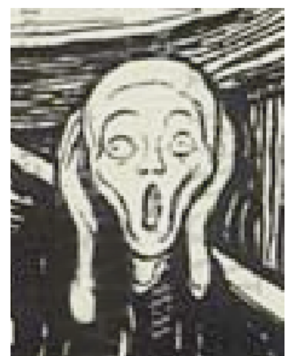
Chapter 1 – Phonetics and phonology: understanding the sounds of speech

Robert Kirchner, University of Alberta

Introduction

In most fields of study, language is thought of principally in terms of the written word, for it is in this form that we usually make permanent records of important ideas. Relatively little attention is spared for something as fleeting and unremarkable as spoken conversation. In linguistics, however, speech, rather than writing, is regarded as more central to human language, for several reasons. First, humans have probably used spoken languages for 100,000 years, perhaps longer. Writing is a relatively recent development, only a few thousand years old. Even today, most of the world's 5,000 or so languages have no established writing system. But there is no society which communicates just by writing, without a spoken language. Furthermore, children learn to speak long before they learn to read and write; indeed, learning of spoken language takes place without formal instruction.

But does ordinary speech really warrant scientific attention? Although we generally take the processes of speech production and recognition for granted, they involve a range of surprisingly intricate mental abilities – part of the knowledge we have of the language(s) we speak. The words that we wish to express seem to emerge inexplicably from our mouths, as soundwaves. These soundwaves then hit the hearer's ear, sending auditory signals to the brain, which are interpreted – again, seemingly magic – as the words intended by the speaker. What kind of mental system might underlie this capacity to produce and recognize speech? Which aspects of this system appear to be common to all humans, and which aspects vary from language to language? And what exactly goes on in the mouth and throat to produce speech? These sorts of questions are the domain of **phonetics** and **phonology** (both from the Greek root *phon-* 'sound'), the two subfields of linguistics concerned with speech sounds. In the remainder of this chapter, we examine some basic observations, terminology, and techniques of analysis used by phoneticians and phonologists to address these questions.


The Scream (detail), by Edvard Munch

Phonetics, phonology – what's the difference?

Traditionally, **phonetics** deals with measurable, physical properties of speech sounds themselves, i.e. precisely how the mouth produces certain sounds, and the characteristics of the resulting soundwaves; while **phonology** investigates the mental system for representing and processing speech sounds within particular languages. In recent years, however, the two fields have increasingly overlapped in scope. For our purposes, the important point is that linguists (whether they're called phoneticians or phonologists) have accumulated some basic observations about how the speech systems of human language 'work,' and these principles have a good deal to do with the physical properties of the speech sounds in question.

I. Forget letters, we're talking sounds

Sound energy is disturbance of air molecules: the disturbance radiates outward from its source, in waves of fluctuating air pressure ('soundwaves'), like ripples from a stone dropped in a pond. When we speak of an individual **sound** of speech, however, we mean something more specific: a portion of the speech in which the sound energy (and the configuration of the mouth to produce that sound energy) remains relatively stable. In the word *so*, for example, the sound energy changes, from a hissing sound at the beginning (with the mouth relatively closed) to a more open, singable sound at the end. But within each of these two portions of the word – the hissing sound of the *s,* and the singable sound of the *o*, there is relatively stability. We can therefore say that *so* is composed of two distinct sounds. Indeed, this decomposition of words into individual speech sounds is reflected to some extent in our writing system, for we spell *so* with two letters.

Nevertheless, it is important to bear in mind, throughout this chapter, that we are interested in the sounds which make up words, not the letters with which they are spelled. The word *fought*, for example, has six letters, but only three sounds: the *f,* followed by a single vowel sound (written with two letters, *ou,* in this word), and the final *t*. The *gh* is, of course, 'silent'; it is not part of the word's sounds, so we disregard it.

In fact, for the purpose of representing sounds, the English spelling system is quite unreliable – as generations of schoolchildren, struggling to memorize English spellings, can appreciate. The letter *c*, for instance, is pronounced like *s* in some words (e.g. *cell*), and *k* in others (e.g. *call*). Similarly, *o* corresponds to one vowel sound in *Robert* and a different one in *robe*.

The inadequacies become even more obvious if we try to transcribe (write down) the words of other languages – as linguists must do. The language might have no established writing system, or it may have sounds which don't occur in English. We might invent our own way of transcribing such sounds, using the closest-sounding letters of English. But how is a Russian linguist going to understand our English-based transcriptions, if she is not fluent in English? And how are we to understand this Russian's transcriptions of an unusual Kurdish dialect, written in the Russian (Cyrillic) alphabet, if we are not fluent in Russian? Linguists need an internationally agreed-upon system of transcription, in which the symbols correspond straightforwardly to sounds, and in which there are enough symbols to represent all the sounds of the world's languages.

This system is called the **International Phonetic Alphabet (IPA)**, first developed in 1886 and since modified in light of subsequent linguistic discoveries. For your interest, the full chart of IPA symbols appears at the end of this chapter. For present purposes however, we'll focus on the symbols needed for the basic sounds of North American English, adding other symbols as needed.

A. Consonants. If your first language is not English, and you are not sure how to pronounce any of the example words in Table 1, check with a native English-speaker.

Table 1: IPA symbols for the basic consonant sounds of North American English

IPA symbol	Example words	IPA symbol	Example words
p	**p**at, hi**pp**y, tri**p**	ʃ	**sh**ip, pre**ss**ure, ra**sh**
t	**t**op, re**t**urn, pa**t**	ʒ	**J**acques, mea**s**ure, rou**g**e
k	**c**at, bi**k**er, sti**ck**	m	**m**ice, le**m**on, hi**m**
b	**b**at, ru**bb**er, sno**b**	n	**n**ick, fu**nn**y, gai**n**
d	**d**ay, a**d**ore, ba**d**	ŋ	si**ng**er, ba**ng,** ba**n**k
g	**g**uts, ba**gg**y, ri**g**	l	**l**ight, ye**ll**ow, fee**l**
f	**ph**oto, co**ff**ee, lau**gh**	ɹ	**r**ice, a**rr**ive, ve**r**y
v	**v**oice, ri**v**er, li**v**e	w	**w**inter, a**w**ay
θ	**th**ink, au**th**or, tee**th**	j	**y**ell, on**i**on
ð	**th**is, wea**th**er, tee**th**e	h	**h**ill, a**h**ead
s	**s**it, re**c**eive, ba**ss**	tʃ	**ch**op, na**t**ure, i**tch**
z	**z**oom, fu**zz**y, ma**ze**	dʒ	**j**udge, re**g**ion, a**g**e

Most of these consonant symbols in Table 1 correspond to familiar letters, and represent their usual sound values. For example, [f] and [h] in IPA are pronounced exactly as an English-speaker would expect from their spellings in *force* and *horse*. Let's consider the less familiar symbols:

[θ,ð]. English has two distinct consonant sounds (***th**eme* vs. ***th**ese*), both spelled with *th*; since these two consonant sounds are not the same, they should each have their own symbol. Moreover, a guiding principle of the IPA is that each individual speech sound corresponds to a unique symbol, and each symbol to a sound; while a sequence of sounds must be represented as a sequence of symbols. We therefore shouldn't use a two-letter sequence, *th*, to represent a single consonant sound, as this could be confused with a true sequence of consonants, e.g. the *th* in *sweetheart*. We therefore require two special IPA symbols, [θ] and [ð].

[ʃ,ʒ]. For similar reasons, the *sh* sound in *ship* deserves its own symbol, [ʃ]. And the middle consonant in *measure* (sometimes represented as *zh* in pronunciation guides) is [ʒ].

[ŋ]. *Stinger* does not really contain an [ng] sequence phonetically: it's a single consonant sound, similar to *n,* but with the tongue in the position of a [g]. If you want to confirm that there's no [g] in *stinger,* compare it to *finger,* which has an [ŋg] sequence.

[tʃ,dʒ]. Why then are the *ch* and *j* (or 'soft *g*') sounds represented as a sequence of symbols? These are actually not single consonants at all: they are [t] plus [ʃ], and [d] plus [ʒ], sequences. *Wheat ship* spoken quickly is indistinguishable from *wee chip*. Similarly, if you say *aid Jacques* quickly, the *d-j* sequence sounds the same as the end of *age*.

[ɹ]. This symbol ('turned *r*') represents the English *r* sound. In IPA, [r] is reserved for the (more common) trilled *r,* as in Spanish ***r**ojo* ('red') or *pe**rr**o* ('dog').

[j]. As in German, [j] represents the sound usually written in English as *y*. [j] is never pronounced as in *just*. (In IPA, the [y] symbol refers to a different sound, not found in the basic sounds of English.)

[g]. This symbol is always pronounced as a 'hard' *g*, as in *get* or *bag,* never as in *gem* or *age*.

By the same token, a number of letters of the alphabet are not needed as IPA symbols for transcribing English consonant sounds. For example, the *qu* in *quick* is the same as [kw], and the end of *tax* is simply a [ks] sequence. As we already noted, either [s] or [k] can replace *c*, depending on the word. These extra letters are used in IPA to denote different sounds, not found among the basic sounds of English.

B. Vowels. The vowels require more careful study, as the symbols are less familiar; and even the familiar symbols generally do *not* have the phonetic values we would expect from English spelling. They're more like the spelling-pronunciation correspondences of Spanish or Italian.

Table 2: IPA symbols for the basic vowel sounds of North American English

IPA symbol	Example words	IPA symbol	Example words
i	s**ee**, funn**y**, b**ea**d	ʊ	p**u**ll, g**oo**d, w**ou**ld
ɪ	b**i**t, s**i**ng, r**i**b	o	g**o**, b**oa**t, p**o**le, s**ew**
e	h**a**ze, gr**ea**t, ob**ey**	ɔ	c**au**ght, d**aw**n, b**o**ss
ɛ	b**e**t, s**e**nd, aff**e**ct	ɑ	c**o**t, D**o**n, f**a**ther
æ	st**a**mp, p**a**ck, y**ea**h	ɐ	sh**u**t, c**o**me, b**u**g[1]
u	l**oo**n, fl**u**te, wh**o**	ə	**a**bout, Albert**a**, el**e**ment

Note that, for many of these vowel sounds, a number of different spellings are used in English. The [ʊ] sound, for example is spelled *oo* in *good,* but *ou* in *would*; nevertheless, the vowel sound is the same in both words: *would* and *good* rhyme, which tells us that the vowel sounds (as well as the final consonants) in these two words are identical.

> Examine the example words for the other vowel symbols as well, to satisfy yourself that the sounds corresponding to each symbol really are the same.

The point of this mental exercise is to develop an awareness of the distinct vowel SOUNDS, independent of their spelling in particular words.

