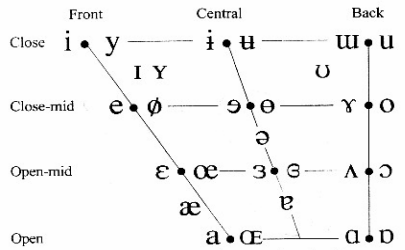
Clicks		Voiced implosives		Ejectives	
ʘ	Bilabial	ɓ	Bilabial	’	Examples:
|	Dental	ɗ	Dental/alveolar	pʼ	Bilabial
!	(Post)alveolar	ʄ	Palatal	tʼ	Dental/alveolar
ǂ	Palatoalveolar	ɠ	Velar	kʼ	Velar
‖	Alveolar lateral	ʛ	Uvular	sʼ	Alveolar fricative

OTHER SYMBOLS

ʍ	Voiceless labial-velar fricative	ɕ ʑ	Alveolo-palatal fricatives
w	Voiced labial-velar approximant	ɺ	Alveolar lateral flap
ɥ	Voiced labial-palatal approximant	ʧ	Simultaneous ʃ and x
ʜ	Voiceless epiglottal fricative		
ʢ	Voiced epiglottal fricative		
ʡ	Epiglottal plosive		

Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary. k͡p t͡s

VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.

SUPRASEGMENTALS

ˈ	Primary stress
ˌ	Secondary stress ˌfoʊnəˈtɪʃən
ː	Long eː
ˑ	Half-long eˑ
˘	Extra-short ĕ
|	Minor (foot) group
‖	Major (intonation) group
.	Syllable break ɹi.ækt
‿	Linking (absence of a break)

TONES AND WORD ACCENTS

LEVEL		CONTOUR	
é͡e or ˥	Extra high	e͡e or ˩˥	Rising
é ˦	High	ê ˥˩	Falling
ē ˧	Mid	e᷄ ˦˥	High rising
è ˨	Low	e᷅ ˩˧	Low rising
ȅ ˩	Extra low	e᷈ ˥˦˩	Rising-falling
↓	Downstep	↗	Global rise
↑	Upstep	↘	Global fall

DIACRITICS

Diacritics may be placed above a symbol with a descender, e.g. ŋ̊

◌̥	Voiceless	n̥ d̥	◌̤	Breathy voiced	b̤ a̤
◌̬	Voiced	s̬ t̬	◌̰	Creaky voiced	b̰ a̰
◌ʰ	Aspirated	tʰ dʰ	◌̼	Linguolabial	t̼ d̼
◌̹	More rounded	ɔ̹	◌ʷ	Labialized	tʷ dʷ
◌̜	Less rounded	ɔ̜	◌ʲ	Palatalized	tʲ dʲ
◌̟	Advanced	u̟	◌ˠ	Velarized	tˠ dˠ
◌̠	Retracted	e̠	◌ˤ	Pharyngealized	tˤ dˤ
◌̈	Centralized	ë	◌̴	Velarized or pharyngealized	ɫ
◌̽	Mid-centralized	e̽	◌̝	Raised	e̝ (ɹ̝ = voiced alveolar fricative)
◌̩	Syllabic	n̩	◌̞	Lowered	e̞ (β̞ = voiced bilabial approximant)
◌̯	Non-syllabic	e̯	◌̘	Advanced Tongue Root	e̘
◌˞	Rhoticity	ə˞ a˞	◌̙	Retracted Tongue Root	e̙
◌̃	Nasalized	ẽ	◌ⁿ	Nasal release	dⁿ
◌ˡ	Lateral release	dˡ	◌̚	No audible release	d̚

The IPA Chart

There are 113 symbols in the IPA chart (84 for consonants and 29 for vowels) and many diacritics to further differentiate small differences among vowels and consonants. The consonants are arranged into columns according to the locations in which they are made in the mouth. For example, bilabial consonants are those consonants made at the lips. The rows, on the other hand, represent the different manner in which the consonants are made. For example, plosives refer to sounds made with a sudden release of air built-up behind a constriction (i.e. between the upper and lower lips). In each column, the symbols on the left represent sounds produced without vocal pulses (voiceless sounds) while the ones on the right are produced with vocal pulses (voiced sounds). Thus, **p** as in the English word 'pin' represents a voiceless bilabial plosive sound while **b** as in 'bin' symbolizes a voiced bilabial plosive sound.

Vowel sounds are arranged according to the relative locations of the tongue body in the mouth during their production. For example, **i** as in 'beat' represents a vowel sound produced with a relatively higher tongue position than **I** as in 'bit'. Similarly, **i** represents a vowel produced with a high tongue body position located in a more forward region of the mouth than **u** as in 'boot'. Thus, [i] is called a high front vowel while [u] is a high back vowel. However, since the observed tongue body's positions for vowels do not exactly match with what's shown in the IPA vowel chart, some phoneticians prefer to view the vowel symbols in the IPA chart to represent the description of how the vowels sound relative to one another rather than the absolute description of the positions of the body of the tongue.

Speech anatomy

The source of power for the majority of speech sounds is the outgoing air from the lungs. The respiratory system pushes air out of the lungs, up through the windpipe or the **trachea** and into the **larynx** (Adam's apple). Inside the larynx are two muscular folds called the vocal folds. With the assistance of cartilage and muscles attached to them, the vocal folds can be brought together (adducted) or moved apart (abducted). When the vocal folds are far apart, the air from the lungs is allowed to pass freely into the pharynx and then the mouth. But if the vocal folds are brought close to each other leaving only a small opening (i.e., **the glottis**) between them, then the air stream from the lungs will set them into vibration. Sounds produced with vocal folds vibration are called 'voiced' sounds and those produced with the vocal folds far apart are called 'voiceless' sounds. Thus, English **v** as in 'van' is a voiced consonant while **f** in 'fan' is a voiceless consonant. Try producing a prolonged **v** and **f** to yourself. You should be able feel the vocal fold vibrations (by putting your fingertips against your larynx) during the production of **v** but not **f**.

The air passage above the larynx is called the vocal tract (see Figure 1). The vocal tract can be further divided into the pharynx, the oral tract and the nasal tract. The parts of the vocal tract used to produce speech sounds are called articulators.

The articulators on the lower surface of the vocal tract usually move toward the articulators located on the upper surface (Figure 2) of the vocal tract during speech sound production. They are referred to as **active articulators** and **passive articulators** respectively.

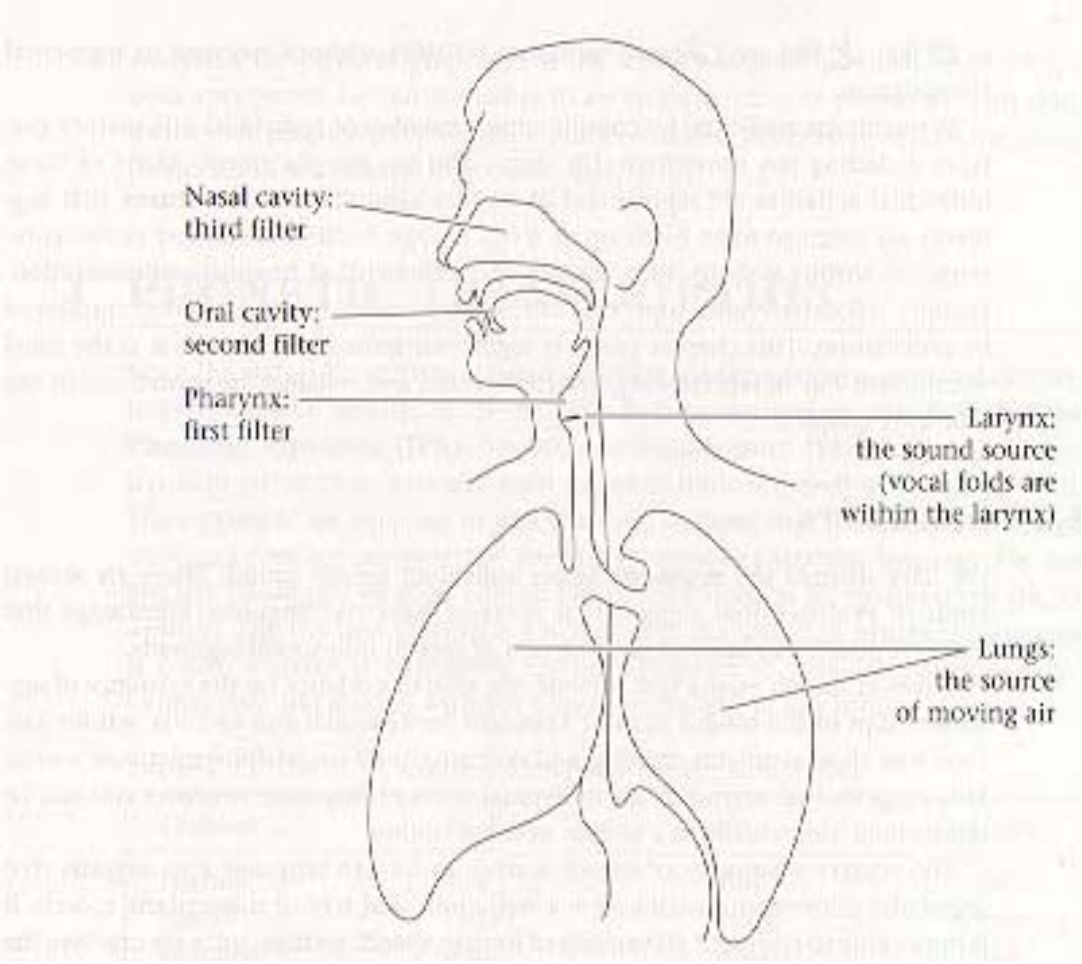


Figure 1. The vocal tract (From Ladefoged, 1993)

The major active articulators on the lower surface of the vocal tract are the lower lip, the tongue and the epiglottis.

As shown in Figure 3, the tongue itself is further divided into subparts namely **the tip, blade, front, center, back** and **root**.

Articulators on the upper surface of vocal tract include the **upper lip**, **upper teeth, alveolar ridge**, **hard palate**, **soft palate** or **velum** and the **uvula**. The velum or the soft palate is a muscular flap that can be raised to press

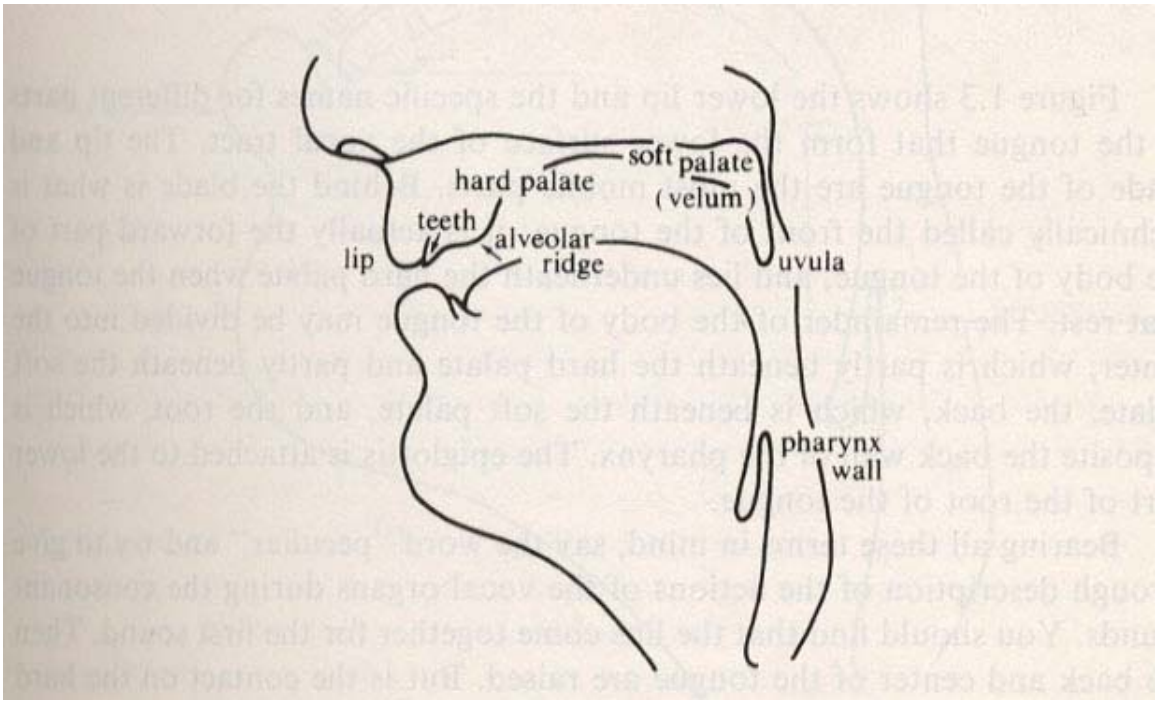


Figure 2. Articulators on the upper surface of the vocal tract

(Ladefoged, 1993)

