The Sounds of Language

When we listen to spoken language, we ultimately ‘know’ what we hear to be a series of discrete individual sound segments. Yet physically (= acoustically) sounds are not discrete but continuous propagating waves, which travel through the air somewhat like the ripples you see on a lake. When these sound waves reach our ear, they cause our eardrums to vibrate, ultimately leading to auditory ‘perception’ of some sort.

If the discrete representations of sound needed to describe our linguistic competence are not part of the acoustic spectrum, then what (where) are they? It would seem that they must internal to our minds, namely, objects in human mental representation. Thus, the systems underlying speech (production and perception) are able to map between certain acoustic events and discrete mental representations of linguistic sounds.

(Note: Acoustics is a branch of physics dealing with sound whereas acoustic phonetics is a branch of linguistics that deals with the acoustic properties of speech and speech perception. In this class, we will not be concerned so much with acoustics or acoustic phonetics, but rather with the way in which linguistic sounds are represented as part of our linguistic competence; i.e., in the mental grammars of humans.)

One of our tasks, then, is to understand the nature of the mental representation of sounds, for example, such as the three discrete sound segments in a word like cat.
If I produce the word *cat*, this speech event gives rise to an acoustic continuum. However, as part of their linguistic competence, every speaker of English ‘knows’ that this word consists of a sequence of three discrete sounds. For example, we know that if any one of these sound segments is replaced by some other, the result can be a completely different word; for example, *cat* ≠ *mat*, *cat* ≠ *cot*, *cat* ≠ *can*. (Sometimes the result may not be an existing word of English, what we might call a *nonsense word*, e.g., *cat* ≠ *cang*, as opposed to an *impossible-word* which is a sound or sequence of sounds that violates one or more of the rules of English sound patterns, e.g., *cat* ≠ *ngat*).

At this point, we pause and take note that mental representations of linguistic sounds **do not** correspond in any direct way to the letters/symbols of a writing (orthographic) system! This is evident by the fact that many languages of the world don’t even have writing systems. We can also easily see how very different these notions are by considering English, in which the relation between sound and spelling is notoriously unsystematic.

Beware of **heard**, a dreadful **word**
That looks like **beard** and sounds like **bird**.
And **dead**: it’s said like **bed**, not **bead**;
For goodness sake, don’t call it **deed**!
Watch out for **mean** and **great** and **threat**.
(They rhyme with **suite** and **straight** and **debt**.)
A **moth** is not **moth** in **mother**,
Nor **both** in **bother**, **broth** in **brother**.
Same spelling, different pronunciation:

\textit{cough} / \textit{tough} / \textit{bough} / \textit{through} / \textit{though}

Same pronunciation, different spelling:

\textit{see} / \textit{sea} / \textit{secret} / \textit{seize} / \textit{Caesar} / \textit{scenic} / \textit{siege} / \textit{ceiling} / \\
\textit{cedar} / \textit{cease} / \textit{juicy} / \textit{glossy} / \textit{sexy}

Clearly, mental representations of linguistic sounds (part of our knowledge of language) cannot be this carefree. And, as it turns out, even having a single symbol per sound does not suffice. We will not deal with this issue right away, but return to it when we discuss phonology later on. For the moment, however, it will be helpful to at least have a more systematic and precise way to represent the sounds occurring in English or any other language. One such system is called the International Phonetic Alphabet (IPA). Here, we will consider only those IPA symbols needed for the description of sounds (consonants and vowels) occurring in English:
Consonants

\[
\begin{array}{llll}
\text{p}^h & \text{pill} & \text{t}^h & \text{till} & \text{k}^h & \text{kill} \\
\text{p} & \text{spill} & \text{t} & \text{still} & \text{k} & \text{skill} \\
\text{b} & \text{bill} & \text{d} & \text{dill} & \text{g} & \text{gill} \\
\text{m} & \text{mill} & \text{n} & \text{nil} & \text{ŋ} & \text{ring} \\
\text{f} & \text{feel} & \text{θ} & \text{thin} & \text{s} & \text{seal} \\
\text{v} & \text{veal} & \text{ð} & \text{then} & \text{z} & \text{zeal} \\
\text{č} & \text{chill} & \text{ɾ} & \text{rake} & \text{š} & \text{shell} \\
\text{j} & \text{juice} & \text{l} & \text{lake} & \text{ž} & \text{measure} \\
\text{y} & \text{you} & \text{h} & \text{heal} & \text{w} & \text{witch} \\
\end{array}
\]

Vowels

\[
\begin{array}{llll}
\text{i} & \text{beet} & \text{u} & \text{boot} \\
\text{ɪ} & \text{bit} & \text{U} & \text{foot} \\
\text{e} & \text{bait} & \text{ə} & \text{sofa} \\
\text{ɛ} & \text{bet} & \text{ʌ} & \text{but} \\
\text{æ} & \text{bat} & \text{a} & \text{pot} \\
\end{array}
\]

(There are also nasal vowels in English, such as in the word *bean*; we will discuss nasal vowels and introduce symbols for them later.)