[ɔ,ɑ]. Except in certain regions, most younger North American English speakers nowadays make no distinction between [ɔ] (as in *caught*) and [ɑ] (as in *cot*), instead using [ɔ] for both cases; or [ɑ] for both cases; or a vowel somewhere between the two ([ɒ]). If you pronounce *cot/caught* and *Don/Dawn* the same, you're in this group of cutting-edge English speakers.

Dialect variation. More generally, bear in mind that the symbols and examples in Table 2 hold true for most dialects of North American English. But if you speak a dialect distinct from the North American mainstream, your vowels may vary significantly, as English dialects differ mainly in the vowels. *Remember: the 'right' way to transcribe a word depends on its pronunciation in the speech you are transcribing, not on any external standard of correctness.*

[1] The symbol [ʌ] is widely used for this vowel of North American English (although it represents a different vowel according to the IPA chart). Either symbol is acceptable for our purposes.

The Queen's English?

We beg your Majesty's pardon, but there is nothing inherently superior about any particular dialect of English – or any other language. The populations of Alabama, Manitoba, and Oxfordshire are equally 'good' English speakers, from a linguist's perspective. Each distinct dialect presents an equally valid object of study. The belief that some dialects are better than others is just another form of the attitude that some ethnic groups or social classes are better than others (more simply, 'prejudice'). For we tend (often unconsciously) to attach prestige to the dialects of groups we admire, and to stigmatize the dialects of groups we look down upon.

Diphthongs. English also has a few 'vowels' that are really a sequence of two vowels. These are called **diphthongs** (from Greek *di-* 'two' + *phthongos* 'sound'). The most common diphthong is the sound in *hide* or *eye*. It begins something like [ɑ], and moves smoothly into [ɪ]. If you say *eye* slowly, you can hear the one vowel change into the other. Because the sounds of a diphthong change from beginning to end, they are transcribed in IPA with two vowel symbols, as shown in Table 3.

Table 3: Diphthongs

IPA symbol	Examples
ɑɪ	h**i**de, **eye**, s**i**gh
ɑʊ	h**ow**, r**ou**nd
ɔɪ	b**oy**, av**oi**d

Vowel + [ɹ] sequences. When a vowel appears before [ɹ] in North American English, the [ɹ] has a strong effect on the vowel's sound, making identification of the vowel tricky, in some cases, for beginners at phonetic transcription. Here, then, is a list of examples.

Table 4: Vowel + [ɹ] sequences

IPA symbol	Examples	
eɹ	h**air**, c**are**d, wh**ere**, b**ear**	
iɹ	h**ere**, w**eir**d, **ear**, b**eer**	([ɪɹ] *in many dialects*)
ɑɹ	b**arr**ed, f**ar**, **ar**m	
ɔɹ	b**or**n, st**ore**, p**our**, sh**ore**	([oɹ] *in a few dialects*)
ʊɹ	t**our**, p**oor**, s**ure**	([uɹ] *or* [ɔɹ] *in many dialects*)

Check this list carefully, thinking about how *you* pronounce these words. Are the vowel + [ɹ] sequences in the examples on each row the same for you? Are the sounds of each row different from those of the other rows? For example, do you pronounce *pour* and *poor* differently, or the same?

A generation ago, many dialects of North American English had even more distinct vowel + [ɹ] sequences. The author's father, for example, pronounces *Mary*, *merry*, and *marry* differently: [meɹi], [mɛɹi], [mæɹi]. The author himself pronounces *Mary* and *merry* both as [meɹi], while sometimes observing a distinction between these and *marry*. But most university-age speakers now pronounce all three as [meɹi].

What about the vowel sound in *her* or *bird*? In fact, there is no distinct vowel + [ɹ] sequence in these words. In the examples in Table 4, there is a clear change from the vowel into the [ɹ], much like the change in the diphthongs in Table 3, as you can confirm by pronouncing the examples slowly. But in *her* and *bird*, there is just a single vowel sound: that is, the [ɹ] itself is 'serving as' the vowel. *Her* and *bird* should therefore be transcribed simply as [hɹ], [bɹd].

Stress. Consider the word *refund* [ɹifɐnd]. As a verb, the second vowel is **stressed** (it is a bit louder, longer and higher in pitch) than the first, whereas as a noun, the first vowel is stressed. This difference is reflect in IPA with a vertical accent mark immediately before the stressed **syllable**.[2] Thus, the sentence, *Will you refund me the money?* is transcribed [wɪl ju ɹiˈfɐnd mi ðə ˈmɐni?]; whereas the sentence, *I got my refund* is transcribed [ɑɪ gɑt mɑɪ ˈɹifɐnd].

C. Summary. A few remaining points: Never capitalize IPA symbols. For example, use [g], not [ɢ], for the initial consonant in *get*, even at the beginning of a sentence, and even in names. [ɢ] stands for a different sound, not found in English. Also, take care to keep your IPA symbols distinct: it can be difficult for a reader to distinguish a sloppily written [ʊ] from [u], or [ə] from [ɔ], [ð] from [d], etc.

No system of transcription can reflect all the minute differences between two utterances. There will be, for example, some differences in pronunciation between a forty-year-old man singing *Happy Birthday,* and a ten-year-old girl singing the same song. Indeed, even two forty-year-old men singing this will have individual voice characteristics, making the sounds somewhat different. For this level of detail, you need a recording. The IPA nevertheless provides a compact, low-tech, reasonably precise way of notating how the words of a language are pronounced. A transcriber has to decide how much detail is needed, depending on the uses to which the transcription will be put: for a precise description of the sounds, a lot of detail is needed (**narrow transcription**); for a description of word-order in sentences, much less detail is needed – just enough to distinguish one word from another **(broad transcription**). Narrow transcriptions are enclosed in **[square brackets]**, broad transcriptions in **/slashes/**.

---

[2] Syllables can be thought of as (meaningless) 'mini-words' into which longer words can be broken down. Each of these mini-words contains a single vowel (or diphthong), together with any consonants that can be grouped with that vowel. Thus, *appendicitis,* for example, can be broken down into [ə.pɛn.dɪ.sɑɪ.tɪs].

Finally, a warning: beginners often approach IPA transcription by trying to translate directly from English spellings into IPA symbols: e.g. *ea* (as in *treat*) = [i] in IPA. But the *ea* in *great*, for example, is not [i], it's [e]. As we've noted above, English spellings of sounds are notoriously inconsistent, making this strategy fundamentally unworkable. Rather, **always be aware of how the word *sounds*** in any exercise involving phonetic transcription**.**

Here's the paragraph to the left, in IPA:
[fɑɪnəli, ə wɑɹnɪŋ: bigɪnɹz ɔfən əpɹotʃ ɑɪ pi e tɹænskɹɪpʃən bɑɪ tɹɑɪŋ tə tɹænzlet dɑɪɹɛktli frəm ɪŋglɪʃ spɛlɪŋz ɪntə ɑɪ pi e sɪmbəlz. bɐt ðə i e ɪn gɹet, fɹ əgzæmpl, ɪz nɑt i, ɪts e. æz wiv notəd əbɐv, ɪŋglɪʃ spɛlɪŋz əv sɑʊndz ɹ notɔɹɪəsli ɪŋkənsɪstənt, mekɪŋ ðɪs stɹætədʒi fɐndəmɛntəli ɐnwɹkəbl. ɹæðɹ, ɔlwiz bi əweɹ əv hɑʊ ðə wɹd sɑʊndz ɪn ɛni ɛksɹsɑɪz ɪnvɑlvɪŋ fənɛtɪk tɹænskɹɪpʃən.]

Now here is the opening paragraph of this chapter in IPA. See if you can read it without referring back to page one:

> ɪn most fildz əv stɐdi, læŋgwədʒ ɪz θɔt əv prɪnsəpli ɪn tɹmz əv ðə ɹɪtən wɹd, fɹ ɪt ɪz ɪn ðɪs fɔɹm ðæt wi juʒəli mek pɹmənənt ɹɛkɹdz əv ɪmpɔɹtənt ɑɪdiəz. ɹɛlətɪvli lɪtəl ətɛnʃən ɪz speɹd fɹ sɐmpθɪŋ əz flitɪŋ ənd ənɹimɑɹkəbəl əz spokən kɑnvɹseʃən. ɪn lɪŋgwɪstɪks, hawɛvɹ, spitʃ, ɹæðɹ ðən ɹɑɪtɪŋ, ɪz ɹigɑɹdəd əz mɔɹ sɛntɹəl tə hjumən læŋgwədʒ, fɹ sɛvɹəl ɹizənz. fɹst, hjumənz həv pɹɑbəbli spokən læŋgwədʒəz fɹ fɪfti θɑuzənd jiɹz, pɹhæps mɐtʃ lɔŋgɹ. ɹɑɪtɪŋ ɪz a ɹɛlətɪvli ɹisənt dəvɛləpmənt, onli ə fju θɑuzənd jiɹz old. ivən tədе, ðeɹ ɹ səsɑɪətiz witʃ spik læŋgwədʒəz wɪθ no ɹɑɪtɪŋ sistəm, ɔɹ weɹ ɹɑɪtɪŋ ɪz sɛldəm juzd. bɐt ðeɹ ɪz no səsɑɪəti wɪtʃ kəmjunɪkets dʒɐst bɑɪ ɹɑɪtɪŋ, wɪθɑʊt ə spokən læŋgwədʒ. fɹðɹmɔɹ, tʃɪldɹən lɹn tə spik lɔŋ bifɔɹ ðe lɹn tə ɹid ənd ɹɑɪt; ɪndid, lɹnɪŋ əv spokən læŋgwədʒ teks ples wɪðɑʊt fɔɹməl ɪnstɹɐkʃən. bɐt dəz ɔɹdɪneɹi spitʃ ɹili wɔɹənt sɑɪəntɪfɪk ətənʃən? ɔlðo wi dʒɛnɹəli tek ðə pɹɑsɛsəz əv spitʃ pɹədɐkʃən ənd ɹɛkəgnɪʃən fɹ gɹæntəd, ðe ɪnvɑlv ə ɹendʒ əv sɹpɹɑɪzɪŋli ɪntɹɪkət mɛntəl əbɪlətiz – pɑɹt əv ðə nɑlədʒ wi hæv əv ðə læŋgwədʒəz wi spik. ðə wɹdz ðət wi wɪʃ tu əkspɹɛs sim tu imɹdʒ ɔtəmætɪkli fɹəm ɑʊɹ mɑʊðz, əz sɑʊnd wevz. ðise sɑʊnd wevz ðɛn hɪt ðə hiɹɹz iɹ, sɛndɪŋ ɔdətɔɹi sɪgnəlz tə ðə bɹen, wɪtʃ ɹ ɪntɹpɹətəd – əgɛn, simɪŋli ɔtəmætəkli – əz ðə wɹdz ɪntɛndəd bai ðə spikɹ. wɐt kɑɪnd əv mɛntəl sistəm mɑɪt ɐndɹlɑɪ ðɪs kəpæsəti tə ɹɛkəgnɑɪz ənd pɹədus spitʃ? wɪtʃ æspɛkts əv ðɪs sistəm əpiɹ tə bi kɑmən tu ɔl hjumənz, ənd wɪtʃ æspɛkts veɹi fɹəm læŋgwədʒ tə læŋgwədʒ? ənd wɐt əgzæktli goz ɑn ɪn ðə mɑʊθ ənd θɹot ɪn ɔɹdɹ tə pɹədus ðis sɑʊndz? ðiz sɔɹts əv kwɛstʃəns ɹ ðə domen əv fənɛtɪks ənd fənɑlədʒi (boθ fɹəm ðə gɹik ɹut *fon*- 'sɑʊnd'), ðə tu sɐbfildz əv lɪŋgwɪstɪks kənsɹnd wɪθ spitʃ sɑʊndz. ɪn ðə ɹimendɹ əv ðɪs tʃæptɹ, wi əgzæmən səm besɪk ɑbzɹveʃənz, tɹmənɑlədʒi, ənd tɛkniks əv ənæləsɪs juzd bɑɪ fonətɪʃənz ənd fənɑlədʒɪsts tu ədɹɛs ðis kwɛstʃənz.