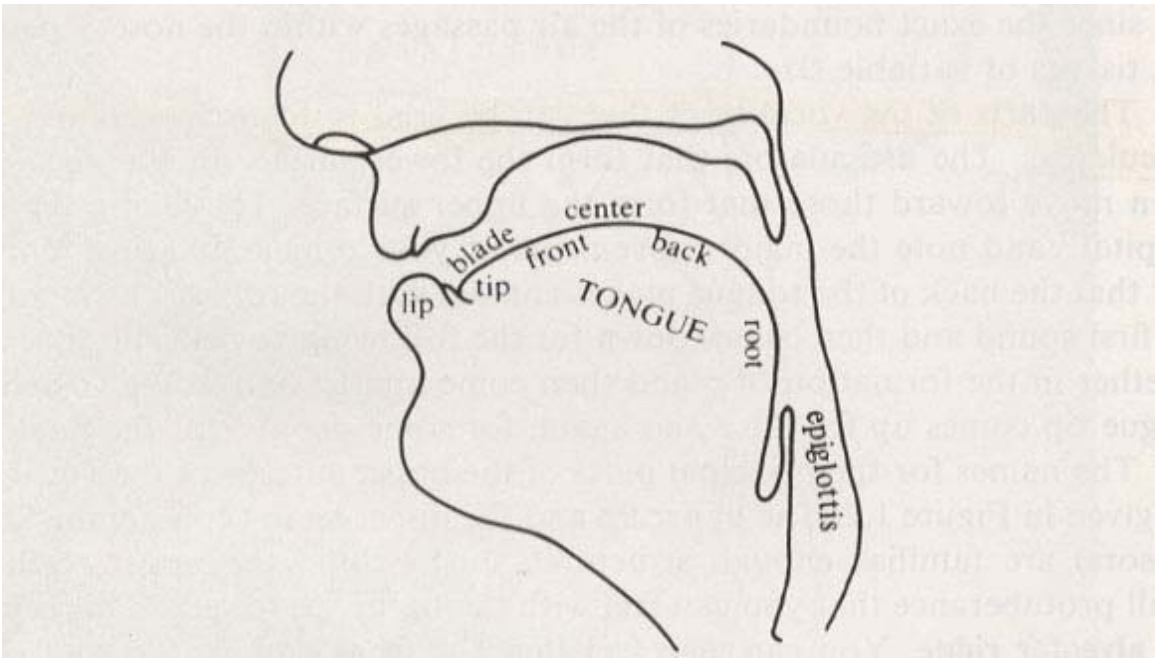


Figure 3. Articulators on the lower surface of the vocal tract

(Ladefoged, 1993)

against the back wall of the pharynx (forming a **velic closure**) to close off the nasal tract preventing air to escape through the nose. The velic closure thus separates between the oral and the nasal tracts. Sounds produced with a raised velum are 'oral' sounds and sounds produced with a lowered velum are referred to as 'nasal' sounds. The small appendage at the end of the velum is known as the **uvula**.

Airstream Mechanisms

Pulmonic Airstream Mechanism

As mentioned earlier, most speech sounds are initiated by the lung air being pushed outwards. The air from the lungs can be pushed outwards by the downward movement of the rib cage and/or an upward movement of the diaphragm, resulting from a contraction of the abdominal muscles. When the body of air from the lungs is moved, it is said that there is a **pulmonic airstream mechanism**. This type of airstream mechanism is the most common source of power for the production of both vowels and consonants in the world's languages. English sounds are typically produced using pulmonic airstream mechanism.

Glottalic Airstream Mechanisms

Different bodies of air are moved to produce speech sounds in some other languages. In other words, other types of airstream mechanisms are used in some languages in the world. For example, some consonants in Hausa, a language of Northern Nigeria are produced by moving the body of air that's trapped inside the vocal tract above the larynx while the glottis (the opening

between the vocal folds) remains closed. The closed glottis can be moved upward to push the air out of the mouth (**egressive**). It can also be moved downward to pull the air into the mouth (**ingressive**). When either of these actions occurs, it is said that there is a **glottalic airstream mechanism**, denoting the central role of the glottis in this process.

An **egressive glottalic** airstream mechanism is used to produce **ejective** stop consonants. Several languages in the world including American Indian languages, African languages and languages spoken in the Caucasus have ejective consonants in their sound systems. Figure 4 illustrates how a voiceless glottalic egressive velar ejective stop [kʼ] is produced. The IPA diacritic indicating an ejective is an apostrophe [ʼ].

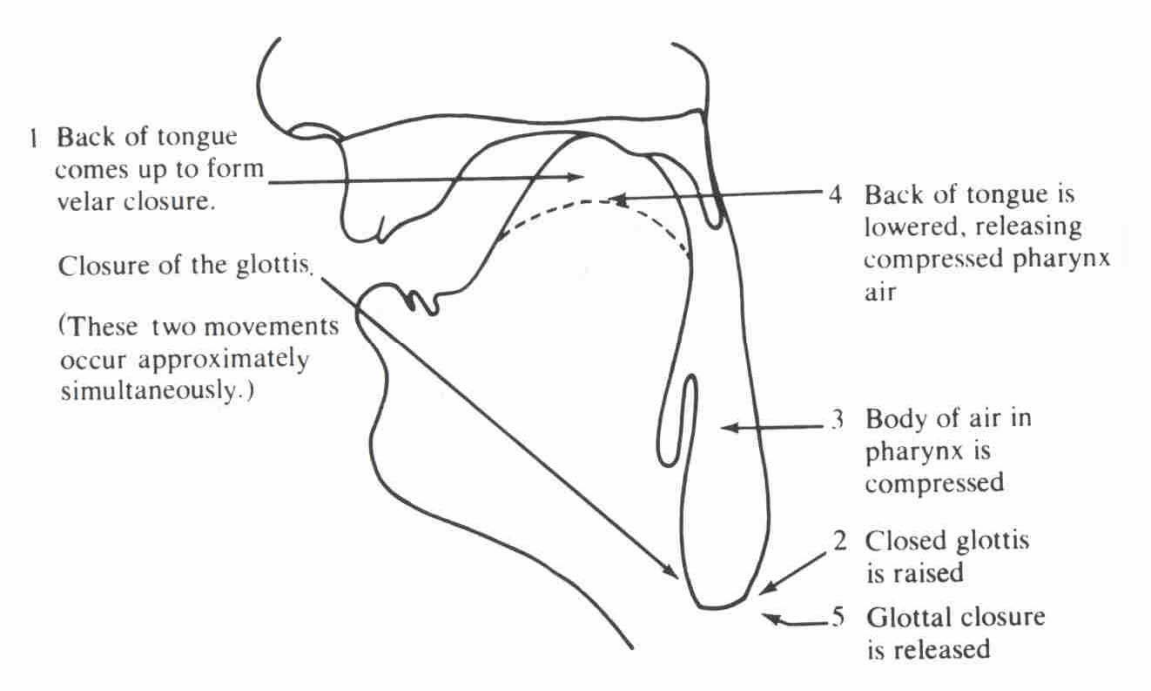


Figure 4. The sequence of events that occurs in a glottalic egressive velar [kʼ]. (Ladefoged, 1993)

An **ingressive glottalic** airstream mechanism is involved in the production of **implosive** stops. Figure 5 shows the sequence of events that occur when a bilabial implosive [ɓ] is produced. Several languages in the world including Vietnamese, Cambodian, Sindhi etc. include implosives in their sound systems.

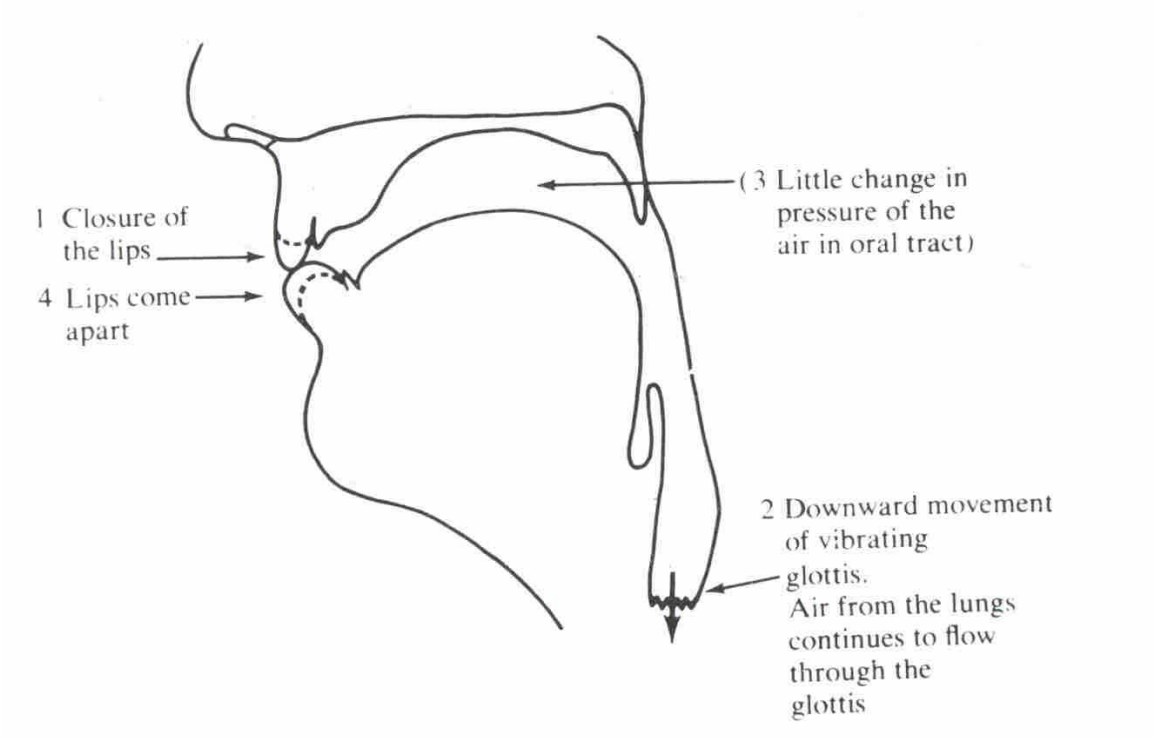


Figure 5. Estimated sequence of events that occurs in a Sindhi bilabial implosive [ɓ]. (Ladefoged, 1993).

Velaric Airstream Mechanism

Another type of airstream mechanism used to produce consonants in a few of the world's languages is the **velaric airstream mechanism**. This is the airstream mechanism used to produce a class of stop consonants called 'clicks.' In the production of clicks, there is a velar closure in which the back

of the tongue is raised to make a complete contact with the soft palate, and the body of air involved in click production is in front of this closure. Figure 6 shows the sequence of events during the production of a dental click. Notice that it's the air pocket trapped between the velar closure and the dental closure that is involved. Specifically, when the tongue body is lowered decreasing the air pressure, the trapped air is rarefied. Thus, when the dental closure is released, higher pressured air from outside the mouth rushes inside producing a click sound quality.

