Pronunciation of American English
Consonants and Semivowels

Paco Gómez

Contents

1 Introduction ..................................... 2

2 Anatomy of speech organs ...................... 3
   2.1 From lungs to larynx .......................... 3
   2.2 The upper vocal tract ....................... 6

3 Consonants and semivowels ..................... 10
   3.1 Phonemes and allophones .................... 10
   3.2 Description of consonants and semivowels ... 12
   3.3 Voicing of consonants ...................... 15

4 The American English consonants .......... 18
   4.1 Stops ........................................ 18
      4.1.1 Bilabial sounds \(\text{p}^{h}\), \(\text{p}\), \(\text{p}\), and \(\text{b}\) ......................... 20
      4.1.2 Alveolar sounds \(\text{t}^{h}\), \(\text{t}\), \(\text{t}\), and \(\text{d}\) ......................... 21
      4.1.3 Velar sounds \(\text{k}^{h}\), \(\text{k}\), \(\text{k}\), and \(\text{g}\) ......................... 23
      4.1.4 Glottal sound \(\text{?}\) ......................... 24
      4.1.5 Glottalized sound \(\text{t}^{\prime}\) ...................... 26
   4.2 Fricative sounds .............................. 27
      4.2.1 Labio-dental sounds \(\text{f}\) and \(\text{v}\) ................. 27
      4.2.2 Dental sounds \(\text{ð}\), \(\text{ð}\) and \(\theta\) ................. 28
      4.2.3 Alveolar sounds \(\text{s}\) and \(\text{z}\) ..................... 29
      4.2.4 Post-alveolar sounds \(\text{s}^\prime\) and \(\text{z}^\prime\) ............. 32
      4.2.5 Glottal sound \(\text{h}\) ......................... 33
   4.3 Affricate consonants ......................... 35
      4.3.1 Consonants \(\text{t}^\prime\) and \(\text{d}^\prime\) .................... 35
   4.4 Nasal consonants ............................. 36
1 Introduction

This article is the natural continuation of the study of American English consonants. All the remarks made at the outset of that article still hold true. Repetition, if not founded on the right pillars, is futile. Teaching pronunciation by mere repetition is methodologically fruitless and conceptually amiss. In no other subject we find such a powerful enemy, our own brain, willing to interfere student’s learning. As students of English as a second language, we all are well equipped with our mother tongue. Our brain is slothful, craves for the familiar, and adamantly rejects the unknown. Yet we all have the same articulatory organs. The difficulty in learning pronunciation lies in psychological factors, not in physiological factors. Before the unexpected, our brain tries to adapt the new sounds to the familiar repertoire of sounds of our mother tongue. Therefore, in order to attain a good pronunciation new articulatory habits must be taught. Those habits have to be instilled by explaining what articulatory organs are involved and how they produce those new, funny sounds. In the words of Peter Roach [Roa09], “pronunciation is an acquired skill.” Once those new sounds have been internalized, then repetition is the key to success, the mother of learning. A grammar rule can be understood instantaneously and, with some extra effort, master it. When learning a new sound, our brain has to familiarize itself with it, study its features, understand it in context, hear it pronounced with different accents; and then, but not before, as perception precedes production, our brain is all set to properly pronounce the new sound. Hence, practice from knowledge, be aware of each gesture, of each sound, of each organ. With time, the new sounds will become old friends, you will forget the technical explanations and just feel whether a produced or heard sound

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.4.1</td>
<td>Stops [m], [n], and [ŋ]</td>
<td>36</td>
</tr>
<tr>
<td>4.5</td>
<td>Approximant Consonants</td>
<td>38</td>
</tr>
<tr>
<td>4.5.1</td>
<td>r-Consonants [ɾ], [ɾ], and [ɻ]</td>
<td>39</td>
</tr>
<tr>
<td>4.5.2</td>
<td>Lateral sounds [l] and [ɭ]</td>
<td>42</td>
</tr>
<tr>
<td>4.6</td>
<td>Semivowels</td>
<td>44</td>
</tr>
<tr>
<td>4.6.1</td>
<td>Palatal semivowel [ʃ]</td>
<td>45</td>
</tr>
<tr>
<td>4.6.2</td>
<td>Velar semivowel [w]</td>
<td>46</td>
</tr>
<tr>
<td>5.1</td>
<td>Other pronunciation problems of Spanish language</td>
<td>47</td>
</tr>
<tr>
<td>5.2</td>
<td>Syllabic consonants</td>
<td>48</td>
</tr>
<tr>
<td>5.3</td>
<td>Classification of consonants</td>
<td>48</td>
</tr>
<tr>
<td>6.1</td>
<td>Diaphragmatic breathing</td>
<td>50</td>
</tr>
<tr>
<td>6.2</td>
<td>Speech articulators</td>
<td>50</td>
</tr>
<tr>
<td>6.3</td>
<td>Tongue shapes</td>
<td>50</td>
</tr>
</tbody>
</table>
is conformable to language’s nature.

Description of consonant sounds implies a more complicated interaction of articulatory organs than in the case of vowel sounds. For that reason, in the next section we will examine the anatomy of speech organs. Building awareness of the organs involved in speech and how they participate in sound production will handsomely contribute to a good pronunciation. After that, we will study the American English consonants in detail. This will include, among others, variants of sounds /p/, /t/, /r/, and glottal stops. Finally, we will look at semivowels /j/ and /w/, which are sounds sharing features of both vowels and consonants. Throughout this article the symbols adopted by the International Association of Phonetics (IPA) will be used to represent sounds; see their official website [Ass] for more details on the IPA alphabet or [Wik]. Definitions will be given in boldface, examples and spellings will be given in italics, and sound transcriptions will go within forward slashes / / for broad or phonemic transcriptions and within square brackets [ ] for narrow or phonetic transcriptions (see Section 3.1).

To illustrate the sounds explained in this article, we have again borrowed the excellent videos from Rachel’s English [Eng12]. The symbols used in those videos are mostly IPA symbols. Unless stated otherwise, all phonetic transcriptions will be given in American English. As some complicated anatomy words will appear in the text, we will provide the reader with the phonetic transcription as appropriate.

Experience has taught us that the articulatory habits of a student’s mother tongue cannot be ignored when teaching a new language. They inevitably interfere in the acquisition of the new articulatory habits. That interference is explained through many causes such as orthographic interference, different consonant and vowel distributions, differences in phonemes and allophones, contrastive features not used in one of the languages, disparate use of voicing, among others. Throughout the description of the consonants in this article we will point out to the main problems that can be found in the pronunciation of native Spanish speakers. Our purpose idea is to illustrate how to correct pronunciation in an effective manner.

2 Anatomy of speech organs

In many languages speech sounds are produced by modifying the airstream coming from the lungs. These modifications may take place in many ways, but two are of utmost importance: voicing, or making the air vibrate through the vocal cords; and disturbances caused by the organs in the upper vocal tract, such as tongue, lips, teeth, etc. We will divide our study of the organs involved in speech production into two steps, the first being the study of the organs from lungs to larynx, the second being the study of the upper vocal tract.

2.1 From lungs to larynx

Air is stored in our lungs. When we want to breath out, our chest is contracted, helped by the diaphragm /ˌdaɪəfræm/, an umbrella-like, powerful muscle, and air is expelled from
the lungs, passing through two short tubes, the bronchi/ˈbronkɪ/; see Figure 1 (it has been taken from Antranik.org). Air comes in two different streams from the lungs and both merges at the bottom of the trachea. The trachea/ˈtrelkiə/, also called the windpipe, is a tube joining the bronchi and the larynx. Inside the trachea the air flows freely as it is a hollow tube.

![Organs of the Respiratory System](image)

Figure 1: Diaphragm, lungs, bronchi, trachea, and larynx.

When the airstream reaches the end of the trachea, it then passes through the larynx. The larynx/ˈlærnks/ is a sort of hollow box made of cartilage designed to contain the vocal cords, also called vocal folds. The front of the larynx is more prominent than the back (even more in the case of males) and is called Adam’s apple. The vocal cords consist of two sheets of muscle tissue, running parallel to each other, placed perpendicularly to the trachea, and spanning from back to front of the larynx; see Figure 2 on the left (Figure taken from Anatomy and Physiology Lab website at Florida Atlantic University). The space between the vocal cords is called glottis/ˈglatɪs/. Three positions can be adopted by the vocal cords: a close position, where the vocal cords are brought together and seals off the larynx cavity (again Figure 2 on the left); an open position, where the vocal cords are completely pulled apart and let the airstream go through with no obstruction (Figure 2 on the right); and

---

1The singular of bronchi/ˈbronkɪ/ is bronchus/ˈbronkəs/, a case of irregular plural. This word is of Latin origin and its plural was preserved when incorporated to English. Many anatomy words share this feature.

2Plural of trachea: tracheae/ˈtrelkiː/.

3Plural of larynx: larynges/loʊˈrɪŋks/.
near-close position, where the vocal cords are almost in contact, but in a loose way (see the video in Figure 3). The close position occurs when swallowing; that prevents foreign bodies to enter into the lungs and cause choking. The open position occurs when breathing normally. The near-close position occurs when speaking and singing. The vocal cords may take up other particular positions, as in whispering or murmuring; see [ODK97], Chapter 2, for a description.

Figure 2: Larynx and positions of the vocal cords.

The vocal cords have freedom to move subject to certain limits. They are attached to two cartilages: posteriorly to the arytenoid/ˈɑːrɪtənɔɪd/ cartilages, and anteriorly to the thyroid/ˈθaɪrɔɪd/ cartilage. The arytenoid cartilages can move so that vocal cords are pulled apart when necessary. In Figure 3 there is a video showing how vocal cords work (stroboscopy of a healthy woman; video taken from Voicedoctor.net). Vocal cords in the video are the pearly white stripes (they are white due to scant blood circulation).

Figure 3: Video showing how vocal cords work.

In the video we can observe several interesting behaviours of the vocal cords. The length of the vocal cords is changed by the action of the larynx cartilages, which affects pitch. The longer the vocal cords, the higher the pitch. Although it cannot be appreciated in the video, the vocal cords can change their tension and therefore their resistance to the airstream.
Those tension changes also modulate the pitch. The greater the tension, the higher the pitch. The vocal cords vibrate controlled by the **vagus**/vegəs/ nerve. Actually, the final pitch is a combination of tension and length adjustments made by the vocal cords.