Note that the pronunciation of particular words in a phrase may vary from their pronunciation in isolation, e.g. *and* as [ənd] rather than [ænd].

For your convenient reference, we repeat, in consolidated form, the IPA symbols discussed above:

Table 5: IPA symbols for the basic sounds of North American English

Consonants	
p	**p**at, hi**pp**y, tri**p**
t	**t**op, re**t**urn, pa**t**
k	**c**at, bi**k**er, sti**ck**
b	**b**at, ru**bb**er, sno**b**
d	**d**ay, a**d**ore, ba**d**
g	**g**uts, ba**gg**y, ri**g**
f	**ph**oto, co**ff**ee, lau**gh**
v	**v**oice, ri**v**er, li**v**e
θ	**th**ink, au**th**or, tee**th**
ð	**th**is, wea**th**er, tee**the**
s	**s**it, re**c**eive, ba**ss**
z	**z**oom, fu**zz**y, ma**ze**
ʃ	**sh**ip, pre**ss**ure, ra**sh**
tʃ	**ch**ip, fu**t**ure, sti**tch**
ʒ	**J**acques, lei**s**ure, rou**ge**
dʒ	**j**erk, proce**d**ure, e**dge**
m	**m**ice, le**m**on, hi**m**
n	**n**ick, fu**nn**y, gai**n**
ŋ	si**ng**er, ba**ng,** ba**n**k
l	**l**ight, ye**ll**ow, fee**l**
ɹ	**r**ice, ve**r**y, b**ir**d, h**er**, f**ur**
w	**w**inter, a**w**ay
j	**y**ell, on**i**on
h	**h**ill, a**h**ead

Vowels	
i	s**ee**, funn**y**, b**ea**d
ɪ	b**i**t, s**i**ng, r**i**b
e	h**a**ze, gr**ea**t, ob**ey**
ɛ	b**e**t, s**e**nd, aff**e**ct
æ	st**a**mp, p**a**ck, y**ea**h
u	l**oo**n, fl**u**te, s**ou**p, wh**o**
ʊ	p**u**ll, g**oo**d, b**oo**k
o	g**o**, b**oa**t, p**o**le, s**ew**
ɔ	c**au**ght, d**aw**n, b**o**ss
ɑ	c**o**t, D**o**n, f**a**ther
ɐ	sh**u**t, c**o**me, b**u**g
ə	**a**round, Albert**a**, el**e**m**e**nt

Diphthongs	
ɑɪ	h**i**de, **eye**, **I**, s**igh**
ɑʊ	h**ow**, r**ou**nd
ɔɪ	b**oy**, av**oi**d

Vowels + ɹ	
ɑɹ	b**arr**ed, f**ar, ar**m
eɹ	h**air**, c**are**d, wh**ere**
iɹ	h**ere**, w**eir**d, b**eer**
ɔɹ	b**or**n, st**ore**
ʊɹ	t**our**, m**oor**

Exercises

1. Each word below (as pronounced by a native speaker of any dialect of English) has one clear mistake in its transcription. Circle the specific part of the transcription where the mistake occurs, and show what the correct symbol(s) (if any) should be. (Ex: honest, [hɑnəst], *should be* Ø (i.e. nothing); rain, [ɹɑɪn], *should be* e.)

Written:	*IPA:*	*Should be:*	*Written:*	*IPA:*	*Should be:*
a. shine	[shɑɪn]	____________	b. beauty	[bjuty]	____________
c. wrench	[wɹɛntʃ]	____________	d. paper	[papɹ]	____________
e. jumping	[jɐmpɪŋ]	____________	f. savage	[sævædʒ]	____________
g. user	[uzɹ]	____________	h. shed	[ʃed]	____________
i. teacher	[tichɹ]	____________	j. his	[hɪs]	____________

2. Give the English words corresponding to the following IPA transcriptions:

a.	[bɑks]	____________________	b.	[bled]	___________________
c.	[sin]	____________________	d.	[kloð]	____________________
e.	[sɐŋ]	____________________	f.	[steɹ]	____________________
g.	[but]	____________________	h.	[ɹod]	____________________
i.	[jæm]	____________________	j.	[itʃ]	____________________

3. Transcribe the following English words using IPA, based on your own pronunciation; if you are not a (near-) native speaker of English, use the pronunciation of a friend who is.

a.	board	____________________	b.	touch	____________________
c.	queen	____________________	d.	graph	____________________
e.	feelings	____________________	f.	laundry	____________________
g.	crime	____________________	h.	thigh	____________________
i.	shoot	____________________	j.	belong	____________________

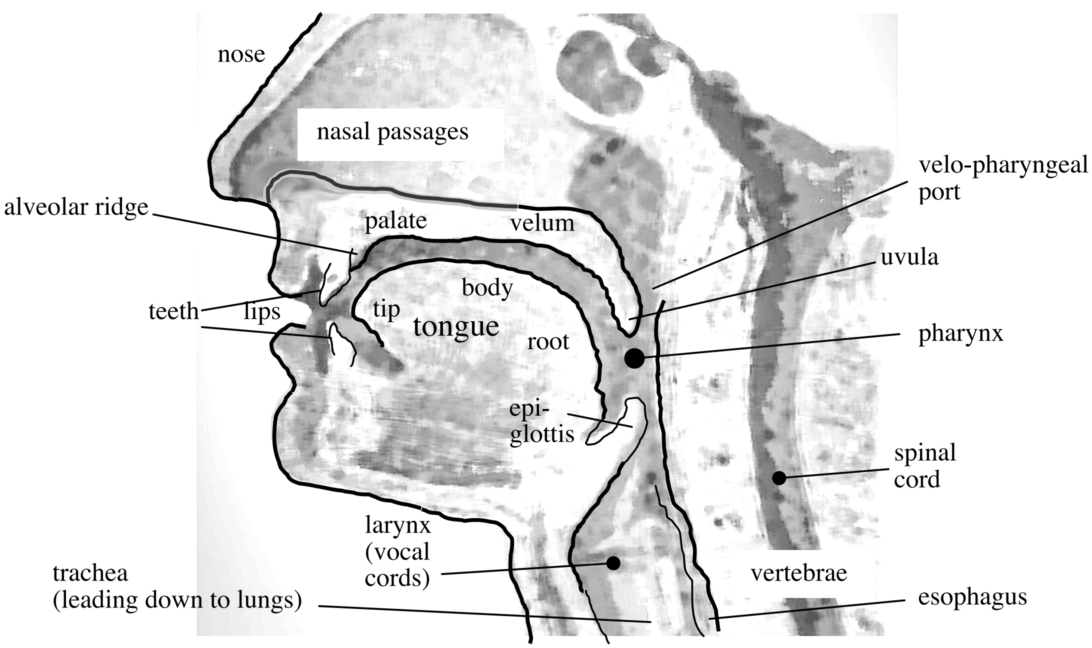
4. What does this passage say? Write it in English spelling.

[waɪ ɪz ɪŋglɪʃ spɛlɪŋ so ɪŋkənsɪstənt? ɪŋglɪʃ spɛlɪŋ ɹilaɪəbli kɔɹəspandz tə pɹənɐnsieʃən – əz ðə læŋgwədʒ wəz spokən faɪv hɐndɹəd jiɹz əgo, ðæt ɪz. ðə pɹənɐnsieʃən əv wɹdz həz tʃendʒd dɹəmætɪkli ɪn ðæt taɪm, bɐt spɛlɪŋz hævənt bin sɪstəmætɪkli ɐpdetəd tə riflɛkt ðiz saʊnd tʃendʒəs. ɛksɛpʃənz ɔlso əɹaɪz dju tə baɹəwɪŋz fɹəm ɐðɹ læŋgwədʒəs. ðə spɛlɪŋgz aɹ ɔfn baɹod tu, ənd ðə ɹəzɐlts fɹikwəntli kənflɪkt wɪð nɔɹməl ɪŋglɪʃ spɛlɪŋ ɹulz, fɹ ɛgzæmpəl, ɪtæljən tʃelo (weɹ ðə lɛtɹ si ɪz pɹənaʊnst tʃ) vɹsəz ɪŋglɪʃ sɛl (weɹ si ɪz pɹənaʊnst s).]

II. Articulatory phonetics

The study of how speech sounds are formed in the mouth (or '**articulated**') is called **articulatory phonetics**. Speech sounds are produced by the **vocal tract**: the mouth, nose, throat, and lungs. Let's take a look inside:

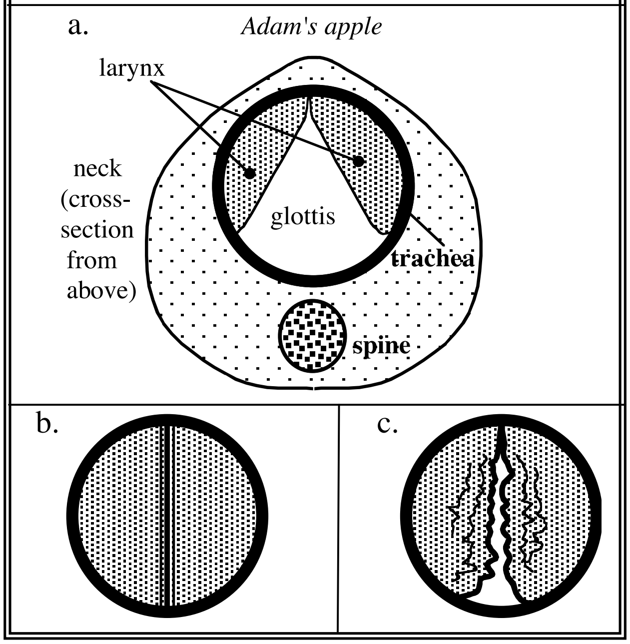
Figure 1: Anatomy of the vocal tract



A. Anatomy. The **alveolar ridge** refers to the gums just behind the upper teeth. The **palate** refers to the 'hard palate,' i.e. the roof of the mouth. The 'soft palate' is called the **velum**, and ends in the **uvula** (this is the fleshy appendage you can see hanging down in the back of your throat). If the velum is raised, this closes the **velo-pharyngeal port**, preventing airflow between the nasal passages and the rest of the vocal tract. The tongue is a mass of muscle, which we can divide into tip (the only part you usually see), body, and root. The **epiglottis** is a flap below the **pharynx** (the back of the throat): it covers the **trachea** (or 'windpipe') when you swallow, so that food goes down the esophagus instead.

