Phonetics, part 2: Families of speech sounds

The interaction of the various articulators and places of articulation described in your first handout yields a very large number of possible speech sounds (though far, far fewer than the number of possible distinct ideas human speakers might wish to express—the latter number probably being unboundedly large, along the lines discussed in class). Each human language offers a selection from this large number of identifiably different speech sounds. There is a small technical point here that’s probably still worth paying attention to: in principle, the number of sounds we can actually produce is probably unboundedly large as well. But the number of sounds we can distinguish is almost certainly finite. There’s not much point in being able to make a great many acoustically different versions of a certain class of sounds if the person you’re speaking to can only distinguish three subvarieties of that sound at most, because of perceptual limitations built into human neurophysiology. So for all practical purposes, we have to live with the fact that the size of our speech sound inventory is going to be finite. But, as we see whenever we look at a given human language, we actually need only a small fraction of that finite number to wind up with a completely sufficient store of phonetic elements.

There are certain key variables that come into play in determining what sounds a language actually makes available to speakers. The following will get us started: for any given sound,

- what airstream mechanisms is employed?
- what articulator is used (primarily; sometime two or even more are involved)?
- what part of the vocal tract is the target of the articulator, i.e., where is the contact made?
- how ‘tight’ is the contact between the primary articulator and the place of articulation—is it complete, or partial, or barely there at all?
- is the airflow entirely through the mouth, or does the nasal cavity play a role as well?

There are other questions that must be answered in many cases before we can be sure that the specification of the sound is complete. But those I’ve listed will give us a useful point of departure.

1 Consonants versus vowels: the major divide

If you pronounce the word *cat* and keep track of what’s happening in your vocal tract as you do so, you’ll most likely be struck by the impression that all is not equal amongst the three sounds you probably think of as making up the word. The *c* and *t* at the beginning and end of the word have almost no independent existence. Try saying them in complete isolation and you’ll realize that it’s almost impossible to identify what they ‘sound like’. The case is very different with the *a* sound that appears in the middle of the written form of the word. This sound is easy to articulate on its own, for as long as you’ve got breath in your lungs. Pronounce the word *cat* again and this time, pay attention to the specific question of how close the contact is between the tongue and some other part of the mouth through the whole word. You’ll notice that during the *k* and *t* parts of the word, the contact is pretty much total: no air can get out while the tongue is in the right position for making either of these sounds. The *a*, however, is pronounced with the tongue far from the roof of the mouth. It really doesn’t feel as though there’s any contact at all between the tongue and the rest of the vocal tract.
Pronounce the word one more time, trying to get a sense of just what the \( c \) and \( t \) consist of. You’ll probably realize, sooner or later, that these sounds don’t have much of an independent existence; they are really nothing more than, respectively, a way of holding the mouth before the \( a \) begins to be pronounced, and a way of holding the mouth as a means of shutting down the air flow to end the word. We hear both the \( c \) and \( t \) as, in effect, two different kinds of ‘coloration’ of the \( a \). The same thing will occur to you if we switch the positions of the two sounds as in the word \( tack \), which is, roughly speaking, just \( cat \) backwards. In a sense, you don’t hear the \( t \) and \( k \) sounds; what you hear is something distinctive about the beginning and about the end of the \( a \).

Things are bit different if you put your tongue tip in the right place for a \( t \) but instead apply enough force to the airstream from you lungs to open a passage between the tongue tip and the upper part of the mouth that it’s in contact with. In \( sack \), you can continue the initial hissing sound indefinitely. But here too, the contact is very close, compared with what you have in the case of \( a \). And if you move your tongue a little further back, and loosen the contact still more, you’ll find yourself pronouncing the work \( yak \). There are, it’s true, some other things going on there, involving the vocal cords, but the degree of closure is considerably less than what happens with \( s \).