A particular sound may or may not be produced with the vocal cords in vibration. The quality of the sound changes enormously when vocal cords vibrate as the sound is enriched with harmonics. A sound is said to be **voiced** if it is produced with the vocal cords in vibration; otherwise, it is called **voiceless** or **unvoiced**. All vowels are voiced. In English, **voicing** – the quality of being voiced, also called **phonation** – is of paramount importance. Most English consonants are paired by voicing; later on we will discuss this in depth. For example, voiceless sound /p/ has a voiced counterpart, sound /b/; the same occurs with other voiceless-voiced pair of sounds: /s/ and /z/, /t/ and /d/, or /f/ and /v/.

It is critical to the student of English to be able to tell when a sound is voiced or voiceless. In English, voicing is **contrastive**, that is, different words can only differ by voicing; compare *pet*/pet/-bet/bet/, *sip*/sip/-zip/zip/, or *fan*/fæn/-van/væn/. After having consolidated its perception, production of voicing is the next step. Try to use sounds that can be sustained in time like /s/-/z/ or /ʃ/-/ʒ/ rather than short sounds like /t/-/d/ or /k/-/g/. For example, alternate /s/-/z/-/s/-/z/-/s/-/z/... It should be clear which one is voiced from the sound. If still in doubt, feel your larynx vibrate by placing your hand on your Adam’s apple while pronouncing a voiced sound, or even better by placing your fingers firmly over your ears.

### 2.2 The upper vocal tract

Once the air is expelled from the lungs and passes through the larynx, it reaches the **upper vocal tract**. This term refers to all organs and spaces from larynx to mouth. The upper vocal tract is divided into the **oral tract**, the mouth and throat cavities, and the **nasal**/nɛzl/ **tract**, the nasal cavity. Next, we will describe all articulatory organs in the oral tract; use Figure 4 for reference and study.
1. **The lips.** They take part in sound production in both vowels and consonants. In vowel pronunciation their position determines the *roundness* of the vowel. For example, /u:/ is a rounded vowel, but /æ/ is unrounded. In the case of vowels, lips can adopt three positions: rounded or lips projected forwards, unrounded or neutral lip position, and spread or lips stretched back. For more details on vowel pronunciation, see [Góm12b].

In the case of consonants, lips are involved in sounds such as /p/ or /m/, as in *pet* or *man*, and also in semivowels such as /w/, as in *wood*. Except for a few consonants, the lips are *passive articulatory organs*, that is, they do not move to produce the sounds.

2. **The teeth.** They take part in consonant sounds such as /θ/ or /v/ (as in *thought* or *van*). The upper front teeth are usually involved in the production of sound rather than the bottom front teeth.

3. **The roof of the mouth.** As a whole, it constitutes an important passive articulatory organ. From front to back, it can be divided into the alveolar ridge, the hard palate and the soft palate.

   (a) **The alveolar ridge**/ælˈviːəl rɪdʒ/, also called the *teethridge* /tiːθriːdʒ/. The roof of the mouth does not make a perfect circular arch. The alveolar ridge starts just at the rear of the upper front teeth and slopes gently towards the interior of the mouth. At certain point there is a bulge, from which the actual arch of the roof of the mouth starts. The alveolar ridge is the area comprised between the rear of the upper front teeth and that bulge. Pull gently your tongue from...
the teeth to the bulge to feel the shape of the alveolar ridge. It is also a passive articulatory organ and appears in the production of sound such as /t/ or /z/, as in ten/ten/ or zoo/zoo:/. 

(b) **The hard palate**/pælət/. After the alveolar ridge the hard palate is found, the arch formed by the roof of the mouth. Within the hard palate two regions are distinguished, the **post-alveolar**/pəʊstələr/ region, found immediately after the alveolar ridge, and the **palatal**/pælətal/ region, the rest of the hard palate. The hard palate is a passive articulatory organ, which is used, for example, to make sound /j/, as in yore/jɔr/. Again, feel it with the tip of your tongue.

(c) **The velum**/vɛləm/ (plural, vela/vi:la/) or **soft palate**. The final part of the roof of the mouth is made of soft tissue, unlike the hard palate, which is made of hard bone. Although the tongue has to be bent quite a lot, it is possible to feel the soft palate with the tip. Sounds like /k/ or /g/, as in can/kæn/ or go/gou/, are produced by using the soft palate. However, apart from sound production, the main role played by the soft palate is sealing off the nasal tract. That allows the distinction between **oral sounds** and **nasal sounds**. Oral sounds are produced when the air only passes through the oral cavity. For that to happen, the soft palate has to be raised so that the entrance to the nasal cavity is sealed off. If the soft palate is at neutral position, just hanging, then part of the sound goes through the nasal cavity. That gives sound a special quality, the so-called nasal quality (**nasality**/næzæləti/).

4. **The uvula**/ju:vələ/. The soft palate ends in a conic projection called uvula. It is described here for the sake of completeness, but the uvula is not involved in any English sound.

5. **The tongue**. The tongue is the articulatory organ. Unlike most organs described so far, the tongue is an **active articulatory organ**, that is, it moves to produce the sounds. The tongue is very flexible and moves in quite subtle ways. It can bend backwards, it can raise or lower several parts, it can move backwards or forwards, it can protrude and retract, it can twist, it can form grooves through its medial axis, and it can touch other articulatory organs in a number of ways; see [HP03] and [SL96] for information on models of the tongue and its movements. The tongue is controlled by a complex and large group of muscles, as shown in Figure 5 (taken from Medica Look), which include horizontal, vertical, transverse, and longitudinal muscles.
In order to describe sound production precisely later on, we will distinguish the following parts in the tongue:

(a) The **tip**, or the thin pointed end of the tongue.

(b) The **blade**. Phoneticians Ladefoged and Maddieson [LM96] gave a definition that has been used often because of its simplicity: “When the tongue is at rest in a closed mouth, the tongue blade is the part of the tongue that lies directly under the alveolar ridge.” That will be our definition, too.

(c) The **front** or **body**. This term is confusing because it is not the front it is referring to but the middle part of the tongue; see Figure 6. The front of the tongue is found after the blade. In sounds such as high, front vowel /iː/, it is the front of the tongue that comes near the roof of the mouth.

(d) The **back**. It is found after the front of the tongue and forms its posterior part. It is involved in sounds such /ʊ/ or /ɡ/.

(e) The **root**. The root of the tongue is the part at the far back and bottom of the tongue. It is also part of the front wall of the pharynx. Again, the root is mentioned here for completeness, but it is not associated to any English sound.
Paradoxical enough, although it is known that tongue shape is perceptually relevant to speech sound production, very few books mention it when describing a sound. **Tongue shape** refers to the shape the tongue takes on when producing a sound. For some sounds, for example sibilants (see Section 3.3), tongue shape is crucial to pronunciation. In this article basic tongue shapes are described.

6. **The pharynx**/ˈfærɪŋks/ (plural, pharynges/foʊˈrɪŋɡəz/). The pharynx is the area delimited by the larynx, the root of the tongue, and the soft palate; refer to Figure 4. The root of the tongue can move and change the width of the pharynx.

3 Consonants and semivowels

3.1 Phonemes and allophones

At this point it is in order to make a finer distinction in the concept of sound. The smallest unit of speech that distinguishes one word from another is called **phoneme**/ˈfoʊnɛm/. For example, *pet*/*pet/* and *bet*/*bet/* are different by one sound; the initial sound in *pet* is a *voiceless* bilabial stop /p/ (precise descriptions later on), whereas in *bet* it is a *voiced* bilabial stop /b/. The presence of one sound or the other definitely changes the meaning of the word. However, in English sound /p/ can be pronounced in at least two ways, one as an aspirated sound [pʰ], which is strong and dry, occurring when the sound is on the syllable carrying the stress, and another as a softer, de-aspirated sound [pʷ] (superscript ʷ is the IPA symbol for de-aspirated sounds). Again, technical details on how to pronounce these sounds will be given later. Strictly speaking, the first sound in *pet* should be [pʰ]. Yet exchanging the aspirated sound with the de-aspirated sound does not alter the meaning of word *pet*. No
native English speaker would contend you are not understood because of your pronouncing pet as [pʰɛt] instead of [pʰɛt]; at most he would say that you speak with a funny accent or your accent is thick. Different sounds that belong to the same phoneme and do not affect meaning are called allophones. Sounds [pʰ] and [pʰ] are allophones of the same phoneme /p/. Allophones are also said to be non-contrastive sounds, whereas phonemes are said to be contrastive sounds.

In order to make the text of this article easy to follow, we will try to use as few symbols as possible as long as clarity and precision is ensured. Some authors write phonemes in forward slashes / / and allophones in square brackets [ ]. We will follow that notation here with some restrictions. Only when narrow transcriptions (at the level of allophones) be really necessary for the discussion, we will use square brackets. Otherwise, we will try to use forward slashes. The context and usage rules will make clear what sounds are being referred to. For example, in general, de-aspirated sounds will not be notated (the superscript = will be dropped), but aspirated sounds will.

**Phonology** is the branch of linguistics concerned with the study of sounds in languages as abstract categories describing function and structure. **Phonetics** is concerned with physical properties of speech sound as well as with its production and perception. Phonology studies the phonemes of a language because they are the categories used by the speakers to extract meaning out of the speech sounds. Phonetics describes how the sounds of a language are produced (articulatory phonetics), how the speech sounds are transmitted from speaker to listener (acoustic phonetics), and how the listener perceives the speech sounds (auditory phonetics). As an example of the distinction phonology-phonetics, consider phoneme /t/ and its allophones in English. This phoneme can be pronounced up to seven phonetically different ways (again, precise details will be given later): as an aspirated sound [tʰ], as a de-aspirated sound [tʰ], as an unreleased sound [t'], as a glottal stop [ʔ], as a glottalized t [tʔ], as a flap t [ɾ], and as an omitted sound [∅] (represented here by the symbol ∅). In Figure 7 several words where those allophones take place are shown. See [PBL+06] and [Hay08] for a deeper discussion on phonemes and allophones.
### 3.2 Description of consonants and semivowels

A common way to define a **consonant** is as a sound produced by a partial or complete obstruction of the airstream (see [AE92], [LM96], [Bal12], for example, where this definition is either given explicitly or simply implied). This definition accounts for most of sounds considered as consonants in the English language, such as /t/, /g/, /f/, or /ɾ/. However, some sounds are produced like vowels, but they always function as consonants; they are the so-called **semivowels**. In the case of the English language, semivowels are sounds /j/, as in *yes/jes/*, and /w/, as in *well/wel*. Surprisingly enough, some sounds that are considered as consonants in a language may be considered as vowels in other languages. For instance, sound /ɾ/ is a consonant in English, both by manner of production and function, but it is a vowel sound in some Slavic languages. A broader definition of consonant is given by phonetician Ian Maddieson ([Sev12], Chapter 1). Following his definition, a consonant is a sound typically occurring at the syllable margins as opposed to vowels, which are sounds typically occurring in the syllable centers; we will come back to this matter in Section 4.6. In this article we will classify consonants both in terms of manner of sound production and function. Therefore, /j/ and /w/ will be regarded as consonants. However, we will still call them semivowels to make emphasis on their manner of production. For an insightful, concise discussion of the definition of consonant, the reader is referred to [Roa09].