Lastly, the **larynx** is a sort of valve, encased in cartilage (the 'Adam's apple,' more prominent in males, but present in all humans), at the top of the trachea. It opens wide during breathing (Fig. 2a); closes when you swallow (b); and when you say a vowel, the two sides draw together, so that they vibrate as air passes through (c). This **voicing** (pulsing of air in the **glottis** as it passes through the vibrating larynx) is what creates the sound of your voice.

Figure 2: Schematic representations of the larynx (not to scale)



B. Consonants. Speech sounds are the result of movements of parts of the vocal tract, particularly the lips, tongue tip, tongue body, and larynx (the major **articulators**) which affect the flow of air as you exhale. **Consonants** are formed with significant **obstruction** of this airflow by one or more of the articulators; whereas in **vowels**, the mouth remains relatively open. We can describe particular types of consonants in terms of *how much obstruction* is involved (**manner of articulation**).

Stops ([p,t,k,b,d,g]) involve a complete blockage of airflow, due to full closure at some point in the mouth.

Nasals ([m,n,ŋ]) involve complete closure in the mouth, but the back of the velum is lowered, allowing the airflow to pass through the velo-pharyngeal port, and out the nose.

Fricatives ([f,v,θ,ð,s,z,ʃ,ʒ,h]) involve a partial constriction in the mouth, such that airflow is forced through a narrow channel, creating a hissing sound.

Affricates is a term sometimes used for stop + fricative sequences made with the same articulator, including ([tʃ,dʒ]).

Approximants ([l,ɹ,j,w]) involve less obstruction than a fricative, but more than a vowel. In an [l], the tip of the tongue often makes full contact with the alveolar ridge, but one side of the tongue is lowered: [l] is therefore called a **lateral approximant**; the others are **central**.

We can also classify consonants in terms of the state of the larynx (**phonation**) during their pronunciation.

> **Feel the buzz!**
> With two fingers on your Adam's apple, say [sssszzzzsssszzzz...]. You should be able to feel a vibration under your fingers, during the [zzzz] parts only. That 'buzz' is **voicing**.

Voiced consonants ([b,d,g,v,ð,z,ʒ,m,n,ŋ,l,ɹ,j,w]) are accompanied by voicing (Fig 2c).

In **voiceless** consonants ([p,t,k,f,θ,s,ʃ,h]), the glottis is more open, as in Fig. 2a, so that air passes through without vibrating.

Finally, consonants can be described in terms of *where* the obstruction occurs in the vocal tract (**place of articulation**).

Bilabials ([p,b,m,w]) involve closure or constriction of the two lips.

Labiodentals ([f,v]) involve constriction of the upper teeth and lower lip.

Dentals ([θ,ð]) involve constriction of the tongue tip and the upper teeth.

Alveolars ([t,d,n,s,z,ɹ,l]) involve constriction of the tongue tip and the alveolar ridge.

Post-alveolars (or palato-alveolars) ([ʃ,ʒ]) involve constriction of the tongue tip and the palate, just behind the alveolar ridge.

Palatals ([j]) involve constriction of the tongue body and the palate.

Velars ([k,g,ŋ,(w)]) involve constriction of the tongue body and the velum. ([w] is considered a velar as well as a bilabial because it involves constrictions both at the lips and velum.)

Glottals ([h]) involve constriction of the glottis (in this case, sufficient constriction to create a fricative, but not enough to cause voicing).

These classifications of consonants are summarized in the following chart:

Table 6 : Classification of English consonants by manner, place and phonation type

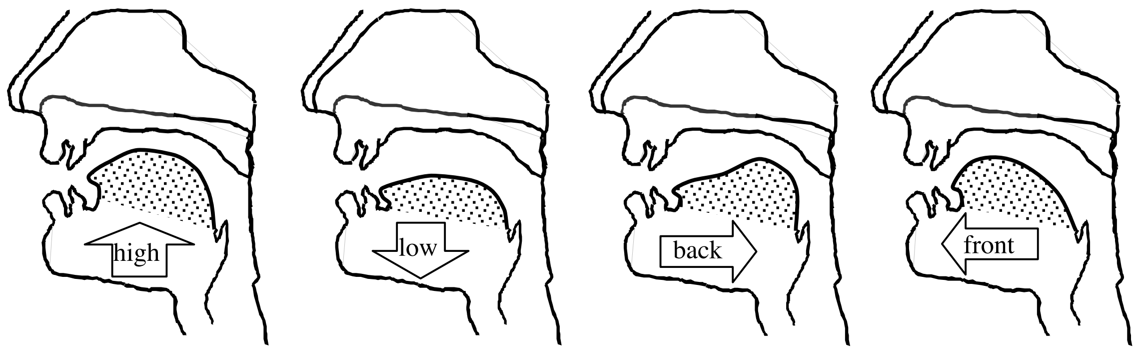
		bi-labial	labio-dental	den-tal	alve-olar	post-alveolar	pala-tal	velar	glot-tal
stops	voiceless	p			t			k	
	voiced	b			d			g	
fricatives	voiceless		f	θ	s	ʃ			
	voiced		v	ð	z	ʒ			h
affricates	voiceless					tʃ			
	voiced					dʒ			
nasals		m			n			ŋ	
approximants	central	w			ɹ		j	(w)	
	lateral				l				

We can thus articulatorily describe [s] as a voiceless alveolar fricative; [ŋ] as a (voiced) velar stop; etc. Likewise, we can refer to the *set* [b,d,g] as the class of voiced stops. A

natural class of sounds is a set such as this, which can be defined in terms of some shared phonetic property or properties.

C. Vowels. Unlike consonants, the various vowels are distinguished by the way the shape of the mouth – in particular, the position of the tongue body – affects the sound of your voice – in particular, the position of the **tongue body**.

Figure 3: Tongue body positions



Using height and frontness of the tongue body (Fig. 3), we can classify the vowels of English as shown in Table 6.

Table 6: Classification of English vowels by height and frontness, etc.

		front	**central**	**back**
high	close	i		u
	open	ɪ		ʊ
mid	close	e	ə	o
	open	ɛ		ɔ
low	close	æ	ɐ	
	open		a	ɑ

Note that the term '**mid**' refers to vowel *height*, while '**central**' refers to the front/back dimension.

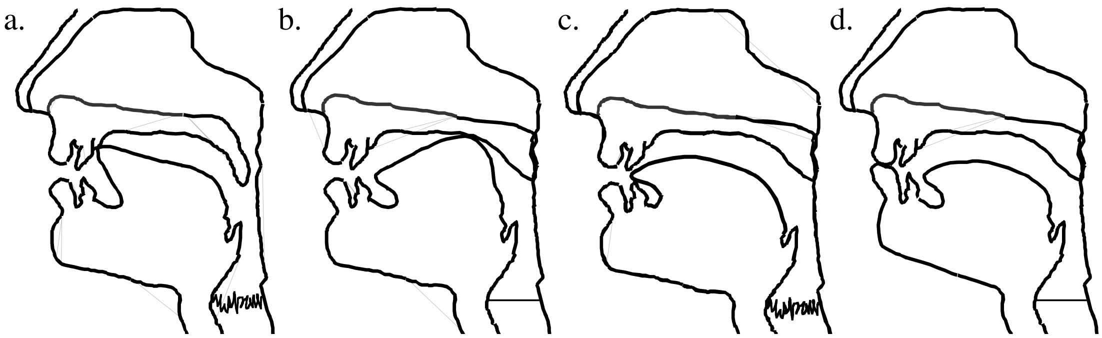
We use '**close**' and '**open**' to further differentiate vowel heights within the high, mid, and low ranges.[3]

In addition to tongue-body position, vowels are affected by **rounding** of the **lips**. The **rounded vowels** of English are enclosed in the oval in Table 6. We can thus describe [u] as a *close-high back rounded vowel*; [ɛ] is an *open-mid front unrounded vowel*; [a] is a *low central unrounded vowel*; etc.

[33] Our terminology here is an extension of the IPA's 'close' and 'open' for the mid vowels. More standardly, [i,e,o,u] are called '**tense,**' and [ʊ,ɪ,ɛ,ɔ], '**lax**', reflecting an early hypothesis that [i,e,o,u] are articulated with greater 'muscular tension' in the vocal tract. But tenseness does not distinguish [ɐ] from [a]; and subsequent research has indicated that the 'tense' vs. 'lax' vowels are not distinguished by any single property, but by height, duration, and centralization, in differing combinations for each of the vowels.

Exercises

1. Give the IPA symbols for the sounds corresponding the articulations shown in the following diagrams. (Voicing is indicated with a zig-zag line at the larynx.)



2. Give the IPA symbols for the sounds with the following articulatory descriptions:

   a. voiceless glottal fricative ________ b. voiced bilabial nasal ________
   c. open-high back rounded vowel ________ d. voiced palatal approx. ________
   e. voiced post-alveolar fricative ________

3. Give the articulatory description for the following sounds:

   a. [ŋ] ____________________________ b. [j] ____________________________
   c. [θ] ____________________________ d. [v] ____________________________
   e. [e] ____________________________

4. The following sets of sounds are natural classes, characterized by shared articulatory properties. For each of the sets, identify these properties. Examples: [t,d] are the set of *alveolar stops*. [m,n,ŋ] are the set of *nasals;* they are also *voiced*, but the voiced set includes other sounds as well, so only *nasals* is correct.

   a. [i,ɪ,e,ɛ,æ] ________________________________
   b. [p,b] ________________________________
   c. [ɹ,l,j,w] ________________________________
   d. [v,ð,z,ʒ] ________________________________
   e. [i,ɪ,u,ʊ] ________________________________

III. Acoustic phonetics

A. Fundamentals of sound. Speech sounds can also be understood in terms of their **acoustic** properties, i.e. properties of the soundwaves. Soundwaves are simply waves of fluctuating air pressure, radiating out from their source. It is the structure of these waves which distinguishes one sound from another.

In a **pure tone** (approximated by the sound of a tuning fork) these ripples of air pressure correspond to a simple sine function, where the x-axis is *time*, and the y-axis is *pressure*. Such a wave has a particular **frequency**, measured in **Herz** (**cycles per second**): the higher the frequency, the *higher* the sound is in pitch. The sine wave in Fig. 4a has a little over 3 cycles per 10 milliseconds, or 300 cycles per second, i.e. 300 Hz.[4] Moreover, the more extreme the fluctuations in pressure, the greater the **amplitude** of the wave (measured in **decibels**), and the *louder* the sound. In comparison to Fig. 4a, the wave in Fig. 4b is higher and quieter. If we *sum* the two waves above, the result is a **complex waveform** (4c). The more individual sine waves we combine, the more complex the resulting waveform.