Diphthongs

\[
\begin{array}{llll}
\text{ay} & \text{bite} & \text{ɔy} & \text{boy} \\
\text{aw} & \text{bout} \\
\end{array}
\]
Standard English Spelling | IPA
---|---
though | ðo
thought | ðɔt
rough | rʌf
through | ðru
would | wʊd
back | bæk
bake | bek
ring | rɪŋ

**Articulatory Phonetics**

To better understand the nature of linguistic sounds and how they are represented in mental grammar, it is useful to understand how they are produced. Articulatory phonetics provides a framework to do this.

**Airstream Mechanisms**

Most speech sounds in human language are produced by pushing air from the lungs out of the body through the mouth and sometimes the nose as well. Since lung air is used, these sounds are called **pulmonic**; since the air is pushed outwards, they are called **egressive**; all of the sounds used in English or in Japanese are pulmonic egressive.
Voiced and Voiceless Sounds

In speech, air is pushed up from the lungs through the **trachea**. In doing so, it passes through the **glottis**, a small opening between the **vocal cords**. If the vocal cords are held apart, the flow of air is able to pass freely into the **supraglottal cavities** (the parts of the vocal tract above the glottis). Sounds produced in this way are called **voiceless sounds**. The sounds represented by the IPA symbols /p/, /t/, /k/, /s/, /f/, /θ/, /ʃ/ and /χ/ are all voiceless.

On the other hand, it is possible for the vocal cords to be held close together, forming a constriction at the glottis. As air forces its way through this constriction, it causes the vocal cords to vibrate. The resulting sounds are called **voiced sounds**. The sounds represented by the IPA symbols /b/, /d/, /g/, /z/, /v/, /ð/, /ʒ/ and /ɣ/ are all voiced sounds.

Oral and Nasal Sounds

Some sounds, called **nasal sounds**, are produced with the **velum** lowered, allowing air to flow freely through both the **oral cavity** and the **nasal cavity**. Other sounds are produced with the velum raised and pressed against the back wall of the **pharynx**. This prevents the passage of air through the nasal cavity, allowing it only to pass through the oral cavity. Sounds produced in this way are called **oral sounds**. English has three nasal consonants, represented by the IPA symbols /m/, /n/, and /ŋ/; all other English consonants are oral. (We will discuss oral and nasal vowels later.)
Places of Articulation

In addition to the voiced/voiceless or oral/nasal category distinctions, further distinctions can be articulated by altering the shape of the oral cavity by moving the lips and/or the tongue. In describing these different categories of sounds, we need to refer to the articulator (lips or tongue) and the place of articulation (lips, teeth, alveolar ridge, hard palate, or velum).

(1) Labial Sounds: Bilabial and Labiodental

The sounds represented by /p/, /b/, and /m/, called bilabial sounds, are articulated by bringing both lips together.

The sounds represented by /f/ and /v/, called labiodental sounds, are articulated by touching the bottom lip to the upper teeth.

(2) Interdental Sounds.

The sounds represented by /θ/ and /ð/, as in the words thin = /θɪn/ and then = /ðɛn/, are articulated by inserting the tip of the tongue between the upper and lower teeth. These sounds are called interdental sounds.

(3) Coronal Sounds: Alveolar and Palatal

The sounds represented by /t/, /d/, /n/, /s/, and /z/ are articulated by raising the tongue to the alveolar ridge. These sounds are called alveolar sounds.
The sounds represented by /ʃ/ and /ʒ/ are articulated by raising the front part of the tongue to a point on the hard palate just behind the alveolar ridge. These sounds are called palatal sounds.

(4) Velar Sounds

The sounds represented by /k/, /ɡ/, and /ŋ/ are articulated by raising the back of the tongue to the velum (also called the soft palate). These sounds are called velar sounds.