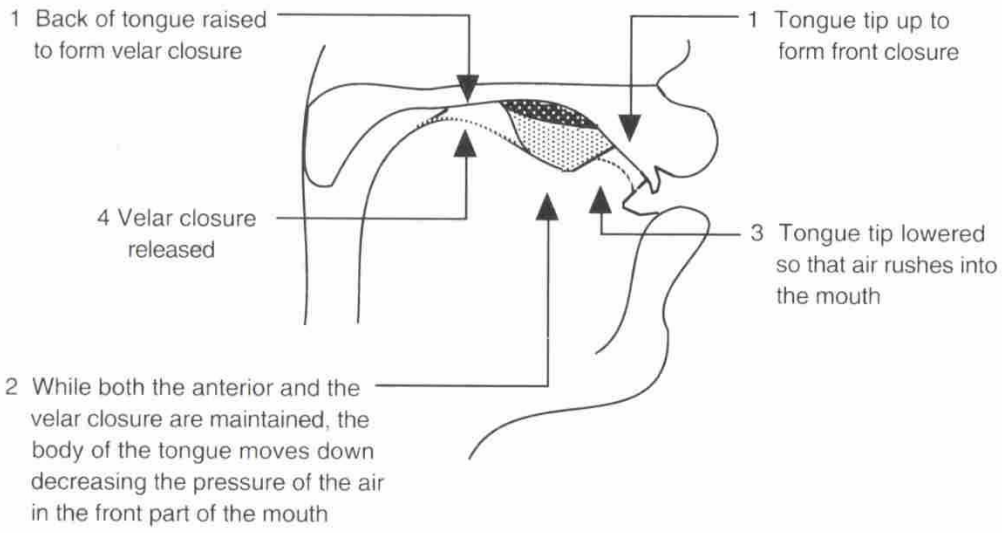


Figure 6.3 *The sequence of events in a dental click. Initially, both the tip and the back of the tongue are raised, enclosing the small pocket of air indicated by the dark shading. When the center of the tongue moves down, the larger, lightly shaded cavity is formed. Then the tip moves down to the position shown by the dashed line, and, a little later, the back of the tongue comes down to the position shown by the dashed line.*

Figure 6. The sequence of events in a dental click (Ladefoged, 1993)

Nasal clicks

Since only the body of air trapped in front of the velar closure is involved in click production, it is possible to produce a velar sound (e.g., velar nasal) using pulmonic airstream mechanism while a click is being made. This is the reason why we can hum continuously while producing a click. The humming (produced by pulmonic airstream mechanism) corresponds to a voiced velar nasal [ŋ]. The co-occurrence of the click and the nasal can be symbolized by using a tie bar [‿] over the click and the nasal symbol. For example, an alveolar click and a velar nasal could be written as [ŋ‿!]. In transcribing click languages, the tie bar is often left off and the co-occurrence between the nasal and the click is often assumed.

Oral clicks

A soft palate can also be raised to block off airflow through the nose to produce oral clicks. Oral clicks can be either voiced or voiceless depending on whether or not pulmonic airstream mechanism is used to keep the vocal folds vibrating during a click. When the back of the tongue is raised to form a velar closure for a click, the articulators are in the position for a voiced velar stop [g]. A voiced alveolar click is therefore a combination of a [g] and a click and could be written as [g‿!]. Similarly, a voiceless alveolar click should be symbolized as [k‿!]. In other words, a phonetic transcription of a click should include a symbol for both the click itself and for other activities (using other airstream mechanisms) accompanying the velar closure, the action central to click production. So we should transcribe a voiced click with

a [g] and the click symbol, a voiceless click with a [k] and a click symbol and a nasalized click with a [ŋ] and a click symbol.

In conclusion, three major airstream mechanisms are used to produce sounds of the world's languages, the most common of which is the (egressive) pulmonic airstream mechanism. This airstream mechanism is involved in the production of **plosive** sounds. **Implosive** sounds are produced by an ingressive glottallic airstream mechanism while **ejectives** are produced with an egressive glottallic airstream mechanism. On the other hand, clicks are produced with a (ingressive) velaric airstream mechanism. In combination with a pulmonic airstream mechanism, voiced and nasalized clicks can be produced. Table 1 below summarizes the description of the main airstream mechanisms.

Airstream	Direction of air flow	Description	Names of stop consonants produced	Examples	Voicing
Pulmonic	egressive	Lung air pushed out under the control of the respiratory muscles.	plosive	p t k b d g	Voiceless or voiced
Glottalic	egressive	Pharynx air compressed by the	ejective	p' t' k'	voiceless

		upward movement of the closed glottis			
Glottalic	ingressive	Pharynx air rarefied by downward movement of the closed glottis; pulmonic egressive airstream may also be involved.	implosives	ɓ ɗ ɠ ʄ	usually voiced by pulmonic egressive airstream, but may also be voiceless
velaric	ingressive	Mouth air rarefied by backward and downward movement of the tongue body	Click	ǀ ǁ ǃ	Voiced, voiceless or nasalized when combined with pulmonic airstream

Table 1 Description of the main airstream mechanisms.

Voicing

The state of the glottis and the action of the vocal folds when air from the lungs passes through the glottis affect the types of **voicing** or **phonation** of the sounds being produced. The common voicing or phonation types found in

the world's languages include voiced, voiceless, breathy or murmur and creaky.

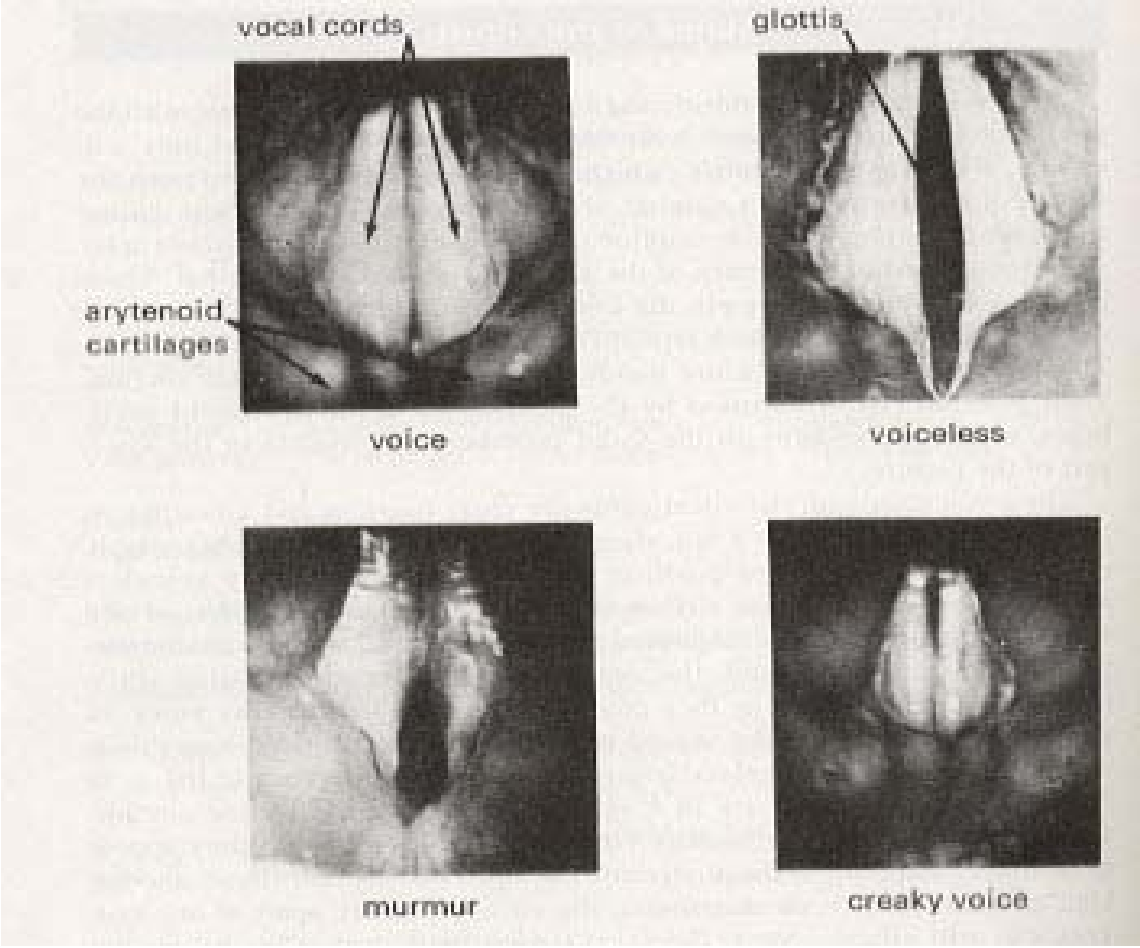


Figure 7. Four states of the glottis. (Ladefoged, 1993)

As shown in Figure 7, in a voiced sound, the vocal folds are closed together and vibrating. In a voiceless sound, the vocal folds are far apart and not vibrating. In a breathy voice sound, the vocal folds are vibrating while remaining apart. This requires a strong airflow through the glottis. One kind of breathy voice is produced with the vocal folds far apart only at one end.

They can still vibrate but at the same time, a lot of air leaks through the glottis. Breathiness is symbolized by [..] two dots underneath the breathy vowel or consonant. In creaky voice, the glottis is closed at one end and open slightly at the other. Thus, the vocal folds can only vibrate at the open end producing a very low-pitch sound. Creakiness is represented by [˷] in the IPA.

Voice Onset Time

Voiceless sounds can be further divided into **voiceless aspirated** and **voiceless unaspirated** sounds depending on whether or not an audible puff of air is produced after the closure is released and before voicing of the following segment (usually a vowel) begins. **Voice onset time** refers to the time interval between a closure release and the beginning of voicing of the following segment. **p** as in '**pin**' is an example of a **voiceless aspirated** sound. When you produce this sound to yourself, and if you put your hand or a piece of paper in front of your mouth, you will feel that a strong puff of air is released as soon as the lips come apart. The vocal folds will then start to vibrate for the following vowel 'i' shortly after. Now, try to produce **p** as in '**spin**'. No puff of air is felt after the release of **p** in this case and the vocal folds begin to vibrate for the following vowel simultaneously or immediately after release of the bilabial closure.

Description of Consonants

In order to produce consonants, the airstream flowing through the vocal tract has to be obstructed in some way. Therefore, consonants are classified according to the manner of the obstruction (manner of

articulation) and where the obstruction is made along the vocal tract (places of articulation). As you may have guessed, being the most flexible articulatory organ, the tongue especially the tip, blade and the back of the tongue together with the lips are the primary articulatory organs used to obstruct the airflow in the vocal tract to produce consonants in most world's languages. **Labial articulation** involves the use of the lips. **Coronal articulation** involves the use of tongue tip or tongue blade. On the other hand, **dorsal articulation** uses the back of the tongue. Below are major places of articulation used to describe consonants in most languages in the world.

Places of Articulation

1. Bilabial. Bilabial consonants are made with the upper and the lower lips approximating each other to obstruct the airflow before it leaves the vocal tract. The first consonants in 'pie', 'by', 'my' are examples of bilabial consonants in English and most of the world's languages.

2. Labiodental. Labiodental consonants are produced with the lower lip raises until it nearly touches the upper teeth resulting in the airflow being almost completely obstructed. The first sounds [f] and [v] in 'five' and 'vie' are examples of such consonants.

3. Dental. In order to form dental consonants, the tip or the blade of the tongue makes contact with the upper front teeth to obstruct airflow. In British English, the first consonants in [θ], [ð] in 'thigh' and 'thy' are produced in this manner. In most dialects of American English, however,

these sounds are produced with the tongue protruding between the upper and the lower front teeth. As such, they may be called 'interdental' consonants to distinguish them from 'dental' consonants.