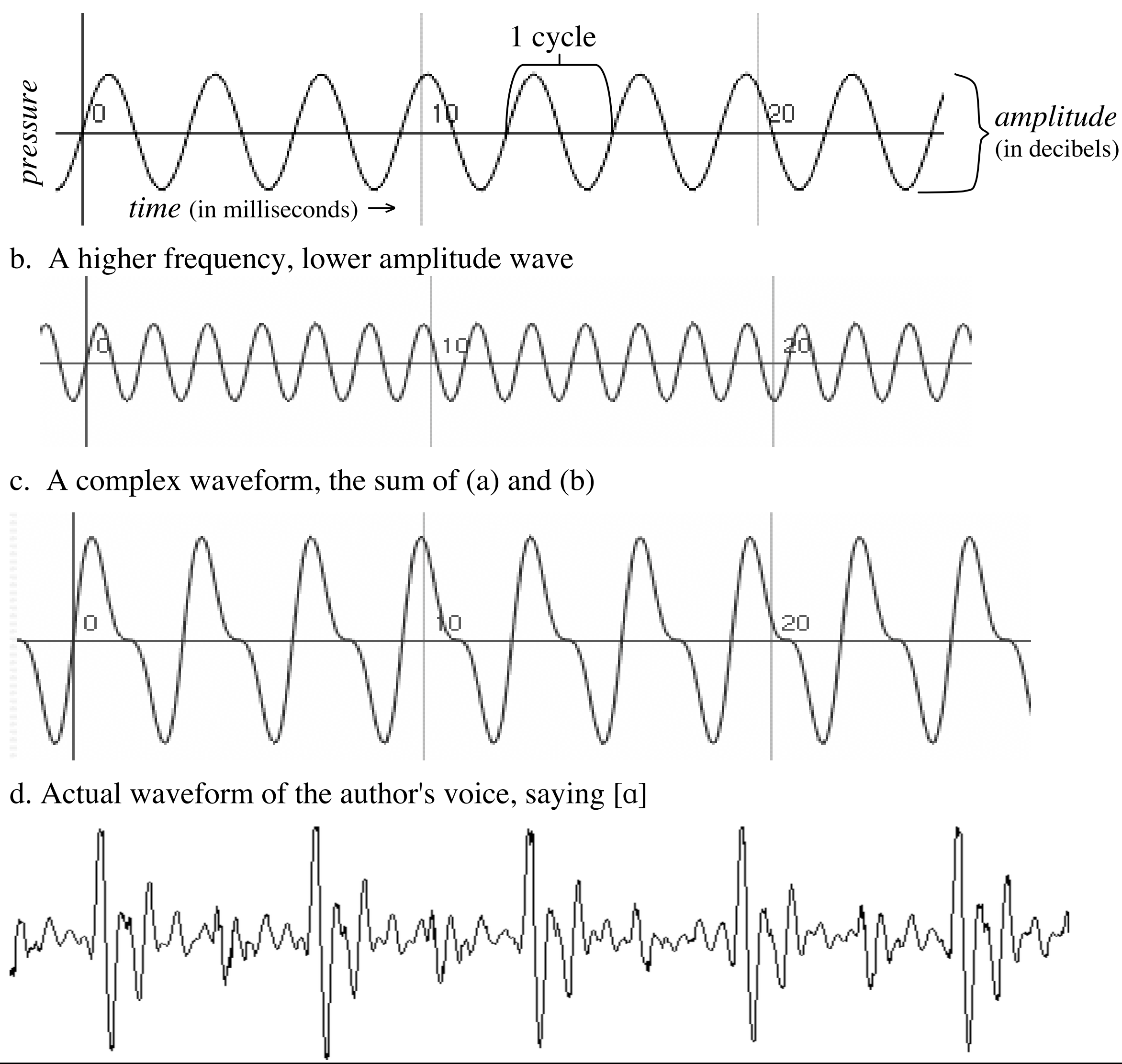
Figure 4

a. A sine function:

b. A higher frequency, lower amplitude wave

c. A complex waveform, the sum of (a) and (b)

d. Actual waveform of the author's voice, saying [ɑ]

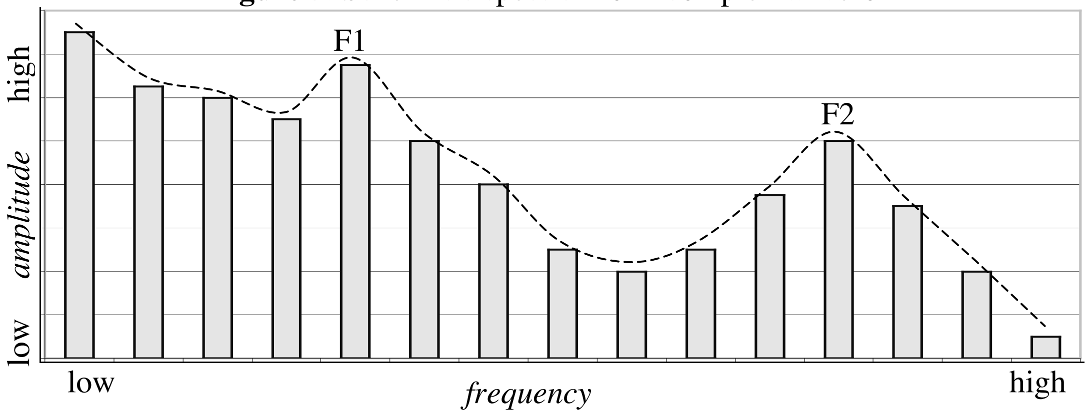


The sound signals of speech are always complex waveforms (see Fig. 4d). But just as we can sum simple sine waves to yield the complex wave in Fig. 4c, we can also take a complex waveform and break it down into simple waves, each with its own frequency and amplitude (a mathematical technique called **Fourier analysis**). The lowest-frequency component of the waveform is called the **fundamental** frequency (**F0**), which

[4] That's roughly D below concert A, for you musicians.

we hear as the basic pitch of the speaker's voice. The higher-frequency waves, all natural-number multiples of the fundamental, are called **harmonics**. In speech, particular harmonics can be louder or quieter, depending on the position of the tongue and other organs of the vocal tract. The amplitude profile (the dotted line in Fig. 5) of these harmonics (the vertical bars of varying heights) forms a **spectrum**.

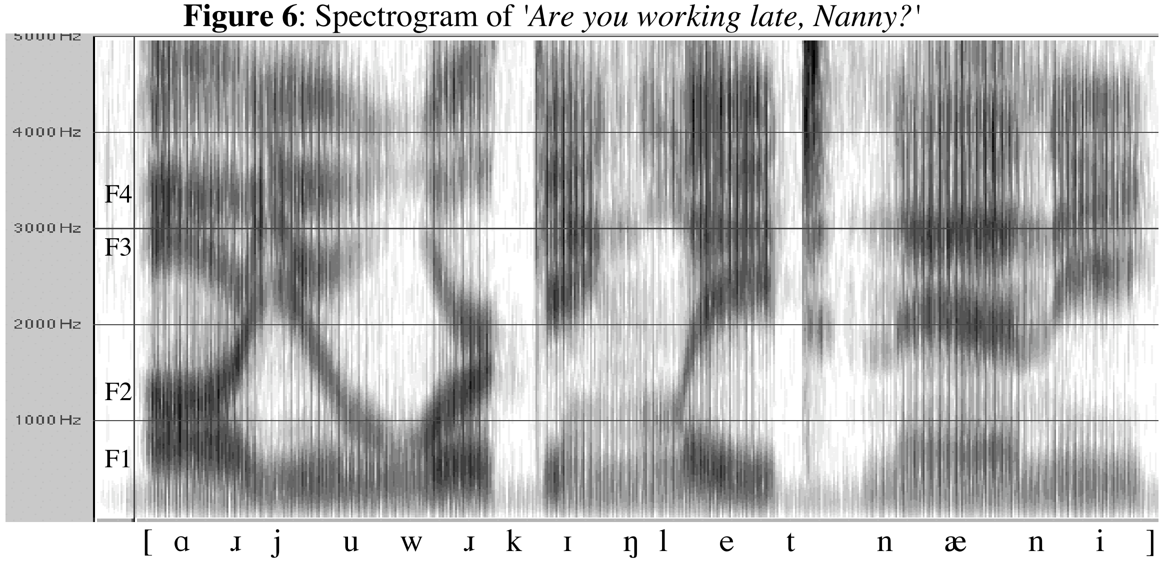
Figure 5: Schematic spectrum of a complex waveform



Peaks in the spectrum are called **formants**: the lowest-frequency peak above the fundamental is called the **first formant**, or **F1**; the next is **F2**, and so on (only the first three formants are relevant for speech perception).

Fig. 5 shows a spectrum of a speech sound at a single point in time. It is more informative, however, to show how the spectrum changes from moment to moment during speech. Such a display is called a **spectrogram** – with *time* on the x-axis, *frequency* on the y-axis, and the *higher-amplitude* frequency regions shown as darker areas (Fig. 6).

Figure 6: Spectrogram of '*Are you working late, Nanny?*'



The thick horizontal bands in Fig. 6 are the formants. The grainy vertical 'ridges' (**striations**) are individual *pulses of voicing*. This display helps us understand which **cues** (acoustic properties) identify particular consonants and vowels.

B. Vowel cues. Vowels are acoustically distinguished principally by the **frequencies of the formants**.

The **higher** the vowel articulatorily, the **lower the F1 frequency**.

The **backer** the vowel, the **lower the F2 frequency**.

Lip **rounding** further **lowers F2**.

The formants smoothly **change** in frequency during a **diphthong**, from the values of the first vowel to those of the second.

C. Approximant cues. Approximants are similar in cues to vowels.

[w,j] are very **similar** in their formant frequencies **to the high vowels** [u,i] respectively, but a bit **shorter** in duration, with a slightly **lower F1**, and a slight **weakening** of the higher formants, particularly in [w].

[ɹ] is similar in formant frequencies to [ə], but with **low F3**.

[l] is similar to [ɹ], but with **high F3**.

D. Fricative cues. Up to this point, we have focussed on **periodic** (humming) sounds. Periodic sounds, such as the vowel shown in Fig. 4d, have a repeating pattern to the waveform. Fricatives, however, involve **aperiodic (**hissing or crackling) noise. Note in Fig. 7 the fricatives [s] and [ʃ], which look like charcoal smudges, vs. the vowels [u] and [i], which have clear vertical striations and clear formants.