From a sound production standpoint, we have established that in consonants there is a partial or complete obstruction of the airstream. That obstruction mainly takes place in the upper vocal tract, and in a few cases, the glottal consonants, in the larynx. Two factors are involved in the consonant production, namely, **what organs** are used to produce the sound—place of articulation —, and **how** the sound itself is produced—manner of articulation and voicing—. This gives rise to the following classification of consonants.
• By **place of articulation**, that is, where in the vocal tract the obstruction of the airstream takes place, and which speech organs are involved; see Figure 8.

  - **Bilabial**/bɪˈleɪbiəl/ **consonants**, when both lips are involved, as in /p/ or /m/.
  - **Alveolar**/ælˈviːəl/ **consonants**, when the alveolar ridge is used, as in /t/ or /z/.
  - **Post-alveolar**/poustælˈviːəl/ **consonants**, when the posterior part of the alveolar ridge is used, as in /ʃ/ or /ʒ/.
  - **Labio-dental**/ˈleɪbiədəntl/ **consonants**, when the consonant is produced by using the front teeth and the lower lip, as in /f/ or /v/.
  - **Dental**/ˈdentl/ **consonants**, when both upper and bottom front teeth are involved, as in /θ/ or /ð/.
  - **Velar**/vɛlər/ **consonants**, when the soft palate is involved, as in /k/ or /ŋ/.
  - **Glottal**/ɡlɒtl/ **consonants**, when the glottis is used, as in /h/ or /ʔ/.

The possible places of articulation are shown in Figure 8. Some places are not used in the English language.

![Figure 8: Different places of articulation (after [LM96]).](image)

• By **manner of articulation**, that is, the actual manner of articulating the consonant.

  - **Stops or plosive**/ˈplɔsɪv/ **consonants**: The airstream is blocked by two articulatory organs, then held for some time, and finally released, which causes the production of the sound. Stops are, for instance, /p/, /d/, /g/, or /ʔ/. Different organs participate in the production of each of those sounds.
- **Fricatives**/frɪkətɪvz/: There is a partial obstruction of the air produced by placing two organs very close together. An example of a fricative is /f/, where the two organs involved are the upper teeth and the lower lip. In this case, the airstream escapes between both organs through a narrow passage and thus the sound is produced. Other examples of fricatives are /s/, /θ/, and /ʃ/.

- **Affricates**/æfrɪkət/: A consonant that begins as a stop and is released as a fricative. In English there are two affricate sounds, /tʃ/ and /dʒ/.

- **Nasals**/nezl/: In this kind of consonants the air can escape freely from the nose as well as from the mouth. Nasal consonants in English are /m/, /n/, and /ŋ/.

- **Approximants**/ɔˈpræksɪmənt/: The sound is formed by the coming of two articulators close together but not as much as in the case of fricatives. Within the approximants, two class of sounds can be distinguished in English, the **r-consonants** and **lateral approximants**. Both will be described in detail below. Examples of approximants in English are /ɹ/ (r-consonant) and /l/ (lateral approximant).

- **By voicing**. Before reaching the upper vocal tract, where the consonant sound is produced, the air passes through the vocal folds. If they vibrate during the articulation of the sound the consonant is **voiced**; otherwise, it is said to be **voiceless** or **unvoiced**. This distinction is very important not only in terms of the quality of sound (different timbre) but also for grammatical reasons, as we will see later.

Table 1 shows the consonant sounds found in English. When consonants are shown in pairs, voiceless consonants appear to the left of the black bullet and voiced consonants to the right. Main allophones have also been included in the Table, which will be explained in detail later.
Table 1: English consonants and semivowels.

<table>
<thead>
<tr>
<th>Consonants</th>
<th>Bilabial</th>
<th>Labio-Dental</th>
<th>Dental</th>
<th>Alveolar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nasal</td>
<td>● m</td>
<td></td>
<td></td>
<td>● n</td>
</tr>
<tr>
<td>Stops</td>
<td>pʰ, p, pʰ, ● b</td>
<td></td>
<td>tʰ, t, tʰ, tʰ, ● d</td>
<td></td>
</tr>
<tr>
<td>Fricatives</td>
<td>f ● v</td>
<td>0 ● ð, ð</td>
<td>s ● z</td>
<td></td>
</tr>
<tr>
<td>Affricates</td>
<td></td>
<td></td>
<td></td>
<td>● j</td>
</tr>
<tr>
<td>Central</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximants</td>
<td></td>
<td></td>
<td></td>
<td>● r</td>
</tr>
<tr>
<td>Flaps (taps)</td>
<td></td>
<td></td>
<td></td>
<td>● l, lʰ</td>
</tr>
<tr>
<td>Lateral</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximants</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Consonants</th>
<th>Post-Alveolar</th>
<th>Palatal</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nasal</td>
<td></td>
<td>● n</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stops</td>
<td></td>
<td></td>
<td>kʰ, k, kʰ, ● g</td>
<td>●</td>
</tr>
<tr>
<td>Fricative</td>
<td>f ● v</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affricates</td>
<td>lʃ ● lʒ</td>
<td></td>
<td>h ●</td>
<td></td>
</tr>
<tr>
<td>Approximants</td>
<td></td>
<td>● j</td>
<td>● w</td>
<td></td>
</tr>
<tr>
<td>Semivowels</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retroflex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximants</td>
<td>● l</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.3 Voicing of consonants

The distinction of consonants by voicing is essential in English and therefore crucial to acquiring good listening skills. Minimal pairs are pairs of words that only differ by one sound, such as pull/pul/-pool/pu:l/ or eat/i:t/-it/ːt/. Minimal pairs based on voicing are very common in English. Those are pair of words where all sounds are identical except for one whose voicing is different. Consider pairs pet/pet/-bet/pet/, sip/sip/-sip/zip/, ten/tend/-den/den/, or chin/ʃɪn/-gin/ʃɪm/; the only difference between each pair is the voicing at the initial sound, enough to alter their meaning. John Higgins has compiled a complete and illustrative table of minimal pairs for British English; see [Hig]. Regrettably, we do not of any comprehensive compilation of minimal pairs for American English. The following table shows the English consonants paired by voicing (voiceless are on the left and voiced on the
Table 2: Minimal pairs based on voicing.

Consonant sounds that are not paired by voicing are the following:

- Voiced sounds: nasals /m/, /n/, and /ŋ/; r-consonants /ɹ/, /l/, and /r/; lateral approximants /l/ and /ɹl/; and semivowels /j/, and /w/.
- Voiceless sounds: fricative /h/ and stop /p/.

In addition to the semantic distinction given by minimal pairs, voicing also governs other important cases. Voicing may determine grammatical function. For instance, *use/*juːs/* is a noun pronounced with voiceless sound /s/, but as a verb *use/*juːz/* is pronounced with voiced sound /z/*. Another important case is that of pronunciation of plurals, possessives and third person of present simple. Again, they are also formed according to voicing. In order to explain how those grammatical cases are formed, we need to introduce a special type of consonants called *sibilant* /sɪbɪl/ consonants. A sibilant consonant is either a fricative or an affricate consonant where the tongue makes a groove through which the airstream is directed. The groove is central so that the airstream flows across the centre of the mouth over the tongue. Table 3 lists the sibilant consonants found in English.

Table 3: Sibilant consonants.

Having defined sibilant consonants, we are now in the position of setting out the rules for the pronunciation of the plural of nouns.

1. If a word ends by a vowel sound, then its plural will be pronounced by adding sound /z/. For example:
   - *Eye/*aɪ/* → *eyes/*εɪz/*.
2. If a word ends by a non-sibilant sound, then its plural will be pronounced by adding sound /s/ when it is a voiceless consonant, and by adding sound /z/ when it is a voiced consonant. For example:
   - *Pet*/pet/* → *pets*/pets/* (voiceless case).
   - *Dog*/dɑːɡ/* → *dogs*/dɑːɡz/* (voiced case).
   - *Folk*/fɑːk/* → *folks*/fɑːks/* (voiceless case).
   - *Coin*/kɔɪn/* → *coins*/kɔinz/* (voiced case).

3. If the word ends by a sibilant sound, then the plural is pronounced by creating a new syllable /iz/. For example:
   - *Beach*/biːʃ/* → *beaches*/biːʃiz/*.
   - *Bridge*/brɪdʒ/* → *bridges*/brɪdʒiz/*.
   - *Bush*/bʌʃ/* → *bushes*/bʌʃiz/*.
   - *Garage*/gærɪ:ʒ/* → *garages*/gærɪ:ʒiz/*.
   - *Bus*/bʌς/* → *buses*/bʌςɪz/*.
   - *Rose*/rəʊz/* → *roze*/rəʊzɪz/*.

Exceptions to the previous rule can also be found (irregular plurals):
   - *House*/haʊs/* → *houses*/hauzɪz/*.
   - *Mouths*/maʊθ/* → *mouth*/maʊðz/*.
   - *Bath*/bæθ/* → *baths*/bæðz/*.

Moreover, the same pronunciation rules are applied to possessives:
   - *Mark*/mɑːrk/* → *Mark’s*/mɑːrkz/*.
   - *Joe*/dʒɔi/* → *Joe’s*/dʒɔiz/*.
   - *George*/dʒɔrɛɡ/* → *George’s*/dʒɔrɛɡɪz/*.