4. Alveolar. To produce alveolar consonants, the tongue tip or tongue blade touches the alveolar ridge to momentarily obstruct the airflow through the vocal tract. The first consonants [t, d, n, s, z, l] in 'tan', 'Dan', 'nine', 'sign', 'zeal' and 'lie' are examples of alveolar consonants for native speakers of English. However, in some cases, some alveolar consonants are produced as dental consonants. For example, while the [n] in 'ten' may be produced with the tongue tip touching the alveolar ridge, the [n] in 'tenth' is produced with the tongue touching the upper front teeth [tɛn̪θ].

5. Retroflex. To form retroflex consonants, the tongue tip raises and curls such that the underside (of the tongue tip) touches or approaches the back part of the alveolar ridge. Retroflex sounds are not common in English. However, some native speakers may produce r in 'rye', 'ride', 'ire' or 'air' as a retroflex with the tongue raised. Retroflex consonants are common in languages spoken in India including Hindi, Tamil, Malayalam, etc.

6. Palato-alveolar (Post-alveolar). Tongue blade and the back of the alveolar ridge are involved in the articulation of palato-alveolar consonants. Try say the words 'shine', 'shore' or 'shop'. You will notice that during the

articulation of the first consonants [ʃ] in these words, the back of your tongue is close to the back of the alveolar ridge.

7. Palatal. The front of the tongue raises toward the hard palate during the articulation of palatal consonants. In English, the first consonant [j] in 'you' is an example of a palatal consonant i.e., a palatal glide. Other palatals sounds include a palatal nasal [ɲ] as in Spanish 'niño' (baby) or in French 'agneau' (lamb). Voiceless palatal fricative [ç] occurs in German as in 'ich' (I). For some English speakers, [ç] is found as the first sound in 'hue' or 'huge'.

8. Velar. In order to form velar consonants, the back of the tongue raises toward the soft palate. The [k] in 'kite', 'king', the [g] in 'good', 'gang' and the final sound [ŋ] in 'sing' are examples of velar consonants in English. Voiced and voiceless velar fricatives [ɣ] and [x] are found in such languages as Vietnamese as in [ɣà] ' chicken ' and [xá] 'very'. Both of these sounds are said to occur in some dialects of Spanish as in "ojo" [oxo] 'eye' and "pago" [paɣo] 'I pay'. According to Ladefoged (1993), the Spanish [ɣ] is more of an approximant than a fricative and thus should be transcribed with the velar approximant symbol [ɰ].

9. Uvular. To produce uvular sounds, the back of tongue has to raise toward the uvula. Uvular sounds do not occur in English. French, however, has both voiced and voiceless uvular fricatives [ʁ] and [χ] as in [ʁoz] 'rose' and [lɛtχ] 'letter'.

10. Pharyngeal. Pharyngeal sounds are not very common. They are produced by pulling the back of the tongue back toward the back wall of the pharynx. A complete contact between the root of the tongue and the pharynx wall is very difficult to achieve for most people. Thus, pharyngeal stops have not yet been attested in the world's languages. Pharyngeal nasals are also impossible since the approximation between the root of the tongue and the pharynx wall would essentially block the air from flowing through the nose. The only pharyngeal sounds attested so far are voiced and voiceless pharyngeal fricative [ʕ], [ħ] both of which occur in Arabic as in [ʕamm] 'uncle' and ' [ħammaam] 'bath' (Ladefoged, 1993).

11. Glottal. Like pharyngeal sounds, glottal sounds are not very common. They are produced with a constriction deep down in the throat and thus only a few types of consonants can be made. A glottal stop [ʔ] is made with a constriction at the glottis. That is, the vocal folds are brought together to momentarily shut of the airflow. A glottal stop can occur word or syllable initially in a language like Thai as in [ʔaːn] 'saddle'. In English, a glottal stop is an allophone of [t] as in [bʌʔn̩] 'button' in some speakers.
Epiglottal sounds are produced with a constriction between the root of the epiglottis and the lowest part of the pharynx. Pharyngeal fricatives can be produced in this location. According to Ladefoged (1993), some speakers of Arabic produce epiglottal fricatives instead of pharyngeal

fricatives and that some of the languages of the Caucasus contrast between these two sounds.

Manners of Articulation

In the previous section, we learned that consonants can be articulated at different locations along the vocal tract. However, consonant articulation can be achieved in several manners in each of these places of articulation. For example, the articulators involved may make a complete a contact with each other, thus completely blocking of the airflow for a period of time. Or they may approximate each other closely leaving only a narrow gap between them such that the airflow is partially obstructed. Finally, they may just approach each other without making any contact leaving a relatively wider gap between them allowing the airstream to flow through the mouth without being severely obstructed. Below are manners of articulation commonly found in the world's languages.

Stops

Production of stop consonants involves a complete closure of articulators involved. Thus the airflow cannot escape through the mouth at least momentarily. There are two types of stop consonants

Oral Stops

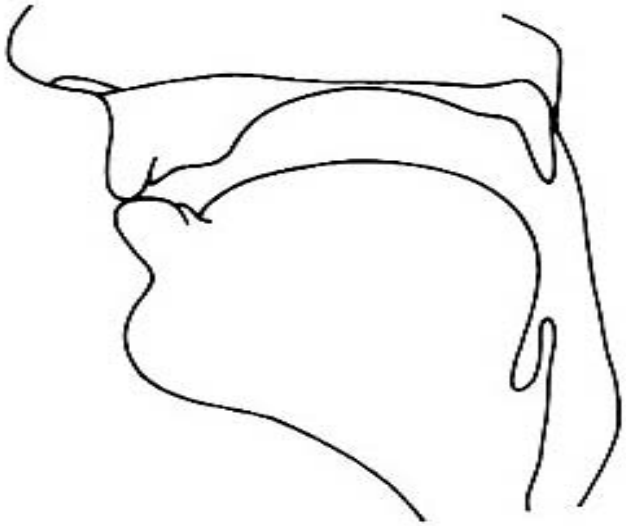


Figure 8. Positions of the vocal organs during an oral stop [b] as in 'buy'. (Ladefoged, 1993)

Figure 8 shows the positions of the articulators involved in the production of the oral stop [b] as in 'buy, book,' etc. As you can see, in addition to the lips being closed to block off the airflow through the mouth, the soft palate is also raised up against the pharynx wall to block off the airflow through the nose (nasal cavity). Thus, the air pressure in the mouth is built up and the oral stop [b] is formed. When the lips come apart, a burst of air will occur. Other examples of oral stops are [d] (alveolar closure) as in 'doll', [g] (velar closure) as in 'guy' etc.

Nasal Stops

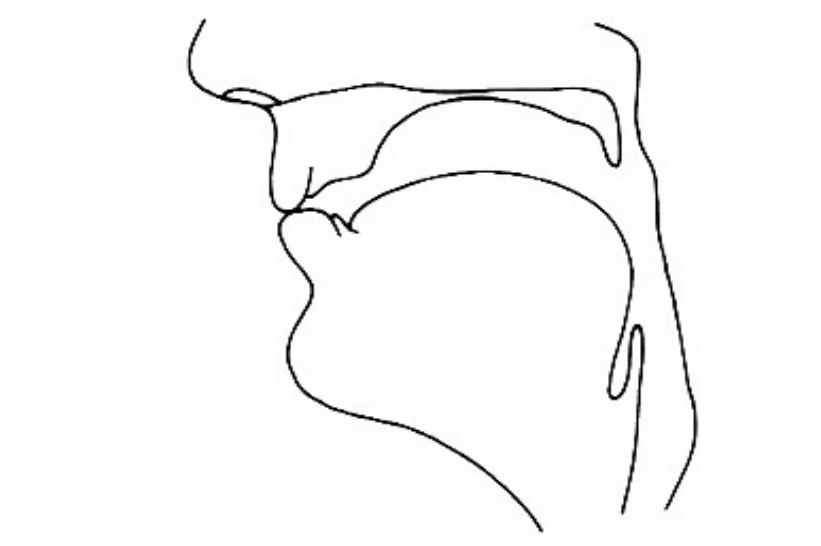


Figure 9 The positions of the vocal organs during the bilabial nasal. (Ladefoged, (1993).

Figure 9 shows the positions of the articulators involved in the production of the bilabial nasal stop [m] as in 'man'. Notice that the only difference between the [m] production and the [b] production is in the position of the soft palate. That is, while the soft palate is raised in the [b] production, it is down in the [m] production. Thus, the pressure built up in the mouth due to the closure of the lips can be release through the nose producing the bilabial nasal stop [m]. Other nasal stops occur in English are [n] (alveolar closure) as in 'nest', [ŋ] (velar closure) as in 'sing'. As shown on the IPA chart, nasal stops produced at other places of articulation also exist in other world's languages.

This includes a palatal nasal and a uvular nasal. One should note, however, that even though both nasal and oral sounds just mentioned are both 'stop' consonants, the term 'stop' alone is usually used to refer to oral stops and the term 'nasal' indicates a nasal stop.

Fricatives

Fricative consonants are produced with two articulators approximating each other closely so that the airflow is partially obstructed and (with a sufficiently high pressure inside the mouth) a turbulent airstream is produced and a 'hissing' sound can be heard. The first consonants in the English words 'sign' [s], 'five' [f], 'thigh' [θ], 'thy' [ð], 'zip' [z], etc. are examples of fricative consonants. Figure 10 illustrates the production of the voiceless palato-alveolar fricative [ʃ] as in 'shy'. The voiced counter part of this fricative, i.e., [ʒ] occurs in a number of English words such as 'vision' , 'visual', 'azure', etc.

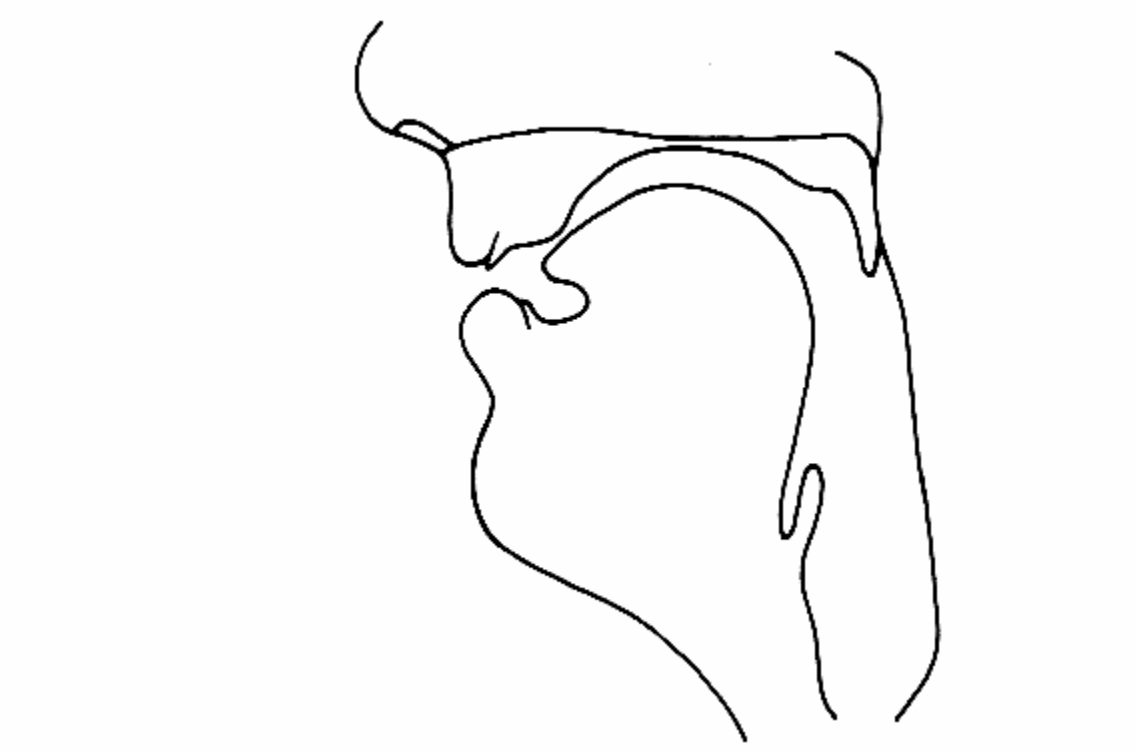


Figure 10. The positions of the vocal organs in the palato-alveolar fricative in "shy."