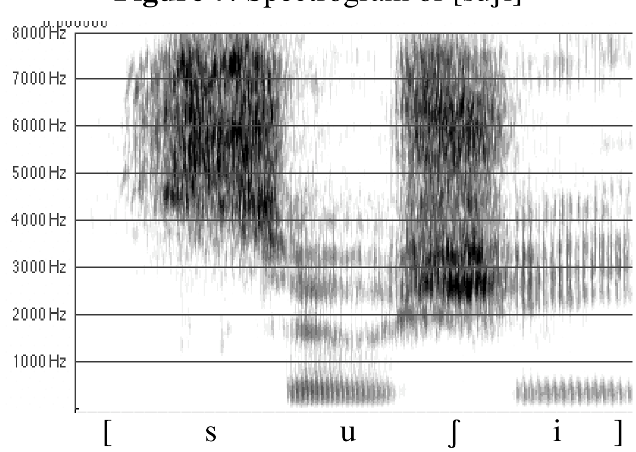
Although they have no fundamental frequency, aperiodic signals can be stronger in some frequencies and weaker in others.

The alveolar [s] has almost all of its noise above 4000 Hz ([z] too), whereas the post-alveolar fricative's noise extends down to 2000 Hz.

Voiced fricatives are generally shorter than the voiceless ones, and may have a band of voicing striations along the bottom of the spectrogram.

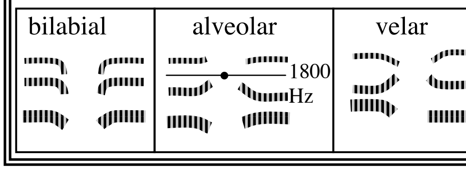
The other fricatives ([f,v,θ,ð,h]) are all much quieter than [s,ʃ,z,ʒ]. The labiodentals ([f,v]) are typically slightly louder than the interdentals, with more noise below 4000 Hz. [h] has bands of aperiodic energy in the same frequency regions as the formants of adjacent vowels.

Figure 7: Spectrogram of [suʃi]



E. Stop cues. The complete articulatory closure in a stop results in an interval of silence, which shows up as a blank column on a spectrogram, followed by a brief burst of aperiodic noise when the closure is released (see the [k] and [t] in Fig. 6). The stops are distinguished from each other by movement of the formants before and after closure (formant transitions) and by properties of the burst.

Figure 7: Formant transitions



- In bilabial stops, all formants move downward heading into the closure, and upward coming out of the closure.
- In alveolar stops, F2 heads towards a frequency of around 1800 Hz moving into closure, and originates from the same frequency coming out of closure. The release burst has considerable energy above 4000 Hz (note the burst after the [t] in Fig. 5).
- In velar stops, F2 and F3 move toward each other heading into closure, and split apart coming out of closure. Velars also frequently have a double burst note the two vertical 'blips' of noise after the [k] in Fig. 5).
- Voiced stops are shorter than voiceless stops, and they may have a narrow band of dark striations (a voicing bar) at the very bottom of the spectrogram. Voiceless stops, particularly in English, have a delay between the release burst and the start-up of full voicing in the following vowel.

F. Nasal cues. Nasals are acoustically somewhat like approximants, and somewhat like stops.

They resemble approximants in that one can see formants and voicing striations during the whole consonant. Nasals have a low F1, and a marked weakening of the higher formants.

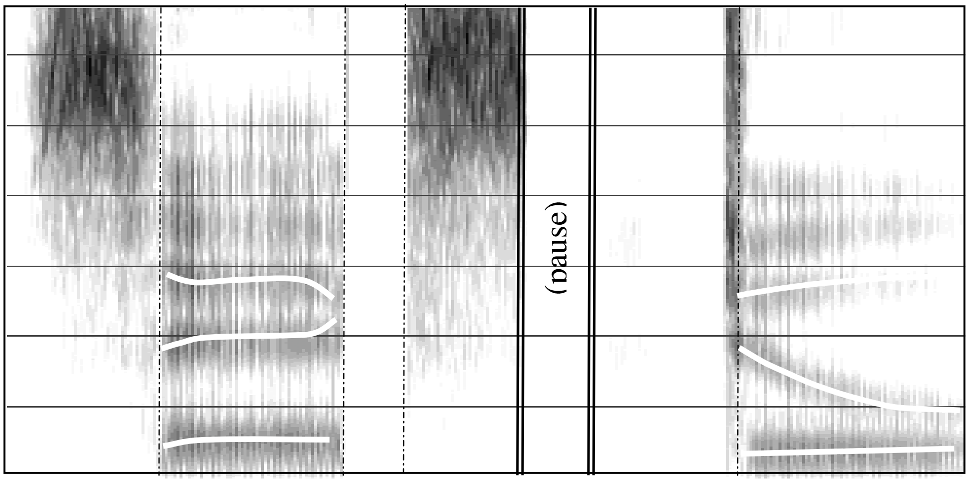
Like stops, identification of the nasal's place of articulation depends on formant transitions, into and out of closure. The transitions in a bilabial nasal are similar to those of a bilabial stop; likewise for nasals and stops at other places of articulation.

Exercises

1. Fill in the blanks:
   a. A soundwave without a repeating pattern is _____.
   b. A graphic display of sound showing changes in formants over time is a _____.
   c. The basic pitch of a voice is its _____ frequency.
   d. Voicing appears on a spectrogram as _____.
   e. Stops appear on a spectrogram as _____.

2. Fill in the blanks:
   a. The differences between [s] and [ʒ], as they appear on a spectrogram, are _____.
   b. The differences between [m] and [ŋ], as they appear on a spectrogram, are _____.
   c. The differences between [e] and [o] , as they appear on a spectrogram, are _____.
   d. The vowel [i] has _____ F1 frequency and _____ F2 frequency *(high or low)*.
   e. The vowel [ɑ] has _____ F1 frequency and _____ F2 frequency *(high or low)*.

3. The following spectrogram contains two single-digit numbers of English. What are they (in order)? To help you, dotted lines are drawn between the sounds, a pause between the words is marked, and F1-F3 are highlighted.



IV. The problem of variation in speech.

One potato, two potato, three potato, four…

Behind this childhood counting-game lies a profound scientific puzzle. For each utterance of *potato* as someone recites this rhyme, the actual soundwaves hitting a listener's ear are somewhat different, depending on such factors as speech rate, loudness, background noise, position within the sentence. If several people say it, there are even greater differences, depending on the age, gender, and dialect of each speaker. More generally, we can say that, like snowflakes or fingerprints (or potatoes, for that matter), no two utterances of a word are ever exactly the same, in English or any other language.

But except in extreme cases, we are able to instantly recognize each utterance as a mere repetition of the same word, *potato*; indeed, we are generally not even aware of the variation. How is it that English-speakers can zero in – without any conscious thought – on the properties of the sound signal which distinguish the intended word, in this case *potato,* from similar-sounding words (e.g. *tomato* or *petunia*), without getting sidetracked by the irrelevant differences? Speakers of other languages show the same facility in recognizing words of their language, despite similar types of variation in the signal. In fact, humans' ability in this regard is far more sophisticated than that of any existing speech recognition software, even running on the world's most powerful supercomputer.

Furthermore, the particular properties of the sound signal which distinguish one word from another vary from language to language. For example, in English, in the word *boot* ([but]), you can draw out the vowel for 400 msec (0.4 seconds), or shorten it to 150 msec (0.15 seconds): such a vowel duration difference is merely an irrelevant detail. However, in Cree, an indigenous language spoken through much of Canada, these distinct vowel durations can mean the difference between one word and another. Conversely, English has a distinction between the vowel sounds in *greet* ([gɹit]) and *grit* ([gɹɪt]), whereas this vowel distinction is absent in French.

In sum, there is no uniform set of sound properties which are relevant for speech across all languages. And since we grow up speaking the language(s) of the society around us (not necessarily those of our biological ancestors), our ability to zero in on the particular set of sound properties which are relevant for our language can't be attributed to our genes, like hair colour. Important aspects of this mental speech system have to be learned. Indeed, this is a crucial part of learning a language fluently. But in the case of a first language, we seem to pick up this knowledge within a few years of birth, without any formal instruction – in fact, without much conscious thought at all.

> ***So what do you know?***
>
> If you speak English fluently, you must already 'know' the phonological rules of English. But how can you 'know' something that you've never even thought about before? Actually you know a great many things, without being at all conscious of that knowledge. For example, you probably know how to pick up a carton of milk, a complex task requiring nearly instantaneous assessment of the weight of the milk vs. the strength of the container, so that you neither drop nor crush it (robots are terrible at this task). But humans do this without conscious thought; and it is difficult to put this knowledge into precise words. Psychologists call this 'implicit knowledge.' Speakers' knowledge of the phonological rules of their language is likewise implicit. We're generally unaware of these rules (outside of linguistics courses). But we instantly detect violations of these rules, e.g. in speech with a foreign accent, or in computer-synthesized speech.

In the remainder of this chapter, we focus on the observation that languages obey **phonological rules** – rules concerning what sounds may occur in the language, and how these sounds may be put together to form words of the language. Have you ever overheard someone speaking a language you don't understand – nevertheless you've been able to recognize the language as French, Spanish, Chinese, etc.? How can you identify a language without being able to recognize any of the words? The answer is that you're recognizing the phonological rules which characterize the language.

A plausible hypothesis is that phonological rules arise as particular languages' responses to this problem of maintaining recognizable words despite variation.[5] Consider the fact that language sound systems (henceforth '**phonological systems**') must be able to convey a broad range of information, with a minimum of confusion, for a broad range of speakers and hearers, across a broad range of situations. This practical consideration introduces two important constraints on phonological systems:

Ease of perception: recovery of meaning must not depend on cues which are
subtle, i.e. difficult to hear, nor
unstable, i.e. not always present in the signal, nor
singular, i.e. differences in meaning are not supported by multiple cues; misperception of just one cue could result in confusion of meaning.

Ease of articulation: recovery of meaning must not depend on cues that require highly effortful or precise articulations.

[5] Although this hypothesis is accepted, in some form, by most phoneticians and phonologists, the question of how *directly* phonetic pressures constrain phonological systems, and whether some aspects of phonological systems are independent of these phonetic constraints, are topics of debate in current linguistic theory.

Weird *phonology*:

A language might consist of nothing but sequences of [f] and [θ], where [fθθfθ] means 'dog', [θffθθf] means 'cat', etc. It might be 'spoken' by singing particular sequences of exact pitches. More imaginatively, Kurt Vonnegut's novel *Slaughterhouse Five* presents a race of extraterrestrials who communicate by tapdancing and making other bodily noises. Clearly, none of these is remotely like a real human language. But what's the difference; and why hasn't any human society ever developed anything like them? A plausible answer is that these imaginary systems seriously violate **Ease of Perception** or **Ease of Articulation**. The [fθfθ] language depends upon accurate perception of quiet fricatives, which are easily confused with each other, and easily masked by background noise. Vonnegut's alien language would be considerably more strenuous (for earthlings, at any rate) than speech. And the singing language would require all speaker/hearers to have perfect pitch (in perception) and flawless intonation (in production).