When the third person of singular of present simple is written, it takes either an *s* or an *es*. The pronunciation of the new added consonant follows the same rules as in the plural.
   - *I lie*/aɪˈliː/* → *He lies*/hiˈlaɪz/* (vowel syllable-final case).
• I want/ˈaɪ.wɑnt/ → He wants/ˈhi.wɑnts/ (non-sibilant voiceless case).
• I read/ˈaɪ.rɪd/ → He reads/ˈhi.ri:d/ (non-sibilant voiced case).
• I teach/ˈaɪ.tʃ/ → He teaches/ˈhi.tʃ/ (sibilant case).

The previous examples should have made clear how significant voicing is. Voicing should be perceived with crystal clarity and produce with great precision if an intelligible pronunciation is pursued.

4 The American English consonants

4.1 Stops

Stops or plosive consonants. The sound is produced in three stages; see Figure 9:

• Closure/ˈkloʊzər/: the oral cavity is completely blocked by the speech organs involved.

• Blockage/ˈblækIz/: The oral cavity is held blocked for some time (some milliseconds). The air from the lungs continues to come into the oral cavity. Therefore, the air pressure inside the oral cavity increases. This stage is also called compression/kəmˈprɛfn/stage.

• Release: Finally, the blockage is released. Due to the air pressure difference between the oral cavity and outside, the air is expelled with a burst or puff (hence, the name of plosive).

According to the organs involved in the closure of the oral cavity, stops are classed as follows:

• Bilabial stops: the airstream is held by using both lips. Bilabial stops are [pʰ], [p], [p′], and [b].

• Alveolar stops: the airstream is held by pushing the tip of the tongue against the alveolar ridge. Alveolar stops are [tʰ], [t], [t′], and [d].

• Velar stops: the airstream is held by pushing the back of the tongue against the soft palate. Velar stops are [kʰ], [k], [k′], and [g].

• Glottal stops: the airstream is held by closing the vocal folds (remember that the space between the vocal folds is the glottis). Glottal stops are [ʔ] and [tʰ]. Sound [ʔ] is a pure glottal stop, not associated to any other sound, while [tʰ] is a co-articulation of [t] and [ʔ] (see Section 4.1.5).
Figure 9: Articulation of plosive consonants (after [Wik12d]).

Stops can be either voiced or voiceless. When they are voiceless, three allophones or variants arise, the **aspirated**/æspərtid/ sound, the **de-aspirated**/dɪːæspərtɪd/ sound, and the **unreleased** sound. Aspiration will be indicated by writing a superscript h, as in [pʰ]; de-aspiration by the absence of superscript; and when unreleased by superscript ‘.

How does aspiration work? Sometimes, a plosive is located between two vowels. If the stop is voiced, the vocal cords will continue vibrating. However, if the stop is voiceless, the vocal cords will stop vibrating for some time. If they start vibrating after the release stage, the stop will be aspirated (see Figure 9, bottom). When the vocal cords start vibrating before the release stage, that vibration disturbs the air coming from the lung, making the air pressure decrease; the burst is not as intense as in the previous case. Therefore, the stop will be de-aspirated (see Figure 9, middle). Similar situations can be found when a plosive sits between other consonant sound, particularly fricatives.

All aspirated sounds follow the same rules for their pronunciation.

1. Voiceless stops are aspirated when they begin a stressed syllable, no matter if it is not the first syllable of the word. Examples: *opinion*[ˈɒpɪˈnɪʃn], *ten*[tɛn], *recall*[rɪˈkɔːl].

2. When it is the first sound of a word irrespective of where the stressed syllable is, as in *tempting*[ˈtempɪŋ], or *potential*[ˈpəʊtəˈnɔːl]

3. Voiceless stops are de-aspirated when the first sound is [s], as in *spot*[spɑːt], *stop*[stɑːp], or *skunk*[skɑŋk] (no superscripts).

4. Word-final voiceless stops optionally can be aspirated. Sometimes this depends on emphasis, speech speed, or the particular accent.

As for rule 3, some authors contend (Rachel herself [Eng12], for example) that in clusters formed by sound [s] plus a plosive, the plosive is still aspirated. We think that in a very
emphatic enunciation that might be the case, but in general we incline to the view that those plosive are de-aspirated. Davidsen-Nielsen \[DN69\] looked into this question carefully. He found that native speakers perceive sounds [p], [t], and [k] as [b], [d], and [g], respectively, when those sounds are isolated from pairs [sp], [st], and [sk]. It seems that in the case of [s] followed by voiceless plosive plus vowel the vocal cords start vibrating much earlier than in other cases. This fact prompted Davidsen-Nielsen to even propose transcribing [sp], [st], and [sk] followed by vowel as [sb], [sd], and [sg]. The tenet in phonetics is that in general plosives before fricatives are de-aspirated, as in \[æft=æt\] or \[æstray=æt\]. For more information on this issue, see \[DN69\], \[Roa09\], \[Gie92\], \[Wel90\], and the references therein.

Finally, let’s describe the third allophone of plosive phonemes, the unreleased stop. In this sound the blockage is still created, but much less air is accumulated behind the speech organs, and the release is not audible. It appears in word-final positions, specially in colloquial speech, as in \[kat\] or in the presence of stop clusters, such as in \[apt\] or \[loged on\]. In the latter case, both stops are somehow merged into one and the release of the first stop is removed. In the video in Figure 11 we can see an example where plosive [p] is unreleased so that the next word can be properly stressed in the sentence. Unreleased stops are frequent in voiceless consonants and to a lesser extent in voiced consonants.

4.1.1 Bilabial sounds [p\textsuperscript{h}], [p], [p\textsuperscript{r}], and [b]

The closure of the oral cavity is carried out by the lips, which come together to stop the airstream. Then the airstream is held just behind the lips, and finally is released; see Figure 10. There are three voiceless allophones, [p\textsuperscript{h}], [p], [p\textsuperscript{r}], and one voiced allophone, [b]. Spellings commonly associated with bilabials are \[p, pp\] for sound [p], as in \[pen\], \[happy\], or \[lap\], and \[b, bb\] for sound [b], as in \[best\] or \[robber\].

![Figure 10: Manner of articulation of bilabial stops [p] and [b] (after \[Bal12\]).](image)

In this video Rachel describes all allophones of phoneme /p/ (she notates the unreleased stop by [p\textsuperscript{r}] instead of the IPA symbol [p\textsuperscript{r}]).

Aspiration does not exist in Spanish and in general Spanish speakers tend not to aspirate bilabial stops in word-initial position. This make their accent unpredictable and of blurry
4.1.2 Alveolar sounds [tʰ], [t], [t̪], and [d]

In alveolar sounds the closure is produced by the tongue blocking the airstream against the alveolar ridge; see Figure 12. In order to do so, the tongue widens to seal off the oral cavity. It is the blade of the tongue that makes contact with the alveolar ridge; the tip of the tongue is in rest position just behind the upper front teeth. As in the case of voiceless bilabial stops, there are an aspirated and de-aspirated allophones.

Figure 12: Manner of articulation of alveolar stops [t] and [d] (after [Bal12]).

In some languages, such as Spanish, phoneme /t/ is pronounced as a dental stop [t], that is, the place or articulation is at the behind the upper teeth instead of the alveolar ridge; see Figure 13. Consequently, in Spanish phoneme /t/ is articulated with the tip of the tongue rather than with the blade. Although pronouncing phoneme /t/ as a dental consonant is not contrastive, it sounds unnatural to native ears and therefore students should attach importance to it. Again, Spanish speakers may not aspirated /t/ in word-initial position.

Common spellings of phoneme /t/ are:
• *t* and *tt*, as in *ten* [tʰən], *stop* [stəp], or *wait* [wet].

• The past simple of a regular verb ends in phoneme /t/ when the last sound of the verb is a voiceless sound different from /t/: *trip* [tʰrip] → *tripped* [tʰrip t], *ask* [æsk] → *asked* [æsk t], *laugh* [lɑːf] → *laughed* [læft], *toss* [tɔːs] → *tossed* [tɔːst], *mash* [mæʃ] → *mashed* [mæʃt]. Past simple endings are often aspirated to clearly distinguish them from other tenses. Past simple endings are seldom unreleased.

• Spelling *th*, although less frequent, can be found, as in *Thailand* [tʰaɪlænd].

In this video Rachel describes all allophones of phoneme /t/ and the pronunciation of letter *t* as an alveolar flap [ɾ]. This sound is actually an r-consonant and will be covered in Section 4.5.1.
In this video Rachel covers the pronunciation of letter t, which certainly includes the allophones studied in this Section. The video is shown here for the sake of completeness.

Figure 15: Video explaining the pronunciation of letter t.

This Rachel video is a follow-up where the reader is provided with many practical examples of how to pronounce letter t, incidentally to pronunciation of phoneme /t/:  

Figure 16: Practical examples of how to pronounce the letter t.

The main problem native Spanish speakers have with alveolar stops has already been pointed out: the pronunciation of alveolar /t/ as a dental /t/. This is an important mis-pronunciation to correct as sound /t/ is common in English and appears in many consonant clusters.

4.1.3 Velar sounds [kʰ], [k], [k’], and [g]

In velar sounds the closure is produced by the back of the tongue leaning against the soft palate, which blocks the airstream; see Figure 17. Similarly to bilabial and alveolar stops,
aspirated, de-aspirated, and unreleased allophones (\([k^h], [k], [k']\)) can be observed. When the stop is voiceless, sounds \([k^h], [k], [k']\) are produced. If the stop is voiced, sound \([g]\) results.

![Figure 17: Manner of articulation of alveolar stops \([k]\) and \([g]\) (after [Bal12]).](image)

Common spellings of phoneme /k/ are:

- Letter c in word-initial, as in *cat*\([k^hæt]\), and also within a word, as in *fact*\([fæk't]\).
- Letter k in word-initial, as in *keep*\([ki:p]\), and also within a word, as in *like*\([lark]\).
- Combinations *ck* and *ch*, as in *black*\([bæk]\) and *school*\([sku:l]\).
- Combination *qu*, as in *quiet*\([k^hwaɪt]\).
- Letter *x*, as in *six*\([siks]\).

Sound [g] is normally spelled g or gg, as in *girl*\([g^3:rl]\) or *egg*\([eg]\).

In the video below only \([k^h], [k], [g]\) are explained.