Approximants

Approximant consonants are produced with one articulator approaches the other but without narrowing the vocal tract to the point that a turbulent airstream is produced. The first consonant in the English word 'yarn' [j] is an example of an approximant consonant produced with the front of the tongue raising toward the palatal region on the roof of the mouth. The first consonants in ' we ' [w] and 'ride' [ɹ] are also examples of approximants in English. [w] is a labio-velar approximant and [ɹ] is an alveolar approximant.

Lateral (approximant)

Lateral approximants are produced in such a way that the airstream at a point along the center of the oral tract is obstructed, but a closure between one or both sides of the tongue and the roof of the mouth is incomplete thus allowing the air to flow freely over the sides of the tongue. The only lateral sound in English is [l] as in 'lie'. When producing this sound, you will notice that the tongue touches the center of the alveolar ridge, while the sides of the tongue are not in contact with the roof of the mouth. It is thus referred to as an alveolar lateral approximant.

Trills

Trill consonants are produced when an articulator is set into vibration by a strong airstream. Spanish 'r' in a syllable initial position as in [roɣo] 'red' is a good example of an alveolar trill, in which the tip of the tongue (raised toward the alveolar ridge) is set into vibration. Alveolar trill [r] also occurs in a careful reading or speaking of Thai as in [rɯa] 'boat or ship'. A uvular trill [ʀ] may occur in some dialects of French instead of a voiced uvular fricative. Bilabial trill [ʙ] occurs in Kele and Titan spoken in Papua New Guinea. For example, the word for 'rat' in Titan is [mʙulei]. (Ladefoged, 1993).

Tap/flap

Unlike trills, a tap or a flap is produced with a single contact between the articulators. In English, the 't' sound in 'better' [bɛɾ] and 'Betty' [bɛɾi] , for example, is produced as an alveolar tap among some speakers. Some

textbooks distinguish a tap from a flap (Ladefoged, 1993). In this case, a tap is produced with the tongue tip moves up to make a contact with the dental or alveolar region of the roof of the mouth and then moves back down to the floor of the mouth along the same path. A flap, on the other hand, is produced with the tongue tip curled up and back (in a retroflex manner) and then strikes the roof of the mouth in the post-alveolar region as it returns to its original position behind the lower front teeth (Ladefoged, 1993). Thus, the distinction between a flap and a tap is that of a place of articulation. Flap is a retroflex while tap is an alveolar or dental.

Affricates

An affricate is a sequence of a stop followed by a homorganic (i.e., produced at the same place of articulation) fricative. The affricates in English are [tʃ] the first sound in 'church' and [dʒ] in 'jog'. The first is voiceless while the second is voiced.

Table 2. Summary of manner of articulation. (Ladefoged, 1993)

Phonetic terms	Description	Example
Nasal (Stops)	Soft palate lowered so that air flows out through the nose; complete closure of two articulators	m, n, ŋ etc.
(Oral) Stops	Soft palate raised, forming a velic closure;	p, t, k, etc.

	complete closure of two articulators.	
Fricative	Narrowing of two articulators so as to produce a turbulence airstream.	f, v, θ, etc.
Approximant	Approximation of two articulators without producing an airstream.	w, j, l, etc.
Trill	An articulator set in vibration by the airstream	r
Tap	Tongue tip hitting the roof of the mouth; a single movement in a trill	
Flap	One articulator striking another in passing	
Lateral	With a central obstruction, so that air passes out at the side.	
Central	Articulated so that air passes out the center.	

Coarticulation

In natural language, individual speech sounds (both vowels and consonants) are strung together to form a larger unit such as a syllable or word. Thus, it is often the case that the articulation of one sound in the same syllable or word will overlap with that of neighboring sounds. This phenomenon is called "coarticulation". In producing the English word 'boy', for example, you will notice that, in anticipation of the mid back rounded vowel [ɔ], the initial [b] is produced with a certain degree of lip rounding. However, your lips remain spread during the [b] in 'beat'. Coarticulation sometimes results in a change of place of articulation of the consonants. The English alveolar stops [t, d], for example, may become dental stops in 'width' [wɪd̪ð] or 'eight' [eɪt̪θ] due to coarticulatory effect with the following interdental fricative [ð, θ].

Phoneticians have identified two types of coarticulation based on the direction of this effect.

1. **Anticipatory coarticulation** occurs when the articulation of a sound changes or becomes more similar to the following sounds. A change from an alveolar [t,d] to a dental [t̪, d̪] is an example of anticipatory articulation.
2. **Perserverative coarticulation** occurs when the articulation of a sound persists to the next sounds. For example, when 'it is' [ɪt ɪz] is elided into 'it's' [ɪts], the voiceless of the [t] persists turning [z] to [s].

Coarticulation is the most common cause of 'assimilation', a process whereby a speech sound is changed into another because of the influence of a neighboring sound. Anticipatory coarticulation is the most common cause of

assimilation. Several types of assimilation have been attested. Place of articulation assimilation is when a sound changes its place of articulation to become more like that of its neighbor as in the case of 'width' and 'eight' above. Voicing assimilation occurs when a sound changes its voicing feature to that of the neighboring segment. Likewise, manner assimilation is when a sound changes its manner to become more like that of its neighbor.

Vowels

Vowels are generally described in terms of three features:

1. Relative degree of tongue height
2. Relative degree of tongue backness
3. Degree of lip rounding

The English vowel **[i]** in 'seat', for example, is produced with the tongue body raised relatively higher than the vowel **[ɪ]** in 'sit'. The highest point of the tongue during the production of **[i]** is also located in a relatively more forward region in the mouth than the vowel **[u]** in 'boot'. Moreover, unlike [u], the lips are spread during the production of **[i]. [i]** is thus referred to as a high front unrounded vowel.

These articulatory-based descriptions of vowels are not, however, satisfactory and people often have difficulty describing the tongue position of certain vowels in their own speech. Unlike consonants, part of the problem in describing vowels is that there is no physical boundaries between one vowel and another. A consonant can be a stop or a fricative, but it can't be something in between the two, but it is possible to make a vowel that is halfway between two other vowels. Thus, vowel sounds form a continuum

rather than a discrete point in the vocal tract. For this reason, some phoneticians prefer to describe vowel sounds in terms of their relative auditory quality to one another rather than to their relative tongue positions in the vocal tract. According to this view, the vowel [i] is 'seat' is called a high front vowel because it has the auditory quality high and front. Similarly, the vowel [u] in 'boot' has an auditory quality that may be called high and relatively back. Of course, the height of the tongue body and its relative location in the vocal tract correspond to the vowel's auditory quality. But the point that some phoneticians try to make is that empirically, the tongue body positions for vowels are not the same as what's shown on the IPA chart, but there auditory qualities (with the support of acoustic measurements) are. According to this view, the terms high, low, front and back describe how vowels sound relatively to one another rather than the absolute descriptions of the positions of the tongue body during vowel productions.

Cardinal vowels

To describe vowels of certain languages or dialects, phoneticians have to rely on their auditory ability. The vowels being described are plotted on vowel charts and the quality of the vowels should be interpretable from their positions on the vowel charts by others who know about the vowel charts. In order for a vowel to be interpretable from such a chart, they have to be plotted with certain reference fixed points and these points should be known to both the people originally plotting the vowels as well as those who are going to interpret them. These reference vowels are called cardinal vowels. They are also called 'extreme' vowels since they are located in the extreme fixed points on a vowel chart. The use of cardinal vowels as

reference fixed points in vowel description was proposed by Daniel Jones, a British phonetician from the late 1800's. Figure 11 below shows the primary cardinal vowels (1-8) and Figure 12 shows the secondary cardinal vowel (9-13).

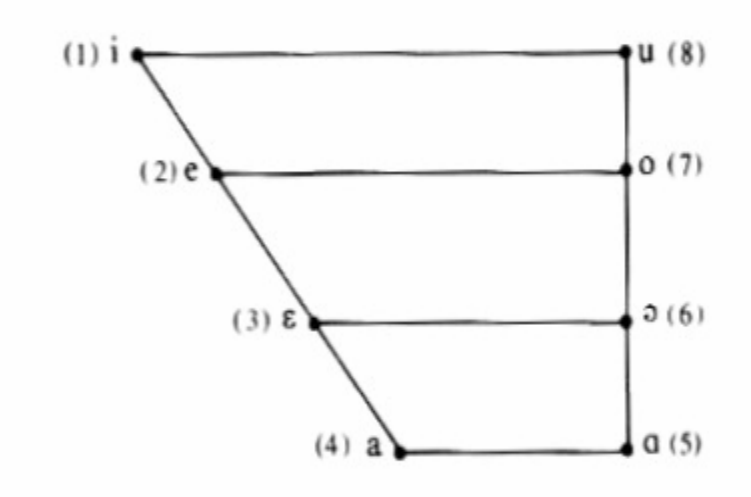


Figure 11.The primary cardinal vowels (Ladefoged, 1993).

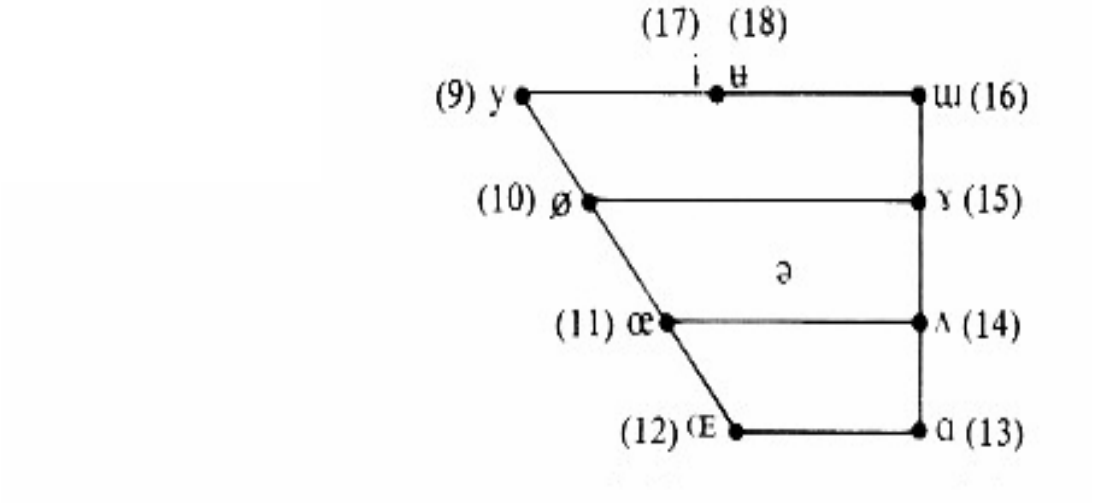


Figure 12. The secondary cardinal vowels (Ladefoged, 1993).

The cardinal vowel system has been used widely by phoneticians to describe the vowel systems of many of the world's languages. However, there are some difficulties associated with its use. First, it was suggested by Daniel Jones himself that the cardinal vowels cannot be learned from written descriptions, but orally from someone who knows them. Thus, unless one has a recording of the cardinal vowels, learning how to produce them from their written description may prove to be less than accurate. Second, according to Daniel Jones, cardinal vowels are auditorily equidistant from each other such that the distance between cardinal vowel 1 to cardinal vowel 2 is the same as the distance between cardinal vowel 8 and 7. Acoustic measurement of the cardinal vowels recorded by Daniel Jones himself showed that the distance between the cardinal vowels (5), (6), (7) and (8) are much closer together than cardinal (1), (2), (3), and (4). Third, it is not certain whether the cardinal vowels are being described in acoustic or articulatory (tongue height) terms. While some phoneticians choose to describe them in terms of tongue height, others prefer to describe them in auditory and acoustic terms.