Sounds such as the \( c,k,s \) just discussed, which involve a high degree of closure, are examples of what are standardly referred to as consonants; the \( a \) we’ve been looking at in the words mentioned above belongs to a group of sounds called vowels. It’s instructive to consider the source of these names. ‘Consonant’ is based on Latin-derived elements \( con \) ‘with’+\( sonant \) ‘sounded’, i.e., ‘sounded with’. The idea is that consonants are pronounced in conjunction with vowels, rather than on their own, along the lines I’ve pointed out earlier. \( Vowel \), on the other hand, is related to the Latin \( vox \) ‘voice’, which is also the term used to describe how the sound wave is affected when the vocal cords are vibrating as air passes through the glottis. You have a voice as long as you’re pronouncing vowels, but consonants need to have vowels alongside them or you don’t hear them. This is quite an oversimplification, but it’s basically correct in general.

All sounds are assigned to one or the other of these two major categories: consonants the sounds involving a high degree of closure, vowels involving much less closure. Within each of these categories, of course, there are significant differences. Some consonants, such as \( t \) and \( k \), involve complete closure, while others involve relatively little (the \( y \) of \( yak \) is a typical example). On the other hand, the initial sound of \( eat \) involves the tongue in much closer proximity to the roof of the mouth than the vowel in the first syllable of \( bother \). The divide is actually slightly arbitrary, but on the whole, it’s fairly clear what’s at stake.

From this point on, we’ll look at consonants and vowels separately. It’s true that consonants involve much more stricture than vowels do—much tighter closure, that is—but on the other hand, a very small difference in the way the tongue makes consonants will yield a far more easily recognizable difference in sound than a comparable difference in the production of vowels. We’ll see below just why this is.

## 2 Consonants

Consonants are subdivided into those which reflect very tight levels of closure (aka obstruants) and those which do not. The term sonorant is used for the latter, but be careful: vowels are also classified as sonorants, since, of course, they involve the lowest degree of closure possible. Strictly speaking, we are talking about a contrast within the consonants of obstruants on the one hand and sonorant consonants on the other. But usually the term sonorant is used on its own, to denote consonants in which there is a fairly free passage of the airstream from inside the vocal tract to the outside.

### 2.1 Obstruants

The \( obstru \) part of the name for this group is the same string of letters that shows up in the word \( obstruction \). The idea is the same: what makes obstruants obstruants is that they involve a significant blockage to the
airflow, resulting in acoustic noise, the result of turbulence. Put two surfaces in close contact and force some kind of substance between them, and the result will be eddys, vortexes, and a lot of complex, highly unstructured movement, which will translate into what strikes our ears as hard-edged, sharp sounds, often associated with a harsh acoustic quality. Within the class of obstruants are two main subcategories:

2.1.1 Stops
Stops are exhibit A amongst the consonants. They exemplify to the extreme limit the obstructive basis for the consonant/vowel distinction. At each point in the vocal tract, a closure which results in a complete blockage to the stream of air yields a very distinctive set of sounds, which are however almost completely dependent, as noted earlier, on the vowel to ‘carry’ the information about where that blockage was made. Stops are universal in human languages, and have been recorded at every major point of closure in the mouth.

2.1.2 Affricates
Affricates could be described as species of stop which loses its nerve at the end: instead of a very sharp, clean release of the closure, the tongue, starting in a tight closure at some point in the mouth, only gradually opens the obstruction it created, leading to a kind of harsh, hissing quality in the end phase of the affricate’s pronunciation. The first and final sound in English judge and church are different versions of the same affricate.

2.1.3 Fricatives
As the name tells you, fricatives are characterized by a marked hissing quality suggestive of the friction that results considerable airstream turbulence—one of their most obvious phonetic properties. Fricatives are sounds such as the s in see, the sh in rash, and the v in very. You’ll probably notice that that kind of harsh noisy quality is much less in the last case than in the first two; the reason is that v is voiced—the obstruction to the air that makes the v sound is accompanied by vibrating vocal cords. This voicing masks the distinctive hissing quality of fricatives to a very great extent.