On the other hand, language doesn't need sound at all. Sign languages (principally used by deaf communities) are complete human languages, independent of the sound-based languages of the societies around them; and they are sight- rather than sound-based. Nevertheless, sign languages are subject to similar functional constraints: they avoid signs which involve extreme physical exertion or dexterity (e.g. walking on one's hands), or which require perception of extremely subtle gestures (e.g. a twitch of the calf muscles).

Each language develops its own particular set of rules, as strategies for satisfying these constraints. This is not to say that anyone ever sat down and consciously *designed* a phonological system. Rather, these systems continually evolve, through the back-and-forth of communication, and miscommunication, between speakers and hearers – including young children learning the language. Moreover, these rules are not **prescriptive** rules, which speakers are explicitly taught that they *should* obey (e.g. *don't say 'ain't'*): speakers follow these rules without even thinking about them. Indeed, it requires careful analysis, and some understanding of phonetics, to be able to figure out what the rules are – even for one's own language.

V. Phonemes and allophones

A. Allophonic variation. Because the organs of the vocal tract generally move in smooth trajectories rather than abrupt jerks, sounds are inevitably influenced by the sounds around them. As a case in point, consider the influence of nasal consonants on preceding vowels in English. In words such as *ran*, *doom*, or *sing*, the velum begins to lower, opening the nasal passages, well before the oral closure in the nasal consonant begins. This results in a significant part of the vowel being **nasalized**. This sort of overlap in movements of the articulators is called **coarticulation**. In a narrow transcription, these examples should therefore be transcribed as [ɹæ̃n], [dũm], [sĩŋ] ([˜] is the IPA **diacritic**, or supplementary symbol, for nasalization).

We thus have two different sets of vowels in English:

nasal ([ĩ,ũ,ɪ̃,ʊ̃,ẽ,ə̃,õ,ɛ̃,ɔ̃,æ̃,ɐ̃,ɑ̃]), and
oral ([i,u,ɪ,ʊ,e,ə,o,ɛ,ɔ,æ,ɐ,ɑ]).

English speakers are generally unaware of this phonetic distinction in their speech, because there are no *words* solely distinguished by nasalization of vowels. We have *bow* ([bo]) and *bone* ([bõn]), but not [bõ]. On the other hand, this nasalization is not simply an automatic physiological consequence of pronouncing a vowel before a nasal consonant. Some languages *do* have this distinction in words, e.g. French [bo] ('handsome') vs. [bõ] ('good'), or Dene Sųłine (an indigenous language of Northwestern Canada) [tɑbil] ('net *for* water') vs. [tɑ̃bil] ('net *in* water').

Languages also cope with variation by enhancing certain phonetic distinctions with additional cues. Consider the close and open mid vowels of English, [e,o] vs. [ɛ,ɔ]. The close vowels generally have lower F1 then the open ones; but this is a slight difference, and it's far from 100% reliable: some [ɛ]'s have lower F1 than [e]'s, even for the same speaker. The height distinction is therefore reinforced by a duration distinction: [e,o] are typically considerably longer than [ɛ,ɔ]. But the duration cue is not reliable either: in fast speech, all vowels shorten, potentially wiping out the difference between long and short vowels. English has one more trick up its sleeve: the close mid vowels are heavily **diphthongized** in most dialects. Words such as *day, fake,* and *so, boat,* are therefore narrowly transcribed [deɪ], [feɪk] and [soʊ], [boʊt] (the extra duration of the close mid vowels is also reflected in this transcription, since there are two vowel symbols rather than one). These three cues, all working together, make the close/open distinction in mid vowels more **robust** (i.e. less likely to be misperceived). Other languages, such as Spanish, avoid these potential problems of variation, and resulting possibilities of miscommunication, by having a simpler vowel system: [i,e,u,o,a]. Since there are no words in Spanish differentiated by the close/open distinction, Spanish speakers' mid vowels can vary between [e] and [ɛ] without risk of confusion.

In many cases, coarticulation and perceptual enhancement are both involved in a particular pattern of variation. The English vowel nasalization coarticulation described above, for example, can also be viewed as a kind of perceptual enhancement: the nasalization of the vowel enhances perception of the following nasal consonant, thus preventing a word such as *bone* from being confused with *bowl* or *bowed*. In sum, this pattern of variation in English can be viewed as having both an articulatory and a perceptual basis:

By allowing the velum to lower sluggishly over the course of the vowel + nasal sequence, rather than abruptly at the beginning of the nasal, less articulatory precision and effort are required.
And by extending the span of the nasal cue into the preceding vowel, perception of the nasal consonant is improved.

B. Phonemic analysis. To help us concisely describe the role of particular cues in particular languages' sound systems, linguists use the following terminology:

A distinction between two sounds (or sequences of sounds) is **phonemic** if it corresponds to a difference in the *meaning* of words, either by itself (e.g. vowel nasalization in French and Dene Sųłine), or as the primary distinction among a set of cues (e.g. the open/close distinction in mid vowels in English).
Otherwise, the distinction is **allophonic** (e.g. vowel duration and diphthongization in mid vowels, and vowel nasalization, in English), from Greek *allo-* 'other' + *phon* 'sound', i.e. a variant sound.

The fact that a particular distinction can be phonemic in one language and allophonic in another, gives rise to a problem for newcomers to a language: how do you determine which distinctions are phonemic? Most language learners eventually figure this out (more or less) through trial and error, with little awareness of what they're trying to learn (particularly in first-language acquisition). Linguists, on the other hand, who are interested in understanding and explicitly describing the structure of languages, tackle this problem using a technique called **phonemic analysis**, examining sets of words – in phonetic transcription, or in spectrograms if more detail is needed – and looking for patterns in the sounds. No further knowledge of the language is required. We will apply this technique to the following data set, from Finnish (the [ː] diacritic indicates that the preceding vowel is long).

(1)	kuːzi	'six'	liːsa	'Lisa'	kade	'envious'
	kadot	'failures'	madon	'of a worm'	kuːzi	'six'
	kate	'cover'	maton	'of a rug'	liːza	'Lisa'
	katot	'roofs'	ratas	'wheel'	radan	'of a track'

Minimal pairs. The first step in solving a phonemic analysis problem is to look for a **minimal pair** in the data set, i.e. two words with different meanings, which are identical except for the phonetic distinction in question: *such a minimal pair establishes that the distinction is phonemic*. Let's say we're interested in determining whether the distinction between [t] and [d] is phonemic in Finnish. In the list above, [kadot] and [katot] are identical except for this very distinction; and these two words clearly have different meanings ('failures' vs. 'roofs'). They therefore count as a minimal pair, establishing that voicing in alveolar stops is phonemic in Finnish. There are other minimal pairs in this data set establishing the same thing ([kate] vs. [kade], [maton] vs. [madon]); but once you've found *one* minimal pair for a given distinction, its phonemic status is conclusively established, and you don't have to look any further.

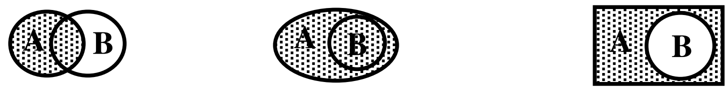
Free variation. Now let's turn to the distinction between [s] and [z] in Finnish. We find the pair [kuːzi] vs. [kuːsi]; but note that they both mean 'six'. That is, we don't have two words with different meanings here, but two transcriptions of the *same word*, with some variation in pronunciation. They are therefore *not* a minimal pair. The same is true for [liːsa] vs. [liːza]. We now have to look for positive evidence of allophonic status. We see in [kuːzi] vs. [kuːsi] that voicing in alveolar fricatives does *not* correspond to a difference in meaning. This kind of allophonic pattern is called **free variation**: either sound is free to occur, in the exact same position in a word, but no difference in meaning results.[6]

Complementary distribution. Recall the discussion of vowel nasalization in English: nasal vowels occur immediately before nasal consonants, and nowhere else; whereas oral vowels can occur everywhere *except* before nasal consonants. That is, one

[6] The diphthongization of English close mid vowels can also be viewed as an allophonic pattern of this type. In normal English, [e] and [o] don't occur by themselves at all; they are always diphthongized to [eɪ] and [oʊ]. Nevertheless, when English speakers hear [med] or [mod] – for example, in foreign-accented English or bad computerized speech – they do not judge these to be unknown words, but (slightly odd) pronunciations of [meɪd] and [moʊd] (i.e. *made* and *mode*). This fact establishes that the distinction between [e,o] and [eɪ,oʊ] is allophonic in English.

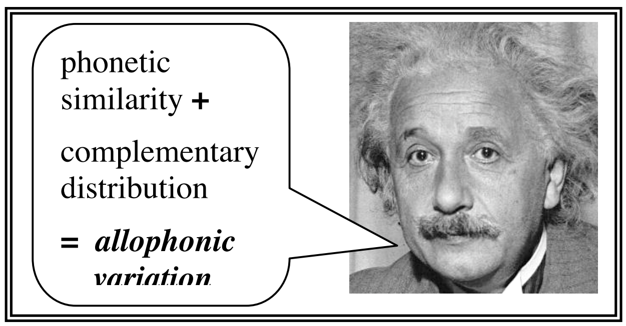
class of sounds occurs in a particular phonetic context, and the other occurs elsewhere. To describe this kind of **distribution** of sounds (i.e. which sounds can occur where), linguists borrow some terminology from set theory in mathematics:

Figure 8: Some set relations



A and B **overlap** B is a **subset** of A A is the **complement** of B

As illustrated in the Fig. 8, the **complement** of a set is *everything that lies outside* that set. Two sounds (or two groups of sounds) are therefore said to be in **complementary distribution** when one group occurs in one set of contexts, and the other group occurs in the complement of contexts – i.e. everywhere else. When two phonetically similar sounds or natural classes are in **complementary distribution**, we may conclude that the distinction between them is **allophonic**.[7] This is clearly the case for nasal and oral vowels in English. You can predict whether any given vowel will be nasal or oral just by knowing the phonetic context it occurs in. The distinction does not correspond to a difference in meaning.



We must further state a **rule** governing the contexts in which each class of allophones occur. We could state that [ĩ,ũ,ɪ̃,ʊ̃,ẽ,ə̃,ɚ̃,õ,ɛ̃,ɔ̃, æ̃,ɐ̃,ã,ɑ̃] occur before [n,m,ŋ], and that [i,u,ɪ,ʊ,e,ə,ɚ,o,ɛ,ɔ,æ,ɐ,a,ɑ] occur elsewhere. But the pattern can be stated more simply and insightfully by referring to the phonetic properties of the natural classes affected by the rule: a vowel is nasal before a nasal consonant, and oral elsewhere. We can express this rule in a visually clear way using the following notation:

vowel → *nasal* / __ *nasal* *vowel* → *oral* / *elsewhere*

vowel → *nasal* should be understood as an implicational statement: if a sound is a vowel, then it is nasalized. Everything following the '/' concerns the *context* in which the rule applies. The blank '__' stands for the position where the sound occurs: '__ *nasal*' means 'before a nasal'; conversely, '*nasal* __' would mean 'after a nasal.'