Advanced tongue root

Degrees of height, backness and lip rounding are sufficient to describe vowel qualities of most of the world's languages. However, some West African languages such as Akan distinguish between two sets of vowels which differ articulatorily in the size of the pharyngeal cavity (Ladefoged, 1993). In one set, the tongue is pulled forward and the larynx is lowered so that the pharyngeal cavity is enlarged. These are called advanced tongue root (or

+ATR) vowels. In the other set, there is no advancement of the tongue nor the lowering of the larynx. These are referred to as –ATR vowels. No vowel pairs in English differ solely in this manner. However, relative degree of tongue advancement may operate in conjunction with variation in vowel height in English. For example, besides tongue body height difference, the tongue root may be more advanced in the high vowel [i] in 'seat' and [u] in 'boot' than in the mid-high vowel [ɪ] in 'sit' or [ʊ] in 'book'. However, the terms **tense** and **lax** have been used to describe the differences between these two sets of vowels such that the high vowels are tense and the mid-high vowels are lax.

Tense and Lax Vowels

English vowels may be divided into two groups referred to as the tense and the lax groups. The distinction is based, not simply on 'articulatory tension', but rather on where these two set of vowels occur in English words. Tense vowels tend to occur in open syllables (i.e., syllables end with a vowel) while lax vowels occur in closed syllables (i.e., syllables end with a consonant), syllables end with a velar nasal [ŋ] or a post-alveolar fricative [ʃ]. A tense vowel tends to be longer and higher than its lax counter part.

Table3. The distribution of tense and lax vowels in stressed syllables in American English (Ladefoged, 1993).

Tense vowels	Lax vowels	Most closed syllables	Open syllables	Syllables closed by [ɹ]	Syllables closed by [ŋ]	Syllables closed by [ʃ]
i		beat	bee	beer		(leash)

	ɪ	bit			sing	wish
eɪ		bait	bay			
	ɛ	bet			length	fresh
	æ	bat		bare	hang	crash
ɑ		hot	pa	bar		slosh
ɔ		bought	saw	bore	long	wash
oʊ		boat	low	(boar)		
	ʊ	good				push
u		boot	boo	tour	hung	
	ʌ, ɜ˞	but		purr		
aɪ		bite	buy	fire		
aʊ		bout	bough	hour		
ɔɪ		void	boy	(coir)		
ju		cute	cue	pure		

Rhotacized Vowels

Some American English vowels are produced with an r-coloring quality. For example, the vowel [ɜ˞] in 'bird', 'herd', 'fur', etc. Just like high-low, front-back, rhotacization described an auditory quality and can be produced in at least two different ways: with the tongue tip raised as in a retroflex consonant or with the tongue tip down but with the tongue body bunched up. These two gestures result in a very similar auditory effect (Ladefoged, 1993). Rhotacization also occurs when vowels are followed by [ɹ] as in 'board', beard', 'tire'. The differences between these vowels and the ones in 'bird', 'fur', and 'fur' is that in the latter, the whole vowels are rhotacized while in

the former, only part of the vowels, usually the later part is rhotacized. Thus, for the former, the original quality of the vowels remains detectable. Try to produce these vowels to see if this is true.

Nasal Vowels

Some languages like French distinguish between nasal and oral vowels. That is, two words in French may differ in meaning depending on whether the vowel is produced with or without nasalization. For example, French distinguishes between [mɛ̃] 'hand' and [mɛ] 'dish' or [bõ] 'good' and [bo] 'good, nice'. Nasalized vowels produced with the soft palate lowered to allow part of the airstream to go through the nasal cavity. In English, vowels are nasalized when followed by a nasal consonant. For example, the vowel [æ] in 'man' is nasalized. In this case, in anticipation of the upcoming nasal **[n]**, the soft palate begins to lower during the vowel. It is important to note that the term nasalization indicates that part of the airstream flows through the nose and part through the mouth. Thus, [m, n, ŋ] are nasals, but they are ***not nasalized*** since the air only flows through the nose during their production.

Semi Vowels/Glides

The labio-velar approximant [w] and the palatal approximant [j] in English are sometimes referred to as 'semi vowels' or 'glides'. Both of these terms capture different aspects of how these two sounds are produced. Similar to vowels (but differ from other consonants), these two sounds are produced with a relatively more open vocal tract in which the articulators merely approximate each other. When they occur in syllable initial position, these

sounds are produced as a rapid glide like movement (thus the name glide) from a high vowel position [i] in the case of [j] and [u] in the case of [w] to the following vowel. In syllable final position, they consist of a glide like movement toward a high vowel position away from the preceding vowel. The term 'semi vowels' also denotes the fact that, despite their vowel-like quality, they never occur at the center (nucleus) of a syllable. Thus, they may be regarded as the 'nonsyllabic' versions of the high vowels. Besides [w] and [j] , there is another semi-vowel, namely the labial-approximant [ɥ] which corresponds to the high front rounded vowel [y]. This sound occurs in French as in [ɥit] 'eight'.

Diphthongs

Diphthong is a vowel that changes its quality during a single syllable. For example, the English vowel [aɪ] in 'five' is a diphthong with the vowel changing its quality from a low front to a high front quality.

Secondary articulation

Every consonant has its own primary articulation. For example, velar articulation is the primary articulation for the velar [k] or [g]. That is, these two consonants are produced with a complete closure between the back of the tongue and the velum region on the roof of the mouth. In some cases, an articulation with a lesser degree of closure is simultaneously added to the primary articulation. This vowel-like add-on articulation is referred to as the 'secondary articulation'.

There are four common types of secondary articulation.

1. **Palatalization** is the addition of a high front tongue position (similar to that of the high front vowel [i], to a primary articulation.) For example, the English [k] in 'key' is palatalized because the contact between the tongue and the roof of the mouth is more front (palatal region) than the [k] in 'car'. Palatalization is not contrastive in English. That is, the meaning of the English word 'key' does not change whether or not [k] is produced with or without palatalization. It may be unnatural and harder to produce and you may sound foreign, but there is no alternative meaning to this odd pronunciation. In English, palatalization often occurs in a high front vowel context and it is symbolized by the raised [^{j}].

   Palatalization is, however, contrastive in Russian and some other Slavic languages. These languages have a series of palatalized consonants that contrast with their nonpalatalized counterparts. For example, in Russian [sok] 'juice' contrasts with [s^{j}ok] 'he lashed', or [domə] 'at home' with [d^{j}omə] 'name' (Ladefoged, 1993).
2. **Velarization** is the addition of a high back tongue position (u-like vowel) to a primary articulation. It is symbolized with [~]. In English, syllable final [l] as in 'feel' is velarized and transcribed as [ɫ]. Similar to palatalization, velarization is not contrastive in English.
3. **Pharyngealization** is the superimposition of a narrowing of the pharynx (cardinal vowel 5 [ɑ]-like). Laryngealization and pharyngealization are very similar and thus not systematically contrastive in most if not all languages in the world. Thus, the same diacritic [~] is used to transcribe laryngealization can be used to transcribe pharyngealization. However, if

a distinction between these two possibilities has to be made, the diacritic [ˠ] is recommended for laryngealization and [ˤ] for pharyngealization.

4. **Labialization** is the addition of lip rounding to a primary articulation. Labialization can co-occur with other types of secondary articulation since it involves the use of the lips rather than the different tongue shapes. In Twi (an Akan language spoken in Ghana) for example, labialization co-occurs with palatalization (Ladefoged, 1993). It is essentially equivalent to adding a front rounded vowel [y] to a primary articulation. As already mentioned, the approximant or semi vowel corresponds to [y] is [ɥ]. Thus, the co-occurrence of palatalization and labialization may be symbolized by a raised [ᶣ]. The name of this language, for example, is transcribed as [tᶣi].

Suparsegmentals/Prosodics

When you listen to a language you have never heard before, chances are that you will hear vowel and consonant sounds that do not exist in your native language. More importantly, you will also notice that it is spoken with a different rhythm from that of your own language. Phonetic characteristics that contribute to differences in rhythm (prosody) often belong to a linguistic domain larger than the individual vowels and consonants. These features are referred to as 'suprasegmental features'. Major suprasegmental features are stress, length, tones and intonations.

Syllables

As mentioned, suprasegmental features belong to a larger linguistic domain than individual vowels and consonants, often this larger domain is a syllable.

The existence and importance of a 'syllable' unit is evident in writing. In Japanese writing systems (Katakana), for example, there is a symbol for each syllable. Moreover, native speakers of a particular language often agree on the number of syllables counted in a word. Children can also be easily taught to identify syllables in their native tongue. Disagreements among speakers do occur and may be attributed to dialectal differences (i.e., American vs. British English) in the way certain words are produced. For example, 'history' may be produced with two syllables in American English, but three syllables among some British English speakers. Similarly, such words as 'troubling' or 'frightening' may be produced as two or three syllables depending on whether they are produced with or without a middle syllabic consonant : [l] in the case of 'troubling' and [n] in 'frightening'.

Despite a fairly good agreement among native listeners, syllables remain one of the most difficult linguistic unit to be defined, both on phonetic and phonological grounds. The notion of a chest pulse as a (articulatory) phonetic correlate of a syllable has been hypothesized. That is, it was proposed that each syllable corresponds to a chest pulse; a contraction of the muscles of the rib cage that pushes air out of the lungs. However, a careful investigation of these groups of muscles revealed that this was not the case. Phonological definition of a syllable has often been circular: a syllable consists of a nucleus (usually a vowel) and a number of preceding and following consonants. But the nucleus of a syllable cannot be identified before the syllable itself. Thus, a syllable nucleus has to be identified on some other grounds.

One proposal has been to use the degree of intensity (sonority) of a segment as the locator of a syllable nucleus: the greater the sonority, the greater

the perceived loudness. Being produced with a more open vocal tract, vowels generally have a relatively greater sonority than consonants. Thus, vowels are more likely to occupy the nucleus position of a syllable. Among vowels, low vowels have higher sonority than high vowels and among consonants approximants and nasal (stops) have greater sonority than oral stops and fricatives. Thus, one possible definition of syllables is that peaks of syllabicity (i.e., nucleus) coincide with peaks of sonority. This would help explain why the number of syllables often coincides with the number of vowels and why people agree on the number of syllables in the majority of words. In 'pocket', 'compensate', etc., there clearly are sonority peaks corresponding to each vowel (two for 'pocket' and three for 'compensate) in these words. A sonority-based theory of syllable cannot, however, explain why a single syllable word such as 'spa' has two sonority peaks; one for [s] and the other for [a]. Suffice to say that while a satisfactory definition of 'syllables' has yet to be discovered, its importance in linguistic analysis is undeniable.

In summary, a syllable may be regarded as the smallest possible unit of speech. Every utterance has to contain at least one syllable. Descriptively, a syllable consists of an **onset** and a **rhyme**. The rhyme consists of the nucleus or the vocalic part of the syllable, which often is a vowel and the coda or final consonants. An acceptable number and types of onset and coda vary from language to language. For example, while German allows such a cluster as [pf] in syllable onset position, English does not. Similarly, while English allows the sequence [sk] in coda position, Thai does not. Differences in syllable sizes and structures in conjunction with other prosodic features to

be discussed below contribute to perceived differences in rhythms in the world's languages.

Stress

Stress is a suprasegmental feature in that it applies not to individual vowels or consonants but to whole syllables. In a given word, the syllable that receives stress is perceptually more prominent or salient than other syllables. Duration, intensity (loudness) and rate of vocal fold vibration (pitch) are the main acoustic correlates of stress. Some languages like English signal stress using all of these three features while others may only use intensity and/or duration.