Manners of Articulation

We have described a number of properties of phonetic articulation that allow us to categorize linguistic sounds into several overlapping classes in terms of voiced/voiceless, oral/nasal distinctions as well as place of articulation. However, we have not yet said enough to describe all of the necessary distinctions between sounds in English. How is /t/, the first sound of the word tell, different from /s/, the first sound of the word sell, for example? They are both voiceless oral alveolar sounds! In order to do this we need to also talk about sounds in terms what is called their manner of articulation:
(1) **Stops and Continuants**

Pulmonic air entering the oral cavity may be (i) stopped completely, (ii) partially obstructed, or (iii) allowed to flow out of the mouth unobstructed. Certain sounds are produced by stopping the airflow completely in the oral cavity for a brief period before it is released. Such sounds are called stops. The sounds represented by /p/, /b/, /m/, /t/, /d/, /n/, /k/, /g/ and /η/ are all stops. All other sounds are called continuants, since the stream of air is allowed to continue through the oral cavity without being completely interrupted.

The sounds represented by /m/, /n/, and /η/ are called nasal stops. In their production, the velum is lowered so that air is allowed to flow through the nasal cavity, although it is stopped completely in the oral cavity. All other stops are called oral stops and are produced with the velum raised and airflow stopped in the oral cavity.

/p/ = voiceless oral bilabial stop  
/b/ = voiced oral bilabial stop  
/m/ = voiced nasal bilabial stop  
/t/ = voiceless oral alveolar stop  
/d/ = voiced oral alveolar stop  
/n/ = voiced nasal alveolar stop  
/k/ = voiceless oral velar stop  
/g/ = voiced oral velar stop  
/η/ = voiced nasal velar stop
(2) Fricatives

In the production of some sounds, the flow of air through the oral cavity is partially obstructed though not completely stopped. When the degree of obstruction in the oral cavity is such that air is forced to push its way through a very narrow passage, friction or turbulence is created. The resulting sounds are referred to as fricatives. The sounds represented by /s/, /z/, /f/, /v/, /θ/, /ð/, /š/, and /ž/ are fricatives. All fricative sounds are categorized as continuants, since the airflow is not completely stopped in the oral cavity.

/f/ = voiceless oral labiodental fricative
/v/ = voiced oral labiodental fricative
/θ/ = voiceless oral interdental fricative
/ð/ = voiced oral interdental fricative
/s/ = voiceless oral alveolar fricative
/z/ = voiced oral alveolar fricative
/š/ = voiceless oral palatal fricative
/ž/ = voiced oral palatal fricative

(3) Affricates

Some sounds are produced as stops followed immediately by a slow release into a fricative. Such sounds are called affricates and include the sounds represented by /č/ and /ǰ/. For example, in the production of /č/, articulation starts out the same as in the production of the stop /t/ but quickly moves into the position for production of the fricative /š/. Thus an affricate is in many ways
like the sequence of a stop followed by a fricative, raising the question of whether an affricate is one sound or two. There are some reasons, however, to treat affricates as single consonantal sounds. This raises the question whether to classify affricates as stops or continuants. Affricates are often said to be stops, since they start with complete closure of airflow from the oral cavity.

(4) Liquids

In the production of the sounds /l/ and /r/ there is some degree of obstruction in the oral cavity, but not enough to give rise to any real friction or turbulence. These sounds are called liquids. /l/ is sometimes called a lateral sound. It is produced by raising the tip of the tongue to the alveolar ridge or to the upper teeth, but with the sides of the tongue lowered so as to allow the flow of air to “escape” out of the mouth. The sound represented by /r/ is produced in several different ways, and varies even amongst speakers of the same language. In English, /r/ is often produced curling the tip of the tongue back behind the alveolar ridge, in which case it is called a retroflex sound. In English, both /r/ and /l/ are usually voiced, but some languages have a voiceless /l/.

(5) Glides

The sounds /y/, /w/, /h/ are called glides. In the production of glides, there is usually a rapid movement, or gliding, of the tongue from the place of articulation for the glide to the position for the production of the following vowel. In this sense glides are “transition sounds.” Also, since glides usually involve little
obstruction in the oral cavity, they are considered to be more **sonorous** than consonant, and are sometimes even referred to as semi-vowels.

The **palatal glide** /y/, as in the first sound of the words *you, yell*, etc., is produced by raising the back of the tongue in the direction of the hard palate, but without causing any real obstruction to the airflow through the oral cavity. It then moves rapidly to the position for the production of the following vowel sound. The glide /w/, as in the first sound of the words *witch, whale*, etc., is produced by simultaneously rounding the lips and raising the tongue in the direction of the velum. It is thus called a **labio-velar glide**. Both /y/ and /w/ are voiced sounds, but some languages also have a voiceless labio-velar glide, which is written as /ʍ/ (in some dialects of English, *witch* and *which* differ in pronunciation, the first sound of the former being voiced /w/ and the first sound of the latter being voiceless /ʍ/).