Spanish speakers also have problems pronouncing sound [g]. This sound is not equivalent to the corresponding [g]-sound in Spanish, which is a *velar approximant* \([\gamma]\). Spaniards unaware of the pronunciation of sound [g] replace this sound with sound \([\gamma]\). This is the reason why many Spaniards do not pronounce perfectly words as simple as *good*. In fact, so soft is the Spanish pronunciation of *good* (the [g]-sound in general) that native English speakers often confuse it with *wood*. This feature has to be taken care of if a polished accent is to be acquired. For more information on this topic, see [Mot11] for technical details (such as actual blockage times), [Yav07] for comparative phonology, or [Wik12a].

### 4.1.4 Glottal sound [ʔ]

A glottal stop is a voiceless sound produced by the closure of the glottal space by the vocal cords. Since it is a stop, there is blockage stage after which the air is released.

The glottal stop is an allophone of phoneme /t/. Normally, it substitutes the de-aspirated sound \([t]\). Rules for such a substitution are the following:
1. The glottal stop appears at the end of words, as in put[pu?] or report[rɪˈpɔːr?].

2. The glottal stop also occurs in the presence of a stressed syllable followed by:

   (a) A nasal vowel, specially with patterns [t+vowel+n] or [tn], as in button[ˈbaʊn], or continent[ˈkɒnɪmənt]. This is the most frequent case both in American and British English.

   (b) A fricative or a stop, as in outside[ˈaʊsaɪd], patsi[ˈpætsi], football[ˈfʊtbɔːl], outfall[ˈaʊtfaːl].

   (c) a semi-vowel [j] or [w], or sound [l]. Examples are right way[ˈraɪtweɪ], can’t you?[ˈkæntju:], or brightly[ˈbraɪtli].

For more information on the glottal stop, see Roa09 (for a good explanation of its anatomy), as well as AE92, Wel12, LM05, Wik12b, LM96. David Brett Bre12 wrote a nice article where he gives several examples of glottal stop, including excellent sound files. The next video by Jennifer Jen12 explains very clearly how glottal stops work in American English. Warning: sometimes, depending on how the video is recording or the sound system where it is played back, it is difficult to perceive the sound of a glottal stop.
Glottal stops are found both in British and American English as allophones of phoneme /t/. In the case of British English the use of glottal stops is extended to other voiceless stops, such as /p/ and /k/ in word-final, as in *rocker/*rəʊər/ or *upper/*ʌpər/. An extreme case is the dialect known as Cockney English, where word-final /t/ is substituted by [ʔ] systematically. In the video below Richard from Linguaspectrum puts forward examples from Cockney accent.

4.1.5 Glottalized sound [tʰ]

A glottalized/*ɡlɒtəlzd/ stop is the occurrence of a glottal stop at the same time or just immediately after another consonant. This is represented in IPA by adding ʔ as a superscript to the main sound. Consonants that are formed by the production of two simultaneous sounds are called co-articulated consonants; glottalized stops are thus co-articulated consonants. In American and British English the most common glottalized stop is [t], which is accordingly represented by [tʰ]. This phenomenon also receives the name of glottal
reinforcement. In a glottalized [tʰ] the stop [t] and the glottal stop [ʔ] are produced at the same time or [ʔ] right after [t]. For its production, this allophone follows the same rules as the glottal stop does. Example of glottalized t are *mutton* [ˈmatn], or *curtain* [ˈkɜːrn].

Glottal reinforcement can be found in British English to strengthen [t] or [tr] at the end of a syllable, among other cases; see [Wel12] for a thorough discussion and [Bre12] for actual examples of glottal reinforcement. For more information about glottal reinforcement and glottalized t, see [Chr52], and [LM96].

4.2 Fricative sounds

In fricative consonants there is a partial obstruction of the air produced by placing two articulatory organs very close together. Normally, one of the organs is a passive articulator (see Section 2.2) and the other is an active articulator (the tongue in most cases). The narrow passage causes a turbulent airstream, which produces the fricative sound. The sound produced by the turbulent airstream is called frication /fraɪˈkeɪʃən/. According to the organs involved in the partial obstruction of the oral cavity, fricatives are classified as follows:

- **Labio-dental sounds**: The frication is caused by loosely placing the upper teeth on the bottom lip. Labio-dental sounds are [f] and [v].

- **Dental sounds**: The frication is caused by placing the tongue between the upper and bottom teeth. Dental sounds are [θ], [ð], and [ɹ].

- **Alveolar sounds**: The frication is caused by moving the blade of the tongue near the alveolar ridge. Alveolar sounds are [s] and [z].

- **Post-alveolar sounds**: The back of the tongue forms a narrow passage with the hard palate (post-alveolar) that results in the frication. Post-alveolar sounds are [ʃ] and [ʒ].

- **Glottal sounds**: The frication is achieved by putting the vocal folds in a near-close position and letting the air go through them. [h] is the only fricative glottal sound in English.

4.2.1 Labio-dental sounds [f] and [v]

The upper front teeth come close to the bottom lip, while the tongue is in rest position. The airstream is forced through the slit formed by those two articulators; see Figure 21. Consonant [f] is voiceless and consonant sound [v] is voiced.
Spellings $f$, $ff$, $ph$, and $gh$ are frequently associated to sound $[f]$; see the following examples: first $[fəːst]$, coffee $[kʰəːfi]$, photo $[fəʊtoʊ]$, or cough $[kʰɑːf]$. The only spelling associated to $[v]$ is letter $v$.

The main problem Spaniards encounter with sound $[v]$ is that this sound does not exist in Spanish. Letter $v$ does exist and is pronounced either as $[b]$ or as a bilabial fricative $[β]$, which causes mispronunciation.

### 4.2.2 Dental sounds $[ð]$, $[ŋ]$, and $[θ]$

The tip of the tongue is placed between the upper and bottom front teeth, making loose contact. The air is forced its way through the narrow passage formed by the teeth and tongue, which causes the frication characteristic of these sounds. There is a voiceless sound $[θ]$, and two voiced allophones $[ð]$ and $[ŋ]$. The latter is used when a word is not fully enunciated (in particular in the context of consonant reduction; see [AE92]). In $[ŋ]$ the tongue does not go through the teeth fully, but just stays behind the teeth and presses them a little bit. This allophone is called a dental approximant. In the video in Figure 23 this sound is covered.
Both phonemes /θ/ and /ð/ are associated to spelling th, as in thanks[θæŋks], these[ðiːz], or Tell me the time[tʰelmiːθætʰaɪm] (definite article the is often pronounced as the dental approximant).

Figure 23: Video explaining the pronunciation of dentals.

Native Spanish speakers tend to devoice phoneme /ð/. The closest sound in Spanish is precisely phoneme /θ/, the voiceless counterpart of /ð/. Also problematic is the substitution of [ð] for [d] (or its equivalent sound in Spanish). This is due to the fact that [ð] never appears in word-initial position in Spanish, whereas [d] always does.

4.2.3 Alveolar sounds [s] and [z]

Sounds [s] and [z] have the same manner and place of articulation; they only differ in voicing. Sound [s] is voiceless, whereas sound [z] is voiced. As for their articulation, the blade of the tongue is raised towards the roof of the mouth. The tip of the tongue rests just behind the bottom front teeth; see Figure 24.

Figure 24: Manner of articulation of alveolar fricatives [s] and [z] (after [Bal12]).
Furthermore, earlier on we defined sounds [s] and [z] as being sibilants. The tongue is thus curved along its medial axis, with its sides actually touching the alveolar ridge and forming a narrow passage or groove through which the airstream passes. That is the way this sibilant consonant is produced. Lips are parted and the corners pull back as in a smile. In Figure 25, which was taken from [LM96], we can see the articulatory gesture for [s] as in *saw* pronounced by Ladefoged. On the right, the solid line shows the position of the tongue, as extracted from x-rays, whereas the grey line indicates the position of the sides of the tongue, as given by palatograms. Between both lines, it is possible to imagine the exact tongue shape for this sound. On the left, it is shown a transverse view taken from the coronal section (indicated by the arrow on the left).

![Figure 25: Tongue shape for alveolar fricatives [s] and [z] (after [LM96]).](image)

Note that in Figure 25 the tip of the tongue is not exactly behind the bottom front teeth as described above, but just behind the upper front teeth. There are certain positional variation for the tongue in this respect.

Consonant [s] has a quite characteristic high-pitched, hissing sound that other similar [s]-sounds from other languages do not to have. For example, compared to the Spanish [s]-sound, some differences with respect to the English [s]-sound can be perceived. Spanish [s] is *apical*[^1] (apikl], that is, it is produced with the tip of the tongue rather than with the blade (these consonants are called *laminal*[^2] (leman]). Some authors even speak of a little retroflex (curling back) of the tongue in the case of the Spanish [s]-sound, which would account for differences in timbre and pitch (Spanish [s] is perceived as lower in pitch than English [s]).

In the video below Rachel explains both sounds [s] and [z] and pays close attention to tongue and lip position.

[^1]: -pikl]
[^2]: -man]]
Common spellings of sound [s] are:

- Letter s in word-initial, as in *son [sən]*, and also within a word, as in *bus [bʌs]*.
- The formation of plural, third person of present simples, and Saxon genitive require adding an [s] if the word-final sound is a voiceless non-sibilant. See Section 3.3 on voicing of consonants above.
- Letter c in word-initial, as in *city [ˈsaɪti]*, and also within a word, as in *pencil [ˈpɜːnsil]*.
- Spelling *sc* is sometimes pronounced as [s], as in *scissors [ˈsɪzərz]*.

Common spellings of sound [z] are:

- Letter z in word-initial, as in *zoo [zuː]*, and also within a word, as in *easy [iːzi]*.
- The formation of plural, third person of present simples, and Saxon genitive require adding an [z] if the word-final sound is a voiced non-sibilant and [tʃ] if the word-final sound is a sibilant, either voiced or voiceless. Again, check out Section 3.3.
- Combination *zz*, as in *dizzy [ˈdɪzi]*.
- Spelling *ss* is sometimes pronounced as [z], as in *scissors [ˈsɪzərz]*.

Another pronunciation problem encountered among native Spanish speakers is that of the distinction of minimal pair [s]-[z]. Sound [z] does not exist in Spanish and is systematically replaced by voiceless sound [s], the apical version of English alveolar sound [s]. This is one of the main problems since voicing is phonemically very important in English.