Word Stress or accent refers to a relatively degree of prominence of one syllable over other syllables in multi-syllabic words. Usually, two to three different degrees of stress are distinguished and syllables are said to have a primary stress, secondary stress or unstressed. It is important to note, however, that these three levels of stress may be identifiable only when words are produced carefully and in isolation. Only primary stress may be perceptible when words are embedded in sentences. Primary and secondary stress are transcribed with the diacritic [ˊ] and [ˋ] respectively. These diacritics are placed in front of the syllable being stressed as in [fə̀nɛ́ɪɪks] 'phonetics'.

Languages vary in their placement of stress. In some languages, stress is always fixed on a certain syllable in a word. For example, stress is usually placed on the first syllable in Czech regardless of how many syllables there

are in the word. On the other hand, stress is always on the penultimate (one before last) syllable in Polish and Swahili. Unlike these languages, stress placement in English is not always predictable and depends, to a certain, extent on the structure of the syllable and the lexical status of the word. Heavy syllables (those that end with a consonant or contain a tense vowel or a diphthong) rather than light syllables (those that end with a vowel or contain a lax or short vowel) tend to attract stress. In bi-syllabic words, nouns are more often stressed on the first syllable while verbs are more often stressed on the second syllables. Exceptions, however, exist for both of these trends.

Sentence Stress (Focus)

Certain syllables in a word (already carrying word stress) may receive additional stress because of the function of that word in the sentence. For example, in the sentence 'The **burgl**ar was caught last night', the syllable 'burg' in 'burglar', the topic of the sentence, receives additional stress besides the word stress it is already carrying.

In some cases, some word stresses may be dropped when said in a sentence. For example, the first syllable in all of these words is stressed when produced in isolation; Susan, younger, sister, wanted, sixty, dollars. But some of these stresses are dropped in the sentence 'Susan's younger sister wanted sixty dollars' such that the first syllable in 'younger, wanted and sixty' are produced without stress.

Emphatic stress occurs when speakers, in order to make clear what they mean, put emphasis on certain words in the sentence. This may result in adding extra stress on an already stressed syllable or in putting stress on

words that are not normally stressed such as function words like *and*, *but*, *both*, *where,* etc. For example, by stressing the function word 'and' in the sentence 'Mary **and** Susan lied to the store manager', the speaker makes clear that Mary was not the only one who lied to the manager, but Susan did as well.

Rhythm

Variations in patterns of stress cause different languages to have different rhythms. Two major types of rhythmic patterns have been proposed to capture rhythms of the world's languages. English and other Germanic languages are said to be a **'stress-timed'** language in which stressed syllables tend to occur at a regular intervals of time throughout the utterance. That is, the rhythm of these languages is determined largely by strong beats falling on the stressed syllables of the words, and the time interval between each beat is said to be more or less equal. For example, in 'Peter is cutting the flowers', the rhythm is determined by the stressed patterned of 'Peter', 'cutting' and 'flower'. 'is' and 'the' are unstressed and compressed. Thus, a typical English utterance will consist of several rhythmic units known as 'feet', each of which is dominated by the beat of the stressed syllable. On the other hand, languages like French and Spanish are '**syllable-timed**' languages in which syllables are said to recur at a regular intervals of time. In these languages, there is no strong pattern of stress or at least not to the extent that unstressed syllables are reduced. The total duration of an utterance depends more on the number of syllables it contains than on the number and position of stressed syllables.

Recently, another type of rhythm has been proposed for languages like Japanese. These are called 'mora-timed' languages. Mora is a temporal unit. Syllables with long vowel or long consonants are said to have more morae than syllables with a short vowel. For example, [kata] has two morae while [ka:ta] or [katta] has three morae. Words with three morae are longer than words with two morae. Thus, the total duration of an utterance in Japanese depends more on the numbers of morae it contains.

Phonetic investigation by means of acoustic measurement suggest that the dichotomy between these two (or three if mora-timed languages is included) rhythmic types may be too simplistic and that rhythmic pattern of a syllable-timed language like French may be affected by stress and the rhythm of an English sentence depends on several factors besides stress. This includes the number of intervening unstressed syllables, the phonetic content of these syllables and more importantly the speaker's intention. Speakers may vary the rhythm of an utterance to maximally transmit his or her message without adhering to the 'equal spacing of stressed syllable rule'. Thus, it is safe to say that an investigation into rhythmic differences among languages in the world is an area that needs further research. A better typology of rhythms has yet to be discovered and rhythmic patterns of the majority of the world's languages have yet to be identified.

Length

Languages may differ in the way they vary the length of the segments (vowels and consonants) in their sound system. In English, variations in segmental length are often contextually induced and thus predictable. For example, you will notice that the vowel [æ] in 'bad' is predictably longer than

the same vowel in 'bat'. This is also true for the vowel [ɛ] in 'peck' and 'peg'. That is, in English, with everything being equal, vowels are longer before voiced than voiceless consonants.

Segmental length is, however, contrastive in some languages. For example, in Thai, long vowels contrast with short vowels. In this language, the length of the vowel cannot be predicted from the consonantal context in which it occurs. Native speakers of Thai have to remember which words are produced with long vowels and which are produced with short vowels. Substituting a long vowel with a short vowel will result in a change in meaning. Length is shown by [ː].

Table 4. Contrast in vowel length in Thai

Short vowel	long vowel
[pan] ' to share'	[paːn] 'equal to, birthmark'
[p^han] ' to wrap around'	[p^haːn] 'a pedestal'
[kan] 'to protect, mutually'	[kaːn] 'business'

Vowel length is also contrastive in Estonian, Finnish, Arabic, Japanese and Danish.

In some languages, consonant length is used contrastively. Unlike vowel length contrast, contrasts between short and long consonants are not so common. Consonant length contrast occurs in Arabic, Italian, Luganda etc. In Italian, for example, 'nonno' [nonnɔ] 'grandfather' contrasts with 'nono' [nonɔ] 'nine'. These long or double consonants (or vowels) are called **geminates**.

Tones

Some languages use voice pitch variation to differentiate word meaning. Pitch is an auditory impression and its physical correlate is the rates of vocal fold vibration such that the faster the vocal folds vibrate the higher the perceived pitch. The tension of the vocal folds and/or the amount of airflow from the lungs are the two main factors affecting the rates of vocal fold vibration, thus the perceived pitch. The higher the vocal folds tension and/or the higher the volume of airflow from the lung, the higher rate of vocal fold vibration, thus the higher the pitch.

Languages that use pitch variations to distinguish word meaning are referred to as **tone languages**. Thai, Chinese, Cantonese, Vietnamese and some African languages are tone languages. The number of pitch variations or tones employed varies from language to language. The simplest tone language has only two possible tones. Bantu languages such as Shona (spoken in Zimbabwe), Zulu or Luganda contrast word meaning using either a high or a low tone. Thai, on the other hand, has five contrastive tones (hi, mid, low, falling, rising). Northern Vietnamese has six contrastive tones while five contrastive tones have been documented for southern Vietnamese. Hmong (spoken in Southeast Asia) is recorded with seven tones.

Tones can be transcribed in many different ways. For languages with only two contrastive tones, a high tone can be marked by an acute accent over the vowel [á] and a low pitch by a grave accent [à]. A combination of an acute accent following by a grave accent can be used to transcribe a falling tone [â], while a grave accent followed by an acute accent can be used to

transcribe a rising tone [ă]. The mid level tone is often left unmarked in many languages.

Different tones may also be transcribed by assigning numerical values to a speaker's pitch range: 5 being the highest and 1 being the lowest. For example, a high level tone may be assigned a value of 55 and a low level tone may be given a 11 value. Similarly, a high rising tone could be numerically transcribed as a 35 tone. Moreover, a graphical representation of pitch levels or **tone letters** can also be used along with numerical pitch level values.

Table 5 Tonal Contrast in Thai

Tone number	Description	Pitch	Tone letter	Example	Gloss
1	Low falling	21	˩	$k^ha:^1$	galangal
2	High falling	51	˥˩	$k^ha:^2$	I, kill
3	High rising	45	˧˥	$k^ha:^3$	trade
4	Low falling rising	215	˩˧˥	$k^ha:^4$	leg
5	mid (level) falling	32	˧˩	$k^ha:^5$	stuck

In some cases, the tone of a word may change due to the influence of a neighboring tone. This phenomenon is called **tone sandhi**. In Mandarin Chinese, for example, the falling-rising tone (Tone3) is changed to a high-rising tone (Tone2) when followed by a falling-rising tone (Tone3). Mostly in African tone languages, there is also a tonal phenomenon known as **declination or down-drift**, in which the pitch level of a given tone is lowered

over a syntactic unit (i.e., a phrase or a sentence). For example, the pitch level of a low tone at the beginning of a phrase will be higher than the pitch level of another low tone occurring at the end of the phrase.

Pitch-Accent

A few of the world's languages are said to have 'pitch accent' including Japanese, Norwegian, Swedish and Serbo-Croatian. Similar to tone languages, these languages use tone contrastively. However, the number of tone is usually limited (e.g., two) and the use of the tonal contrast is limited to a certain syllable in a word. Moreover, this tonal contrast only applies to only some pairs of words in the language. In Swedish, for example, there are two tones or accents and approximately 500 pairs of words are distinguished by this tonal contrast. For example, 'anden' (duck) has a falling tone on the first syllable, while 'anden' (spirit) has a falling tone on both syllables. To native speakers of English, it may sound like 'duck' has stress on the first syllable while both syllables in 'spirit' are stressed. Thus, the pitch accent (or a lack thereof) of the second syllable carries the semantic differences between the two words. This system is constrained since it does not apply to monosyllabic (single-syllable) words.

Differ from Swedish, the variable in the 'pitch accent' system in Japanese is the point at which the pitch falls. For example, in KA(Hi) ki(Lo) ka(Lo) 'Oyster', the first syllable bears the high pitch accent (marked with high tone) and the pitch starts to fall after this syllable. In ka(Lo) KI(Hi) ka(Lo) 'fence' the second syllable is accented and in ka(Lo) ki(Hi) ka(Hi) 'persimmon' is considered unaccented since there is no fall.

Thus, the pitch accent system in Japanese is rather different from that of Swedish. In Swedish, pitch accent is characterized by the contrast between two tones (e.g., high vs. low) in a syllable among multi-syllabic words, while the location where the tone starts to fall is the main feature of the pitch-accent in Japanese. Some researchers argue that Japanese possesses the true characteristics of a pitch-accent language while Swedish does not. Other researchers, on the other hand, argue that both languages may be considered a pitch-accent language and that the system and characteristics of a pitch-accent may vary to some extent from language to language. To date, researchers have yet to agree on a set of defining features of a 'pitch-accent' language.

Intonations

Pitch variation is not used to differentiate word meaning in languages such as English or German, but in these languages, variation in pitch over an utterance (a phrase or a sentence) can indicate whether it is a question or a statement. Variation in pitch across a phrase or a sentence is called intonation. In English, intonation pattern (or pitch movement) of a statement differs from that of a question. Moreover, the intonation pattern of a yes-no question will be different from that of a question that begins with a question word such as what, where, when, why.

The intonation of a statement such as, "I am bored" has a pitch decrease on the last word, thus a **falling intonation**. Questions that require a yes or no answer such as, "Are you hungry?" have a pitch rise on the last word, thus **a rising intonation**. Questions starting with a question word as in, "When are you coming?" are usually produced with the same intonation (i.e., falling) as

that of a statement. Thus, a question can be signaled by a rising intonation of a question word, but not both.

Other intonation patterns exist in English to signal the speaker's attitude or emotional state. Different intonation patterns can be used to signal sarcasm, uncertainty, jealousy, anger, tiredness, definiteness, etc. The use of these intonation patterns is very subtle and is, to some extent, subject to dialectal and individual differences. It is probably fair to say that it is equally difficult for a native speaker of a tone language to learn all the nuances in meaning signaled by various intonation patterns in an intonation language as it is for a native speaker of a non-tone language to learn tones.