2.2 Sonorants
Sonorants convey an impression to the ears of speakers very different kind from what I’ve described for the obstruants. The typical dictionary definition of sonorous is ‘having or producing a full, deep or rich sound’. ‘Richness’ and ‘smoothness’ are connected ideas here; it’s hard to picture a ‘rich’ sound which also merits the description ‘rough’ or ‘harsh’. Sonorants as a class carry just this quality of mellow richness. In this respect, they are the more vowel-like of the consonants.

If you’re not content with this way of identifying the subclass of sounds that are identified as sonorants, there’s another approach you might appreciate: sonorants are consonants which can be the core of syllables, just as vowels can. Recall that we’ve already talked about the need for vowels as the ‘carriers’ of information about stops; for this reason, we think of vowels as crucial to the pronunciation of words. But sonorants can do the same job. Consider the word falter, for example: the first syllable is built around the same vowel as in the word fall, but the second syllable, ter, actually consists of a t sound followed by an r-like sound that happens to take up a whole syllable, and which we write in IPA notation as [ɾ]. We frequently find syllables syllables in English which are built around sonorants—presumably because, like vowels, sonorants can be continued indefinitely, as long as the speaker has breath to spare. While this is also true of fricatives (you can draw the last sound in the word hush out as long as you like), the somewhat chaotic nature of the signal associated with fricatives (due, as noted above, to the turbulent nature of the airflow which produces sound waves corresponding to fricatives) makes these, and obsrue nts in general, much less than ideal as the basis for syllables.

There are several subclasses of sonorants.
2.2.1 Liquids
The term ‘liquid’ again suggests smooth flowing, even though, as we know, liquids don’t necessarily flow smoothly (in a white-water rapids, for example, there’s plenty of evidence for that statement). There are two principle subtypes.

Lateral sounds  Laterals are [l]-like sounds, and there is quite a variety of these in the world’s languages. Laterals all have in common the fact that regardless of where the tongue makes contact with the upper part of the mouth, the sides of the tongue are tilted downward, allowing air to flow along them and out through the mouth (hence the name lateral). The characteristic quality that we associate with l-like sounds is a direct result of this mode of contact between the tongue and the its place of contact in making these sounds.

R-like sounds  Probably the hardest sounds of all to describe are members of the phonetic family which includes the first sound in the word right. One of the difficulties is that, unlike almost all other classes of sound, there is no easily statable common factor in the physical production of the members of this group—the commonalities are actually acoustic in nature, rather than articulatory; that is, they have to do with the properties of the sound wave itself, and the effect on our perception, as opposed to the anatomical actions carried out in the vocal tract to generate the r-sounds. What this amounts to is a kind of common perceptual quality amongs a group of sounds which are typically made in physically very different ways. Some of these sounds are produced with the front part of the tongue, in contact with the (post-)alveolar portion of the roof of the mouth; others are produced at the opposite end of of the oral chamber, in the uvular region. There are r-sounds which involve no more than a quick tap upwards, while others involve a complex trilling action, possibly at the front of the mouth, but also, in some languages, in the very back. English r-sounds have a very different anatomical structure from those in Italian, or Russian, and capturing these differences is one of the most challenging tasks facing descriptive phoneticians.

2.2.2 Nasals
Strange though it seems, nasals—which are clearly members of the sonorant family by any of the usual criteria of classification—are also stops. How can this be? The answer is the somewhat idiosyncratic usage of the term ‘stop’ in phonetics, whereby a sound which can be carried on for as long as the speaker has breath to spare is still a stop as long as there is a complete blockage in the oral part of the vocal tract. Nasal elements, such as the final sounds in the words ram, ran, rang, are stops in this sense: there is a total obstruction of air flow through the mouth at three different point respectively—but the velar flap which connects the oral chamber to the nasal cavity is open, and air flows out through the latter. Try making these sounds on their own, and then pinching your nose—you’ll find that whichever nasal you’re experimenting with stops dead in its tracks. It’s famously true that people with bad colds can’t make nasal sounds properly. What we typically find is that there is a significant parallel between the nasals in a given language and the non-nasal stops; this pattern is clear for English, as we’ll see in the next handout. The rule isn’t ironclad, but it seems to hold up fairly well in the majority of cases.