In Spanish [s] in word-initial is always preceded by a vowel. This makes some native Spanish speakers to insert vowel [e] before [s] in words like *stop*, which is mispronounced [ɛstə:p] .
4.2.4 Post-alveolar sounds \([f]\) and \([z]\)

The two post-alveolar sounds take the same tongue position, \([f]\) being the voiceless consonant and \([z]\) its voiced counterpart. In these sounds the front of the tongue is curled (hence, the term post-alveolar) and approaches the roof of the mouth without touching it; see Figure 27. The hard palate is a relatively large area. Here the tongue approaches the area immediately after the alveolar ridge (hence, the term post-alveolar) as opposed to palatal/\(\text{pa}l\text{æl}/\) consonants, where the approximation takes place at the centre of the hard palate. Since it is also a sibilant consonant, the centre of the tongue forms a groove or channel. The sound is produced by the frication created when the airstream passes through that narrow passage. The tip of the tongue is relaxed and slightly touching the bottom front teeth. The lips are a little bit rounder to help channel the airstream.

![Figure 27: Pronunciation of sounds \([f]\) and \([z]\) (after Bal12).](image)

Taken from the paper by Stone and Lundgerg \[SL96\], the tongue shape corresponding to sound \([f]\) is shown in Figure 28. This is a 3-D reconstruction by using spline interpolation from 2-D measures on the tongue. The groove characteristic of this sibilant sound is quite noticeable. The air is directed by the tongue through this groove along plane A and perpendicularly to plane B. See Section 6.3 for further discussion about tongue shapes and consonant pronunciation.
The most common spelling for \([\text{ʃ}]\) is \(sh\), as in \(\text{shop}[\text{ʃ}][p]\); less common spellings are: \(c\), as in \(\text{ocean}[^{\text{o}}][\text{ʃ}][\text{n}]\); \(ch\), as in \(\text{machine}[^{\text{m}}][\text{ʃ}][\text{i:n}]\); or initial \(s\), as in \(\text{sure}[^{\text{s}}][\text{ʃ}][\text{ɔːr}]\). As for the voiced sound \([\text{ʒ}]\), which is rare in English, it is normally spelled \(s\) or \(si\), as in \(\text{usual}[^{\text{juː}}][\text{ʃ}][\text{ɔːl}]\) or \(\text{Asia}[^{\text{e}}][\text{ʃ}][\text{əsǐə}]\).

Here we have Rachel’s video explaining the pronunciation of these two sounds.

![Video](image)

**Figure 29:** Video explaining the pronunciation of post-alveolars.

### 4.2.5 Glottal sound \([h]\)

This is a complicated sound to describe from a technical point of view. By its manner of articulation it is classed as a fricative, although for some authors this sound lacks some basic characteristics of a consonant and is, therefore, treated as a vowel or a transitional\([\text{tren}'\text{sʃən}]\)
state of the glottis [LM96]. In this article it will be considered a glottal fricative. The sound [h] is produced by the vocal cords coming close together, without touching each another, and vibrating. Since the sound is produced in the larynx, the mouth position usually corresponds to the next sound in the word; see the video by Rachel below. In the video below around minute 0:45 we can see how the vocal cords produce sound [h].

Figure 30: Video showing the movement of the vocal cords when [h] is pronounced.

Sound [h] is usually spelled h, as in here[hɪ].

Figure 31: Video explaining the pronunciation of [h].

In Spanish voiceless glottal fricative [h] is not present. The closest sound is [x], a voiceless velar fricative. Native Spanish speakers tend to substitute [h] for [x]. Since sound [x] is strong and somewhat rough as pronounced in Spanish, its use produces a quite unnatural and thick accent.
4.3 Affricate consonants

An affricate consonant begins as a stop and is released as a fricative. In English there are only two affricate sounds, \[\theta\] and \[\delta\].

4.3.1 Consonants \[\theta\] and \[\delta\]

Sound \[\theta\] is voiceless and sound \[\delta\] is voiced. Both are the combination of an alveolar stop and a post-alveolar fricative. The IPA symbols indicate the combination of the consonant sounds ([t]+[ʃ]=[θ] and [d]+[ʒ]=δ]). In the case of the voiceless affricate \[\theta\], stop [t] is less aspirated than [th]; the manner and place of articulation of the post-alveolar is exactly the same as in [ʃ]. This also applies to the voiced sound \[\delta\]. Common spellings for \[\theta\] are ch, t, and tch, as in choose/[θuːz/; question/ˈkwestən/], or catch/[kæθ]. In the case of \[\delta\], its most common spellings are j in initial position, as in job/ʤəʊb/; letter g, as in general/ˈdʒenərəl/; and combinations ge and dge, as in large/laːdʒ/ and fridge/ˈfrɪdʒ/.

Figure 32: Video explaining the pronunciation of affricates.

The first pronunciation problem with the affricates is associated with stops [t] and [d]. They are pronounced by native Spanish speakers as dental stops [θ] and [δ] instead of their alveolar counterparts (see Figure [13], Section 4.1.2). The second problem is the aspiration of stops, as discussed in Section 4.1.3. Aspiration is not present in Spanish and therefore affricates sound weak or poorly enunciated to English ears. Third, since fricative [ʃ] does not exist in Spanish, it is often replaced by affricate \[\theta\], which does exist in Spanish (not exactly the same). Finally, voiced affricate \[\delta\], not present in Spanish, is often pronounced as voiceless \[\theta\].

Another kind of problem found in native Spanish speakers is the pronunciation of affricate \[\delta\] as semivowel [j], yielding juice rather than use (noun).
4.4 Nasal consonants

In this kind of consonants the air can escape freely from the nose because the soft palate does not lean against the rear wall of the throat but just hangs loosely. Therefore, the air passes through the oral and nasal cavities. Nasal consonants in English are [m], [n], and [ŋ]. However, these consonants are actually stops and during the closure and blockage stages the air only escapes from the nose.

4.4.1 Stops [m], [n], and [ŋ]

Consonant [m] is a bilabial stop, consonant [n] is alveolar stops, and consonant [ŋ] is a velar stop. They can be thought of as the nasal version of the oral consonants [b], [d] and [g], respectively. In Figure 33, the difference in terms of manner and place of articulation between sounds [m] and [p, b] can be observed. The soft palate is lowered allowing the air escape from the nose. These changes several features of the produced sound, as resonance, amplitude (it is lower than in stops or other consonants), and timbre.

![Figure 33: Pronunciation of sound [m] as compared to oral bilabial stops (after Bal12).](image)

Sound [m] is usually spelled m or mm, as in more[moːr] or comb[koʊm]. In the video below Rachel explains the pronunciation of this sound.
Apart from nasality, sound [n] is pronounced as [t] in terms of manner and place of articulation, that is, it is an alveolar stop. Sound [n] is usually spelled n, nn, or kn, as in *not* [nʌt], *sunny* [sʌni], or *know* [nəʊ]. In the video below Rachel explains the pronunciation of this sound.

Last nasal sound is [ŋ], a velar stop like [k] and [g]. In Figure 36 sounds [ŋ] and [k] are compared. In the case of [ŋ] the back of the tongue and the soft palate meet in such a way the nasal tract is not sealed off.
Sound [ŋ] is not contrastive in Spanish and speakers often drop it in word-final position, making *sing* [sm] and *sin* [sn] sound the same.

### 4.5 Approximant Consonants

The definition of an approximant /ˌprɔːkˈsmənt/ consonant is somehow delicate, as observed by several authors (see [LM96] and [MC04]). The sound is formed by the coming of
two articulators close together, neither as much as creating frication nor as being treated as a vowel sound. Within the class of approximants, three subclasses of sounds occurring in English can be distinguished:

- **Central approximants**: The airstream is directed through the centre of the mouth by the tongue. Within this group, we find r-sounds such as alveolar approximant [ɹ], flap approximant [ɾ], retroflex approximant [ɻ] as well as semivowels [j] and [w].

- **Lateral approximants**: The airstream is directed to the sides of the mouth by the tongue. The centre of the tongue (often the tip) actually touches the roof, but the approximation to the alveolar ridge is produced with the side of the tongue; hence, the term for its classification. Sounds [l] and [ɹ] are lateral approximants.

### 4.5.1 r-Consonants [ɹ], [ɾ], and [ɻ]

In English letter r can be pronounced in three different ways, namely, [ɹ], [ɾ], and [ɻ], the three being r-consonants. They are given that name because letter r is usually pronounced as one of these sounds. More technically, they are called rhotics/r-oʊtɪks/; see, for example, [LM96].

Sound [ɹ] is an alveolar approximant, which means that the sound is produced by moving the tongue close to the alveolar ridge, but not as much as creating frication. Here there is a complete description of the tongue movements:

1. The front of the tongue is first pulled back a little bit and then raised.
2. More importantly, the back of the tongue is expanded so that it touches the bottom of the top teeth (near the molars) at both sides. This forms a central channel over which the airstream passes (sound [ɹ] is a central consonant).
3. The blade and tip of the tongue are in neutral position without making contact with other parts of the mouth.
4. The lips are somewhat rounded, more on word-initial positions, as in red[r-ed], and less in word-internal positions, as in camera[kæmərə] or stress[stres].

Note that the term approximant does not mean the tongue does not touch the roof of the mouth at all; it means that where the sound is produced, there is only approximation, but other parts of the tongue may touch the roof of the mouth, as it actually occurs in sound [ɹ]. The approximation in this case takes place at the central channel formed by the tongue. Within this sound there is great variation as the position of the tip is concerned. Some authors contend that the tongue is a little bit curved up rather than in a neutral position. For example, Roach [Roa09] (pages 49–50) speaks of the tip of the tongue “approaching the alveolar area in approximately the way it would for a [t] or [d], but never actually making contact with any part of the roof of the mouth”. Other authors observe that in general the tip and blade of the tongue may adopt the position of the previous sound, as it occurs in combination such as [tr] or [dr], as in train[tɹeɪn] or drain[drɛɪn].
Sound [ɪ] appears at prevocalic positions in a syllable or syllable-clusters, as in red[red], camera[kæmərə], train[træn], confrontation[kənfrəntəʃən], or program[prəˈɡræm].

Rachel summarizes part of the above discussion in the video below.

![Video explaining the pronunciation of [i]](image)

Figure 38: Video explaining the pronunciation of [i].

The second r-consonant is the **alveolar flap** [ɾ]. This is an allophone of phonemes /r/ and /t/ and it only occurs in American English. The manner of articulation is essentially the same as in the sound [ɪ], except for the flap. As in [ɪ], the front of the tongue is pulled back and raised, but then the tip of the tongue makes a rapid movement upward and then downward, slightly touching the alveolar ridge. This movement is called a **flap**.