The part of a sentence over which a particular intonation pattern extends is called a **tone group**. A short sentence may be uttered with only one tone group, while more than one tone group may be needed to utter a longer sentence. For example, 'She gave him her BOOK.' is said with one tone group and with a rising pitch on the last syllable to signal a question. On the other hand, 'I didn't BUY it ‖ because it was exPENsive' is spoken with two tone groups separated by (‖). The emphasis was on the word 'buy' in the first tone group and on the syllable 'pen' in 'expensive' in the second tone groups. These syllables are called the **tonic syllable**. It is the syllable that carries the major pitch in a tone group. It is important to note that there is no fixed syntactic or grammatical unit that corresponds to a tone group, and the number of tone groups in an utterance is determined by the function of the utterance thus the intention of the speaker. Thus, the sentence, 'I didn't buy it because it was expensive' could be spoken with only one tone group 'I didn't buy it because it was exPENsive' with a rising pitch excursion only on 'PEN'.

It should be noted that intonations also exist in tone languages. The interaction between tone and intonation is a complicated one and researchers have just begun an investigation into this issue.

Phonemic Analysis

What we have discussed thus far is the descriptions of possible speech sounds (vowels and consonants) occurring in the world's languages. Characteristics of speech sounds that apply not to individual segment of vowels and consonants have also been discussed. These are called suprasegmental or prosodic features of speech. However, the task of a linguist does not end at describing individual speech sounds. We are also interested in how these sounds work together as a system. That is we are interested in systematic structures of speech sounds (i.e., the phonological systems) in different languages and how they relate to other domain of linguistics such as morphology and syntax in those languages. Thus, after transcribing presumably all possible sounds or phones (speech sounds not yet analyzed into a system) in a language, a linguist needs to perform the so call 'phonemic' analysis on the phonetic data obtained to uncover the language's sound system. The main goal of this analysis is uncover the number of 'phonemes' and their predictable variants called 'allophones' that exist in the language.

Phoneme and Allophones

Phonemes are abstract structural units of sounds that serve to differentiate word or lexical meaning in a language. Phonemes refer not to individual speech sounds but a class of speech sounds that are judged by a

native speaker to be the same sounds. Each phoneme contains a certain number of its predictable variants called 'allophones.'

Consider the following example in English:

Top [t^hap$^\neg$] stop [stap$^\neg$] little [lɪɾl̩] kitten [kɪʔn̩] spot [spat$^\neg$]

You will notice that the letter 't' are produced as [t^h], [t], [ɾ], [ʔ], [$t^\neg$] depending on where it occurs in these words. It is produced as an aspirated alveolar stop [t^h] in syllable initial position in 'tope', as an unaspirated stop in 'stop', as a tap [ɾ] in 'little', as a glottal stop [ʔ] in 'kitten, and as an unreleased stop [$t^\neg$] in 'spot. If you ask native speakers of English, they would tend to agree these are all variants of the same sound, namely 't'. Thus, we may say that [t^h], [t], [ɾ], [ʔ], [$t^\neg$] are predictable variants or allophones of the same phoneme /t/. The occurrence of each allophone is predictable from the context (its position in words or syllables) in which it occurs. Below are the allophonic (or phonological) rules governing the distribution of the various variants of the /t/ phoneme in this data set. Note that phones or allophones are in square brackets and phonemes are in slant brackets.

Allophonic Rules 1. /t/ is aspirated in syllable initial position.

(phonological) 2. /t/ becomes a flap [ɾ] when occurred as syllable initial of an unstressed syllable following a stressed syllable.

3. /t/ becomes a glottal stop [ʔ] when it precedes a syllabic [n] in an unstressed syllable following a stressed syllable.

4. /t/ is unreleased when occurred as syllable or word final.

Please note that while **[t^h]** is an allophone of the **/t/** phoneme in English, it belongs to a separate phoneme i.e., the /t^h/ phoneme in other languages e.g., Thai. In Hindi, aspirated /k^h/ is also a separate phoneme from /k/.

Thai:	[taː]	'eye, grandfather'
	[t^haː]	'to apply (paint)'
Hindi:	[k^həl]	wicked person
	[kəl]	yesterday, tomorrow
	[kap]	cup
	[kap^h]	phlegm

From this data, you can see that unlike English, both aspirated and unaspirated stops can occur in syllable-initial position in these languages.

Identifying phonemes and allophones: The distribution of speech sounds.
The first step in identifying a phoneme is to examine the distribution of sounds or phones that are phonetically similar in some way with regards to the way they are produced. For example, they all are produced at roughly the same place of articulation with the same manner and voicing, etc. The assumption being that phones that belong to the same phoneme should share some of these phonetic characteristics.

Contrastive Distribution

Two sounds that occur in the same phonetic environment (e.g., word or syllable initial, word or syllable final) and substitute one with the other changes the meaning of the word, they belong to two different phonemes. For example, in Hindi [pʰəl] 'fruit' and [bəl] 'strength', both [pʰ] and [b] occur in the same position, i.e., syllable initial, and substituting [pʰ] with [b] will change the meaning of the word from 'fruit' to 'strength'. They are in 'contrastive' distribution and we can conclude that they are allophones of two different phonemes. A pair of words that differ only in one segment like [pʰ] and [b] are called minimal pairs. Minimal pairs provide crucial evidence for the existence of phonemes. Thus, seek them if you can.

Analogous Distribution

A minimal pair may not always be found for certain pairs of sounds. For example, in Jaqaru (Courtesey of Dr. Hardman)

[ʃa]] hello (man to man) [miʃi] cat

[saja] to stand [nisa] in just a moment

[iʃa] no, not (principal cause)

Notice that there are no minimal pairs for [s] and [ʃ]. However, we can see that they both can occur in comparable (or analogous) environments, e.g. word initial in 'hello' and 'stand', (second) syllable initial in 'no', 'cat and 'in just a moment'. This is sufficient to conclude that these two sounds are allophones of different phonemes.

Complementary Distribution

Two sounds that occur in non-overlapping and predictable sets of context and substituting one for the other does not affect the word meaning, then these two sounds are allophones of the same phoneme. Examine the data from Korean below.

[param]	wind
[irɨm]	name
[pal]	foot
[mal]	horse

Minimal pairs can never be found for [r] and [l] in Korean because these two sounds do not appear in the same positions in words: [r] appears only between two vowels, while [l] occurs elsewhere. They are in 'complementary' distribution. Sounds showing this type of distributional pattern are considered to be allophones of a single phoneme.

Free Variation

In some phonetic contexts more than one pronunciation of a given sound may be possible. Consider:

Leap [lijp] leap [lijpˀ]

keep [kijp] keep [kijpˀ]

Both [p] and [pˀ] occur in the same environment; they can both appear at the end of a word. However, the choice between [p] and [pˀ] does not make a

difference in meaning. They are in free variation. They are allophones of the same phoneme.

Transcriptions

Several types of transcriptions can be distinguished, but the two most common ones are:

1. **Broad transcription** is a transcription that does not show a great deal of phonetic details. It usually amounts to a simple phonemic transcription.
2. **Narrow transcription** is a transcription that shows phonetics details to represent allophonic differences by using a variety of symbols as well as diacritics. In English, these phonetic details may include aspiration of initial stop consonants in stressed syllables, vowel length preceding final voiced stop consonants vs. voiceless stop, unreleased stop in word final position, etc.

Common Phonological Processes

1. **Assimilation:** A sound becomes more like a neighboring sound.

   English

   Alveolar stop assimilation: Alveolar stops assimilate to the place of articulation of a following consonant, e.g. tenth /tɛnθ/ > [t^{h}ɛn̪θ], in this /ɪn θɪs/ > [ɪn̪ θɪs]

Vowel nasalization: Vowels are nasalized when occur immediately before a nasal consonant, e.g. man /mæn/ > [mæ̃n]

Voicing Assimilation: English liquids and glides [l, ɹ, w, j] becomes voiceless when they occur after a voiceless consonant, e.g. play /plej/ > [p^{h}l̥ej]; train /tɹen/ > [t^{h}ɹ̥ejn]; queen /k^{h}win/ > [k^{h}ẇ̥ijn]; putrid /pjutɹɪd/ > [p^{h}j̥utɹɪd]

Palatalization: A sound becomes a palatal sound, e.g. did you eat? /dɪd ju it/ > [dɪdʒa ijt] . The sound [d] has become a palatal sound because of the influence of the following palatal glide. Vowels /i/ and /e/ can also cause this change. The following are common types of palatalization: [t] > [tʃ]; [d] > [dʒ]; [s] > [ʃ]; [k] > [tʃ]; [g] > [dʒ]

2. Dissimilation: A sound becomes less like the neighboring sound.

Manner dissimilation in **Greek**: A fricative becomes a stop when preceded by the fricative [s], e.g. in fast speed, /asθenis/ 'sick or weak' can be pronounced as [astenis], or /sxima/ 'schema' > [skima]. That is, /θ/ becomes a stop [t] so that it will be less similar to the preceding fricative /s/, and /x/ becomes [k] so that it will be less similar to the preceding /s/.

English: fifth /fɪfθ/ > [fɪtθ]

3. Insertion: A segment not present at the phonemic level is added to the phonetic form of a word, e.g. dance /dæns/ > [dænts]; strength /stɹɛŋθ/ > [stɹɛŋkθ].

4. Deletion : A segment present at the phonemic level is deleted/eliminated from the phonetic form of a word. This rule is more frequently applied to unstressed syllable and very common in fast speech,

e.g. /h/ deletion in English

He handed her his hat /hi hændəd hɨ hɪz hæt/ >[hi hændəd ɨ ɪz hæt]

Or (children pronunciation) something /sʌmθɪŋ/ > [sʌmpɪm] ([p] is inserted; [θ] is deleted); or fifths /fɪfθs/ > [fɪfs]; clothes /kloðz/ > [k^{h}l̥oʊz]

5. Strengthening (fortition): A sound becomes stronger

English aspiration rule: Voiceless stop becomes aspirated when they occur at the beginning of a stressed syllable e.g. peak /pik/ > [p^{h}ijk]. Aspiration stops are considered to be stronger sounds than unaspirated stops because the duration of voicelessness is much longer.

6. Weakening: A sound becomes weaker

The flapping rule in English is an example of weakening. [ɾ] is considered to be weaker than [t] or [d] because it is shorter and it obstructs air less.

Morphophonemics

The phonetic realization of a phoneme may vary according to the phonetic context in which it occurs. For example, English [l] is velarized when

occurred in word final position immediately after a vowel. Vowels are longer in final voiced than voiceless consonants. Or vowels are nasalized when occurred in front of a nasal. Pronunciation variation is sometime conditioned by morphological factors including word structures. The study of this phenomenon is called morphophonology. This phenomenon is widely attested in the world's language. A good example from English is in the ways the plural suffix (-s) is pronounced. We know that this suffix can be pronounced as [-s], [-z] or [-əz] depending on the final segment of the base (stem) that it is attached to; [-s] after a stem ending in a voiceless consonant besides fricatives and affricates, [-z] after a stem ending in a voiced segment including vowels, [-əz] after a stem ending in fricatives and affricates. We know that all three of these variants belong to the same morpheme because they denote the same meaning (i.e., plurality) and we know that this is a morphophonemic alternation (and not a phonological alternation) because it occurs only in a certain morphologically defined contexts. For example, it is possible to pronounce an [s] after a voiced segment [l] as in 'else', but the plural suffix [-s] is pronounced as [z] after [l] s in 'mills', 'pills', 'sills'.

As conditioned variants of a phoneme are called allophones, variants of a morpheme are called allomorphs. Thus, in this example [-s], [-z] and [-əz] are all allomorphs of the plural suffix morpheme /-s/.

Further readings

1. Ladefoged, Peter (1993). A Course In Phonetics. (3rd edition). Hartcourt Brace College Publishers.
2. Ladefoged, Peter (2001). Vowels and Consonants: An Introduction to the sounds of languages. Blackwell Publishers.

3. McMahon, April. (2002). An Introduction to English Phonology. Oxford University press.