(6) **Glottal Sounds**

The first sounds in the words *house, who*, etc., represented by /h/, although voiceless, are produced with constriction at the glottis causing some turbulence. It is this noise of the turbulent air passing through the open glottis that is heard as /h/. Thus /h/ is often referred to as a **voiceless glottal fricative**. If the air is stopped completely at the glottis, by tightly closing the vocal folds, the result is a **glottal stop**, written as /ʔ/, which is also voiceless. /h/ and /ʔ/ are sometimes classified as glides since
they are produced without any obstruction in the oral cavity.

**Major Classes of Linguistic Sounds**

Very broadly, we can divide linguistic sounds into two major classes: **consonants** and **vowels**. Consonants are produced with some degree of obstruction of airflow in the oral cavity; this is the case with stops (oral and nasal), fricatives, and liquids. Vowels, on the other hand, are produced without any real obstruction in the oral cavity. Vowels also typically serve as the core (nucleus) of syllables. This helps us distinguish vowels from glides and glottal sounds, which are otherwise difficult to classify; they are unlike consonants since they involve little or no obstruction in the oral cavity; whereas they differ from vowels in that they cannot form the core of a syllable. On the other hand, liquids and nasals are similar to vowels in the sense that they can in principle (although they do so only sometimes) form the core of a syllable. Vowels, glides (including the glottal fricative /h/), nasals, and liquids form a class of sounds called **sonorants**. For the moment, these distinctions remain somewhat rough and it suffices to understand them at an intuitive level.

**Vowels**

Although vowels are more difficult to describe in articulatory terms than consonants, the sound quality of an individual vowel is determined by the overall size and shape of the oral cavity during its production. In effect, the oral cavity serves as an
“echo chamber” for air coming up from the lungs through the pharynx. By manipulating our tongues and lips we can alter the size and shape of this chamber, which leads to different sound qualities. The most important articulatory variables determining vowel quality are the following:

1. The vertical position of the tongue: high or low.
2. The horizontal position of the tongue: forward or back.
3. The position/shape of the lips: rounded or unrounded.

**Tongue Position**

Although, there is really a continuum of tongue positions, both along the vertical and horizontal dimensions, it is useful to basically talk about the following parameters in the description of vowels:

<table>
<thead>
<tr>
<th>front</th>
<th>central</th>
<th>back</th>
</tr>
</thead>
<tbody>
<tr>
<td>high</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>low</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The sounds represented by the symbols /i/, /ɪ/, /u/, and /U/, as in the English words *seat, sit, suit, and soot* respectively, are high vowels. /i/ and /ɪ/ are high-front vowels, whereas /u/ and /U/ are high-back vowels.
The sounds represented by the symbols /e/, /ɛ/, /o/, /ɔ/, and /ʌ/, as in the English words *bait, bet, boat, bought,* and *but,* respectively, are all mid vowels. /e/ and /ɛ/ are mid-front vowels whereas /o/ and /ɔ/ are mid-back vowels. /ʌ/ is a mid-central vowel. English also has an unstressed mid-central vowel /ə/, as in the second vowel in the word, *sofa.* This unstressed vowel goes by the name *schwa.*

The sounds represented by the symbols /æ/ and /a/., as in the English words *pat* and *pot,* are low vowels. /æ/ is a low-front vowel whereas /a/ is a low-back vowel.

**Lip Rounding**

The other major articulatory factor affecting vowel quality is the shape of the lips. Basically, the lips can either be rounded or unrounded. When the lips are rounded, or pursed, the shape of the oral cavity is changed, essentially becoming a bit longer, and the shape of the mouth is different as well. These changes have a significant effect on the way a vowel will sound. In English, the back vowels /u/, /U/, /o/, and /ɔ/ are rounded, whereas all other vowels are unrounded. English has no round front vowels, but this is not true of every language. French and Swedish, to give some examples, have both front and back rounded vowels. Also, although in many languages high-back vowels are rounded, in some languages, such as Japanese, there is a high-back vowel that is unrounded.
Nasal Vowels

Vowels, like consonants, can be nasalized by lowering the velum during their production. Instead of introducing a completely new symbol for each nasal vowel, the IPA utilizes what is called a diacritic symbol. Nasal vowels are indicated by the attaching the symbol /~/ to the vowel in question. Thus we pairs of oral and nasal vowels, such as /i/ and /i~/. We will discuss nasal vowels in English in more detail when we turn to phonology.

Diphthongs

Some so-called vowels are probably best described as a sequence of two sounds, a vowel immediately followed by a glide. These *vowel + glide* sequences are sometimes referred to as *diphthongs*. Those occurring in English are: /ay/, as in the word *bite*, /ɔy/, as in *boy*, and /aw/ as in *bout*. 