As an allophone of phoneme /r/, sound [ɾ] appears at intervocalic position when the stress of the word is carried by the first vowel, as in cereal[ˈsɛriəl], or carol[ˈkærəl]. However, notice that at other positions this allophone is not pronounced, as in camera[kæmərə], because the vowel before letter r is not stressed. This alveolar flap also happens when words are linked together, as in Don't jeer at him, which is pronounced [ˈdəʊnt dʒɪəræθm].

Sound [ɾ] is also an allophone of phoneme /t/, perhaps one of the most distinguishing pronunciation features in American English. Mainly, it substitutes allophones [tʰ] or de-aspirated [t]. This change takes place when /t/ is at an intervocalic position when the first vowel is stressed, as in water/ˈwɔːtər/. This phenomenon also applies when words are linked together in a full prosodic unit, as in the sentence *What is this?* [wɔtɪz ˈðɪs].

In this video Rachel gives examples of the pronunciation of [ɾ].
The third r-consonant is the retroflex\[^{[\text{rɛtrəfleks}]}\] approximant \[^{[\text{ɾ}]}\], responsible for the so-called rhotic\[^{[\text{roʊtɪk}]}\] accent or r-coloring. The rhotic accent is one of the main differences in pronunciation between British and American English. For more information on the rhotic accent, see \[^{[\text{Göm12a}]}\] and \[^{[\text{Wik11}]}\] and the references therein. Sound \[^{[\text{ɾ}]}\] is pronounced as follows:

- The retroflex approximant \[^{[\text{ɾ}]}\] always occurs after a vowel sound. Below are described the precise circumstances where that takes place.
- Therefore, the initial position of the tongue is that of the preceding vowel.
- After the vowel is pronounced, the tongue takes a neutral position in the mouth both and the tip of the tongue is curled back towards the alveolar ridge to produce the sound \[^{[\text{ɾ}]}\].

The rhotic accent can be found associated with the following sounds:

- Long vowels \[^{[\text{ɑː]}, [ɔː], [ɜː]}\], as in hard\[^{[\text{hɑːɹd}]}\], borne\[^{[\text{bɔːrn}]}\], and hurt\[^{[\text{hɜːt}]}\], respectively.
- After the short sound schwa \[^{[\text{ə}]}\] in the comparative endings, as in later\[^{[\text{'leɪtər}]}\], or taller\[^{[\text{ˈtɔːlər}]}\].
- A word-final vowel followed by \[^{[\text{ɾ}]}\] is pronounced as vowel followed by \[^{[\text{ɾ}]}\], as in here\[^{[\text{hɪər]}]}\], and hair\[^{[\text{hɛər]}]}\).
- The combination \[^{[\text{ɹʊ]}\], as in cure\[^{[\text{kʌr}]}\], or pure\[^{[\text{pɜːr}]}\].
- After the short sound \[^{[\text{ʊ}]}\], as in poor\[^{[\text{pʊər}]}\], moor\[^{[\text{mʊər}]}\], or boor\[^{[\text{buər}]}\].

Furthermore, rhotic accent is produced according to the following circumstances (here we have followed \[^{[\text{Göm12a}]}\] and the references therein).
• There is rhotic accent when a word is pronounced in isolation or at the end of a prosodic break. For example, *It was very hard* [{\textit{I}tw\textit{a}z\text{'v\text{"e}ni\text{h\text{a}}}z\text{]}\text{].}

• The rhotic accent is lost when the letter r does not belong to the same syllable. Compare *water* [{\textit{w\text{\ae}ra\text{}}}z\text{] and *watery* [{\textit{w\text{\ae}ra\text{}}}\text{]}\text{].

• If within a prosodic unit the last syllable of a word ends by [\text{/}] and the next word begins by a vowel, then the rhotic consonant is substituted by [\text{i}] or [\text{r}], depending on the particular accent. For example, the sentence *That water is cold* is pronounced as [{\textit{ð\æt\text{\text{w}}\text{\ae}\text{t\ae}n\text{z koold}]}; notice the change from [\text{/}] to [\text{i}] in *water*. Also, it could be pronounced as [{\textit{ð\æt\text{\text{w}}\text{\ae}\text{t\ae}n\text{z koold}]}. In the next video Rachel explains the pronunciation of [\text{i}]

![Video explaining the pronunciation of [\text{i}]](image)

Figure 40: Video explaining the pronunciation of [\text{i}].

Common spellings associated with r-consonants are *r*, *rr*, and *wr*, as in *run* [{\textit{r\u{u}n}]\text{, sorry*} [{\textit{s\text{"o}}}\text{]}\text{, and *wrap* [{\textit{wr\ae}p}]\text{, lyrics*} [{\textit{l\text{"i}r\text{"i}}ks}]\text{, parryl*} [{\textit{p\text{"a}r\text{"i}}}l], *water* [{\textit{\text{w\text{\ae}ra\text{}}}z\text{], or *hard* [{\textit{h\text{"a}}}\text{d}].

Some native Spanish speaker experience difficulties when pronouncing sound [\text{i}], specially in placing the tongue at the right position. Sometimes, Spanish trilled [\text{r}] is pronounced instead of the correct sound [\text{i}].

Since the usage rules for the r-consonants are quite clear and complementary, most dictionaries transcribe the three [\text{r}]-sounds as simply [\text{r}] (even though, IPA symbol [\text{r}] is used to represent a coronal trill). In this article we have followed that transcription practice except for the description of allophones.

4.5.2 Lateral sounds [\text{[l]} and [\text{l\text{\u{g}}}]

In lateral sounds, as already mentioned, the airstream is directed to the sides of the mouth owing to the action of the tongue. How does the tongue achieve that? The tongue possesses
muscles that allows it to broaden and narrow its body; check again Figure 5. By narrowing its body and putting its sides down and in the tongue is able to achieve it. Both allophones [l] and [l\textsuperscript{G}] are alveolar lateral approximants. The second sound is a velarized sound, but the first is not. Allophone [l] is pronounced as follows:

- The tip of tongue leans against the alveolar ridge.
- At this point, the sides of the tongue are down and in, forming two passages at each side. Compare the tongue positions for lateral and central consonants in Figure 41. The Figure shows a transverse cross-section of mouth as viewed from the front: on the left, it is a lateral consonant; on the right, it is a central consonant.
- The airstream comes from the lungs and passes through the lateral passages. Because this sound is voiced, the vocal cords are in vibration.
- The lips are in a neutral position, with the lips slightly rounded.

Figure 41: Comparison of tongue positions in lateral and central consonants (after [CM08]).

From this description, it follows that sound [l] is a voiced alveolar lateral approximant.

Allophone [l\textsuperscript{G}] is a co-articulated consonant (see Section 4.1.5), that is, the sound has two articulations. The first articulation is exactly the same as allophone [l]. The second articulation is a velarization,\textsuperscript{[vi:leraɪˈzeɪʃn]} while the tip of the tongue is resting on the alveolar ridge, the back of the tongue raises and come close to the soft palate (but without touching it). The velarization is indicated in IPA by adding the letter \textsuperscript{G} as a superscript to the first sound (symbol [l] is also used). These two allophones [l] and [l\textsuperscript{G}] are also known as light [l] and dark [l], respectively.

In theory, light [l] is pronounced before vowels, as in leaf[liːf], and dark l is pronounced in syllable-final positions, as in feel[fiːl] or selfish[ˈselfɪʃ]. In practice, in American English most of the time only the dark [l] is heard, with the exception of formal speeches or some particular variants.

Phoneme /l/ is spelled l or ll, as in lot[lɒt] or tall[tɔːl]. In the video below Rachel explains both the light and dark l.
Most problems for native Spanish speakers arise from the pronunciation of [l] as sound [l], which is not velarized in Spanish.

### 4.6 Semivowels

Semivowels are sounds that are articulated as vowels —they are not produced by creating a complete or partial obstruction in the upper vocal tract—, but function as consonants. Before going into the precise description of the semivowel sounds, we will elaborate further on the phonological features of semivowels in order to understand this apparent paradox. In English the structure of syllable consists of a initial consonant cluster, called onset (sometimes not present), followed by vocalic cluster (a single vowel, a diphthong or a triphthong, sometimes not present), followed by a final cluster called coda (sometimes not present). The vocalic cluster is called the nucleus of the syllable (see [Roa09] for more information). The term consonant cluster here is quite generic and may refer to up to four consonants or none. Therefore, the general structure of the English syllable is (C)+V+(C), where C stands for consonant and V for vowel cluster. Here we have several examples with different consonant structures:

1. V: *awe*[ɔ:], *owe*[ou]. Although rare, there are some words in English containing no consonants.
2. C+V: *Key*[kiː], *no*[nou], *lie*[lai], *true*[truː]. In these examples the coda is not present and the onset is composed of one consonant.
3. V+C: *Ease*[iːz], *ought*[ɔt], *ash*[æʃ], *aim*[aim]. The onset is not present and again the consonants have only one sound.
4. C+V+C (simple examples): *Caught*[kɔt], *rain*[ram], *man*[mæn], *feel*[fiːl]. In these examples both the onset and coda are present in the syllable.

---

4 Plural of nucleus: nuclei/*nuːkliaʊ/*
5. C+V+C (complex examples): **Play**[pʰleɪ], **ask**[æsk], **stop**[staːp], **spray**[sprem], **fifths**[fɪfθs], **strain**[stren], **stops**[staːps], **strands**[stændz], **prompts**[pʰprɒmpt]

6. Examples including semivowels:

(a) Semivowel+V: **you**[juː], **woe**[wʊ], **yes**[jes].

(b) C+semivowel+V: **few**[fjuː], **pjuː**, **spew**[spjuː].

(c) Semivowel+V+C: **year**[jɪr], **your**[jʊə], **were**[wɜːr], **wore**[wɜːr], **one**[wʌn], **what**[wɑːt].

(d) C+semivowel+V+C: **pure**[pʰjuː], **thwack**[θwek], **cure**[kʰjuː], **sweet**[swiːt], **swear**[swɛə].

(e) In the middle of longer words: **popular**[pʰpɒlər], **quietly**[kʰwɪtli], **beauty**[bjuːtɪ], **earthquake**[ɜːθkwɑːk], **circular**[sɜːkjʊlər].