2.2.3 Semivowels
The term semivowel is a hint about the nature of these sounds (a good deal more so than the alternative name, glide, which is sometimes used). In English, we have two such elements: the initial sounds in yet and wet respectively. If you pronounce these words with careful attention to their very first phonetic components, you’ll see that there is only minimal obstruction to the air steam offered by the tongue. In particular, the highest point on the tongue in each of these two cases does not actually make contact with the roof of the mouth. There is, essentially, unimpeded air flow over the tongue and out the mouth, in a way very reminiscent of garden variety vowels. This impression will be reinforced if you say the following pair of words together in fairly rapid succession: eon (as in geological era)/yon (as in fake-archaic English yon fair village). What you’ll observe is that, in terms of the position of the tongue in the mouth, there is no
difference between the initial sounds. The sole difference is that the high-frequency first sound in *eon* takes up a whole syllable, while that in *yon* takes up no more phonetic time than the first sound in *gone*. In effect, the syllable-length version of this sound is a vowel, while the stop-length version is a consonant. Thus, semivowels in effect stake out the far frontier between consonants and vowels, where the phonetic difference between the two is reduced to the minimum.

3 Vowels

We now come to the vowels proper. Vowels are a good deal harder to describe than consonants, for the same reason that they allow a freer, more sonorous airflow than consonants: minimal obstruction corresponds to a minimal specificity of location in the mouth. The greater the degree of obstruction, the more exact the placement of the articulator; conversely, the more openness, the more approximate the placement. As the highest point in the tongue becomes lower and lower, the less easy it is to identify it as being in a particular location; and the more variation in the sound itself we expect. While it’s always true that a particular configuration of the vocal tract produces a particular acoustic signature for the sounds we hear, our ability to discriminate amongst such sounds becomes considerably less as the tongue’s connection to some part of the roof of the mouth becomes more remote.

For this reason, descriptions of vowels are far less detailed than descriptions of consonants, and have to be taken more as general guidelines than as firm specifications. For the most part, vowel descriptions are given in terms of a three × three grid setup, with (high,mid,low) and (front,mid,back) forming all nine possible combinations. This is a bit idealized as a descriptive framework, implying as it does that there is a single ‘front’ location, in which high, mid and low tongue positions are possible giving rise to vowels which differ only in tongue height. But the truth is that some front vowels are a good deal further front than others, and similarly with back vowels.

There are other important possibilities of variation which have nothing directly to do with tongue position. An important independent distinction sorts vowels into two classes: tense and lax vowels, with the distinction reflecting fairly literally a contrast in muscular tension in the vocal tract. As it happens, tense vs. lax vowels do contrast in terms of height, but there’s good reason to believe this difference is secondary, and that the state of the vocal tract is the key contrast. We’ll see just how the tense/lax distinction finds expression in English when we survey the vowel system of American English in detail in a later handout.

4 What else is going on?

What I’ve described in this handout is just the major families of speech sounds, based on manner and position of articulation, but there is a lot more going on than just this. A whole range of so-called secondary articulations is possible with members of a given class of sounds. A given sound may be accompanied by nasalization (opening of the nasal cavity), extra length, lip-rounding, voicing (full or partial) and many other added physical accompaniments that change the nature of the perceived speech signal, often in predictable ways. These are quite important, but I’m going to postpone discussion of secondary (and tertiary) phonetic properties till we get to the phonology discussion. In the next handouts, I’ll present the sounds of American English consonants and vowels, with some discussion of their specific articulation and the IPA symbols that are used for them.