After having examined the examples above, we conclude that semivowels in English are never part of the nucleus of a syllable. Hence, they act as consonants. However, as we will see next in detail, they behave as vowels in terms of articulation. The following is a list of the main features of semivowels:

1. **Glides**: Semivowels are always glides, that is, continuous transitions between two sounds, where the first sound is always the semivowel. The target sound is the second sound; the semivowel is a very short sound. See, for example, **you**[juː], **sweet**[swiːt], or **your**[jʊə].

2. **Consonant boundaries**: In all the examples listed above the semivowel appears as the last sound of the onset. Semivowels are never encountered in the middle of a consonant cluster or in the coda.

3. **Differences with diphthongs**: Semivowels are not diphthongs. While in diphthongs the main sound is the first, in semivowels the main sound is the second; compare the following pairs diphthong-semivowel [au]-[jʊ], [ə]-[jə], or [ɔ]-[wɜː]. While diphthongs always function as vowels and belong to the nuclei of syllables, semivowels never do.

### 4.6.1 Palatal semivowel [j]

This semivowel is pronounced in a similar way to vowel [iː], but semivowel [j] is very short. Another difference is that in the case of vowel [iː] the tongue does not touch the roof of the mouth, while in the case of semivowel [j] contact is made. In the articulation of [j] the front of the tongue is raised and then the sides of the tongue touch the hard palate. A groove in the middle of the tongue is formed over which the sound travels. The tip of the tongue are down, just behind the bottom front teeth. Notice that [j] is a central consonant as the air is directed through the centre of the mouth.

Sound [j] is usually spelled *y*, as in **year**[jɪr].
Figure 43: Video explaining the pronunciation of [j].

Semivowel [j] is sometimes replaced by palatal lateral approximant [ʎ] by native Spanish speakers. This also happens with affricate [ts]. In English [j] behaves as a consonant found at the boundary of the onset, whereas in Spanish it is considered as part of a diphthong. This might explain the difficulties in properly pronouncing this sound. Not only does the phonetic description of a sound matter to its pronunciation, but also its distribution and function.

4.6.2 Velar semivowel [w]

Semivowel is another co-articulated sound. It is articulated as a voiced approximant at the velum and at the same time is articulated with the lips. More precisely, this sound is termed as a **labio-velar approximant**[^1]. The articulation of [w] is carried out as follows:

- The lips are projected forward and take on a rounded position.
- The back of the tongue is pulled back and raised and comes close to the soft palate. This position is similar to that of vowel [u:], but the back of the tongue is closer to the soft palate.
- The vocal cords then vibrate and the lips are retracted to the position of the next sound in the word.
- Because the tongue has to move up towards the soft palate, the root of the tongue also moves and narrows a little bit the pharynx. This gives the sound a characteristic timbre.

Sound [w] is usually spelled *w* or *wh*, as in *way*[weɪ], or *where*[weə]. In the video below Rachel explains the production of this sound and gives examples.
There is an equivalent sound to [w] in Spanish, but it is not as labialized as in English. Furthermore, [w] in English appears in non-existing combinations in Spanish, as in would[wud], where [w] is followed by [u]. Some native Spanish speakers may pronounce would as [gud].

5 Further material on consonants

In this Section we will cover a few topics that were left out either because the material would otherwise be lengthy, or because discussions would be too technical for the intended reader at that moment, or because of expository reasons. In particular, we favour the idea of presenting consonants in detail and at the end show a general, encompassing classification.

5.1 Other pronunciation problems of Spanish language

Throughout this article we showed the reader the main pronunciation problems associated with each English consonant that native Spanish speakers may be confronted with. However, there are a couple of pronunciation problems not associated to any consonant in particular and that are still worth mentioning:

- **Final consonant clusters.** English has a larger and more diverse inventory of consonant clusters than Spanish does. English may have up to four consonants in a row in word-initial or up to three in word-final, as in texts[teksts], asks[æskz], felt[felt], stress[stres], skunk[skæn]. This consonant complexity, along with an imprecise manner of articulation, specially in phonemes /s/ and /t/, make the pronunciation of many native Spanish speakers obscure.

- **[d] versus [ð].** Phoneme /d/ is realized in Spanish as a dental stop [d] or a dental fricative ð. The latter is always pronounced when letter d is between two vowels;
the former, in the remaining cases. In other words, letter d changes its pronunciation depending on its position; it is a positional variant in Spanish. Native Spanish speakers not aware of this situation tend to apply this rule to English pronunciation. The result is that word this is mispronounced [ðIs] (or even worse, [dIs]) instead of the correct version [ðIs].

- [b] versus [β]. In Spanish letter b at intervocalic position is pronounced as a bilabial fricative [β]. If not explained properly, a native Spanish speaker may pronounce lobby as ['lɑːβi] instead of ['lɑːbi]. Most Spaniards could not tell the difference between [b] and [β].

### 5.2 Syllabic consonants

In Section 4.6 we described the structure of the English consonant as (C)+V+(C), where C stands for consonant cluster, V for vocalic cluster, and parenthesis are used to indicate optional elements. However, that description, strictly speaking, was not complete. In English it may happen that a syllable is composed of no vocalic cluster; of course, in that case something else must be part of the syllable. Syllabic consonants are consonants that form a syllable on its own. Syllabic consonants are never stressed and are always voiced. When a consonant acts as a syllable, it is marked by a central subscript, as in l or m; this is the IPA diacritic to denote syllabic consonants. In English syllabic consonants are [n], [m], and [l]. For example, the following words have syllabic consonants: seven[sevn], awful[ɔːfl], rhythm[ˈrɪðəm]. Unaware students may pronounce these words by inserting a schwa in between the two consonants to form a syllable with vocalic nucleus, as in even[i:vən], but that would still be incorrectly pronounced.

### 5.3 Classification of consonants

We have deferred a classification of consonants until the end of this article because we believe that presenting a complete classification to the student at beginning would be to some degree misleading. The student at that stage would not possess the knowledge and experience to understand what that classification is for. Such a categorization should be given much later, and that is the reason to introduce it at the end.

A consonant was defined as a sound produced by a partial or complete obstruction of the airstream (see Section 3.2). That definition, which refers to manner of articulation, can be refined by splitting the group of consonants into two broad categories, obstruents[ˈɒbstruənts] and sonorants[ˈsɒnərənts]. Actually, this classification is based on the degree of stricture[ˈstrɪktʃər], how narrow the gap is between the active articulator and the passive articulator at the narrowest point in the vocal tract [Col12]. The degree of stricture can be complete, with a complete blockage of the airstream, as in stops; complete approximation, with partial blockage resulting in frication, as in fricatives; and open approximation, with no frication, as in approximants and vowels. See [Man12] for more details on stricture. In obstruents, the sound is either stopped or interfered with in the vocal
tract. Obstruents are stops, fricatives, and affricates. In Figure 45 a complete classification of obstruents is given (sounds in the boxes are sibilant consonants; sounds to left to the bullet are voiced and to the right voiceless).

Figure 45: Classification of consonants: obstruents.

A sonorant is a consonant produced in a continuous manner, as opposed to stops, and without turbulence in the airstream. All sonorants in English are voiced (and in most of world’s languages). In Figure 46 the classification of sonorants is shown.

Figure 46: Classification of consonants: sonorants.

The two previous classifications provide a good example to illustrate the difference between phonology and phonetic features. A phoneme is an abstraction of sound features through which speakers and listeners extract meaning. A phoneme has a set of allophones associated; for example, phoneme /p/ has [pʰ], [pʷ], and [p] as its allophones. In Figures 45 and 46...
and the phonemic features are on the top and as we go down, the phonetic features start appearing.

6 Further material

6.1 Diaphragmatic breathing

Learning how to pronounce a second language may result in more demands on your voice. It is fundamental to be able how to do diaphragmatic breathing. The diaphragm is a thick, umbrella-shaped muscle located below the lungs. During inhalation the diaphragm opens up and enlarges the thoracic cavity, drawing the air into the lungs. See a clear illustration of how the diaphragm works at the website Voice and Speech Source. If diaphragm is used fully and efficiently, we will be able to talk for longer and, which is more important, in a more relaxed way. Learning how to do diaphragmatic breathing can be achieved through a well-chosen set of exercises. Check out the website Speech Therapy Information and Resources for reliable information.

6.2 Speech articulators

The website SPAN, the Speech Production and Articulation Knowledge Group at the University of Southern California, has a list of very instructive videos where we can watch the vocal tract of a young lady while she talks. Magnetic resonance technology has been used to highlight all main articulators, lips, jaw, tongue, soft palate; place and manner of articulation can be observed as her speech progresses. The best videos are Spontaneous speech, and Automatic contours (the contour of the tongue and the roof of the mouth). We strongly recommend the reader to watch them.

6.3 Tongue shapes

Epstein and Stone define four categories for tongue shapes that are relevant to speech sound production: back raising, front raising, continuous groove, and two-point displacement. In Figure the four tongue shapes are depicted. Back raising appears in vowels such as [ʌ] or [ø]. Front raising occur in sounds as [i:] or [j]; notice that these two sounds require a close approximation to the hard palate. Continuous groove is typical of sibilants, as already established, but also appears in vowel sounds, as in [æ]. Finally, two-point displacement is associated to lateral sounds, as [l]; in the Figure below we can notice the points where the tongue makes contact with the hard palate and how the sides of the tongue are put down and in. If we compare the groove created in [j] with the one created in [s], we will realize that in the latter case is deeper and longer. Therefore, in the case of [s] the jet of air seems is channel into the teeth at a high speed, which may explain the characteristic hissing sound of this consonant. Visualizing these tongue shapes may be of great help to the student of English.
Other authors have given alternative classification of tongue shapes; see [Wik12c]. For example, among English sibilants, two tongue shapes are worth mentioning.

- **Grooved tongue shapes**: They appear in sibilants sounds such as alveolar fricatives [s] and [z]. The groove runs down the medial axis of the tongue and tends to be deep and long. In general, these tongue shapes correspond to high-pitched, piercing sounds.

- **Palato-alveolar tongue shapes**: They occur in post-alveolar sounds such as [ʃ] and [ʒ]). The tongue is “domed”, that is, it adopts a convex and moderately palatalized position.

I exhort pronunciation teachers to include these kind of information in their courses, specially in advanced ones. The more exact and vivid the description of a sound is, the better and the faster it will be learnt.

**References**