As a further example, consider the following data set, from Canadian English:

(2)	taɪ	'tie'	həɪk	'hike'	ləɪf	'life'	tɹəɪp	'tripe'
	kəʊtʃ	'couch'	səʊθ	'south'	əɪs	'ice'	ɹaɪð	'writhe'
	ɹəɪt	'right'	spəʊt	'spout'	əlaɪv	'alive'	həʊs	'house'
	tɹaɪb	'tribe'	gaʊdʒ	'gouge'	bɹaʊz	'browse'	laɪm	'lime'
	faʊnd	'found'	haʊ	'how'	faɪl	'file'	maʊnt	'mount'

[7] Note that if the sounds are in complementary distribution, there logically *cannot* be a minimal pair. For there to be a minimal pair, both sounds would have to occur in exactly the same context, in which case the distribution is not complementary, but **overlapping**.

In narrow transcription, the diphthongs /ɑɪ/ and /ɑʊ/ are more accurately transcribed [aɪ] and [aʊ]. That is, they begin as a central low vowel [a], rather than a back low vowel [ɑ]. More interestingly, these diphthongs, [aɪ] and [aʊ], each have an allophonic counterpart, [əɪ] and [əʊ] respectively: these are similar to the diphthongs, but start with a central mid vowel rather than a low one, therefore this pattern is called *raising*.[8] But which allophones occur where? A useful strategy is to make a **context chart** (see Table 7), listing the adjacent sounds for each allophone. (The symbol # indicates a word boundary.) Looking at the sounds *preceding* these diphthongs, there is no common element: we find all manner and place of consonants, or no consonant at all. Looking at the *following* sounds, however, a generalization emerges: the raised diphthongs [əɪ, əʊ] *only* occur before voiceless consonants (circled in Table 7); while [aɪ, aʊ] *never* occur before them – complementary distribution. We can conclude that the distinction between [aɪ, aʊ] and [əɪ, əʊ] is allophonic in Canadian English. The rule can be stated as follows:

Table 7: Context chart for data set (2)

[aɪ] vs. [əɪ]	
[aɪ]	[əɪ]
t__#	ɹ__t
ɹ__b	h__k
l__v	l__f
f__l	#__s
ɹ__ð	ɹ__p
l__m	

[aʊ] vs. [əʊ]	
[aʊ]	[əʊ]
f__n	k__t
ɹ__b	s__θ
m__n	p__t
l__v	#__s
g__d	ɹ__p
h__#	
ɹ__z	

$$\begin{bmatrix} \textit{central} \\ \textit{diphthong} \\ \textit{ending high} \end{bmatrix} \rightarrow \begin{bmatrix} \textit{begins} \\ \textit{mid} \end{bmatrix} \Big/ \_\_\textit{voiceless} \qquad \begin{bmatrix} \textit{central} \\ \textit{diphthong} \\ \textit{ending high} \end{bmatrix} \rightarrow \begin{bmatrix} \textit{begins} \\ \textit{low} \end{bmatrix} \Big/ \textit{elsewhere}$$

Note that this statement of the rule claims that both diphthongs raise before any voiceless sound, though we have no evidence in the data set that [aʊ] raises to [əʊ] before [k], or that [aɪ] fails to raise before [ð]. On the other hand, this rule is not *contradicted* by any of the data; and the broader formulation of the rule is in accordance with our strategy of forming the most general hypothesis that the data permit (see sidebar).

Words enough and time ...

How can you be sure that your analysis wouldn't be contradicted if you just had more data? The best strategy is to form as general and far-reaching a hypothesis about the sound patterns of the language as the current data set permits. *For present purposes, you can assume that any data set you're given is fully representative of the sound patterns of the language.* In real linguistic fieldwork, once you've collected enough words to show each consonant and vowel in a range of positions, you can form a reasonably confident analysis. But you can never be certain that your analysis will hold up in the face of further data. Scientific theories (including linguistic theories) allow us to make predictions about future data, by making sense of the data we have, and assuming future data will behave in the same way. But no scientific theory offers certainty as to how future data will turn out.

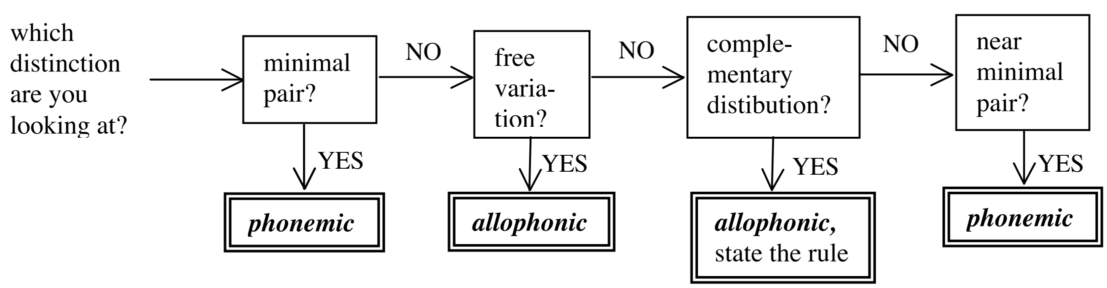
[8] Indeed, this raising, together with use of *eh*, are the two most distinctive features of Canadian English, vs. the dialects of the United States. There is some dialect variation within Canadian English: some speakers raise their diphthongs in other contexts as well.

In fact, this rule is not too broad for English. The vowel + [ɹ] sequences (which could be regarded as a kind of diphthong) are excluded from the rule, by its reference to a *high* vowel as the second half of the diphthong. The rule does not apply to [ɔɪ], [eɪ], or [oʊ], because these are not central. MORAL: you have to think carefully about how you formulate the rule, and the phonetic properties it refers to, so as to include the classes of sounds you want to include, and exclude the others.

The raising rule appears to have an articulatory basis: all vowels in English tend to be significantly shorter before a voiceless consonant. A diphthong beginning with a low vowel and ending with a high vowel involves considerable movement of the tongue body. When the tongue has less time to make this movement, due to the shortening induced by the following voiceless consonant, it 'cheats' by starting from a higher position, [ə] rather than [a]. Observe, however, that we can identify the pattern of allophonic variation without considering its phonetic basis at all.

D. If all else fails. What do you conclude if you can find neither a minimal pair nor evidence of allophonic variation (free variation or complementary distribution)? This is a tricky issue, because there may be a pattern of complementary distribution which you have simply not spotted yet. But assuming that there really is no pattern, it must be the case that you can find, if not an exact minimal pair, then a near minimal pair. For example, imagine that the Finnish data set (1) did not contain the exact minimal pairs [katot] vs. [kadot], [kate] vs. [kade], nor [maton] vs. [madon]. This leaves us with [ratas] vs. [radan]. Though they are not strictly identical but for the [t/d] distinction, this pair shows that [t] and [d] both occur in the context /a__a – that is, the distribution overlaps. Nor is it plausible that the other distinction, the final [s] vs. [n], could play any role in the [t/d] distinction; because we see in other words (e.g. [maton]) that a [t] can occur in a word ending in [n]. The pair [ratas] vs. [radan] can therefore be treated as equivalent to a minimal pair, establishing that the [t] vs. [d] distinction is phonemic in Finnish. Solving a phonemic analysis problem thus involves the following procedure:

Figure 9: Phonemic analysis



Exercises

1. The following data are from North American English. [pʰ,tʰ,kʰ] are *aspirated* allophones of /p,t,k/ respectively (i.e. the stop's release is accompanied by a strong puff of air). Identify the context in which the aspirated stops occur, and state a rule governing their distribution.

tʰæp	'tap'	tʰɪpĩŋ	'tipping'	kʰɹĩm	'cream'	pʰɪkɹ	'picker'
stɑp	'stop'	stup	'stoop'	pʰoʊk	'poke'	tʰɛ̃n	'ten'
skɹĩm	'scream'	tʰwik	'tweak'				

2. The following data are from North American English. [ɫ] is *velarized* allophone of /l/ (i.e. it involves tongue body raising). Identify the contexts in which the plain vs. velarized lateral approximants occur, and state a rule governing their distribution.

fiɫ	'feel'	lod	'load'	dɐɫ	'dull'	mɪɫk	'milk'
slip	'sleep'	p^{h}ɪkɫ	'pickle'	k^{h}ɐɫtʃɹ	'culture'	p^{h}uɫ	'pool'
lɑɹdʒ	'large'	p^{h}lɐ̃m	'plum'	splɪt	'split'	jɛɫ	'yell'
fɑɫtɹ	'falter'	wizɫ	'weasel'	neɪɫ	'nail'	p^{h}æɫ	'pal'

3. The following data are from Québecois French. [y] represents a high front *rounded* vowel (like [i] with your lips in position for [u]). Is the distinction between [i] and [y] phonemic or allophonic in Québecois French? If phonemic, support your answer with examples from the data set. If allophonic, state a rule governing the distribution of [i] and [y].

pətsi	'little'	tsy	'you'	tryi	'sow (pig)'	by	'drank'
trɛ̃ʃ	'slice'	lu	'wolf'	drapo	'flag'	du	'sweet, gentle'
dam	'lady'	bu	'mud'	pɛrdzy	'lost'	tu	'all'
dzimɛ̃ʃ	'Sunday'	dzy	'of'	sortsi	'exit'	ly	'read'
to	'early'	temwɛ̃	'witness'	pɑt	'paw'	kry	'believed'

4. Same data set as the previous question. Is the distinction between [t,d] and [ts,dz] phonemic or allophonic in Québecois French? If phonemic, support your answer with examples from the data set. If allophonic, state a rule governing the distribution of [t,d] vs. [ts,dz].

5. The following data are from Dene Sũłine, an Athabaskan language widely spoken in Northwestern Canada. [ɬ,ɣ] are voiceless lateral and voiced velar fricatives respectively. [t',k',ts',tθ',tɬ'] are *ejective* stops and affricates. Vowels marked with [´] are pronounced with a high tone. Determine whether the ejective vs. aspirated distinction in the stops and affricates is phonemic or allophonic in Dene. If phonemic, support your answer with examples from the data set. If allophonic, state a rule governing the distribution of ejection vs. aspiration.

t^{h}u θe k^{h}ɑ̃	'there's the water'	t^{h}ɛn	'ice'
tɬ^{h}ɛs	'lard/oil'	bɛk'ṍ ɬ̃ilijɑw	'don't you know that one?'
nɑ́h̃ilt'ɪ	'2 people'	tθ'ə́i	'cup'
k'ɑbĩ́	'morning'	ts'i	'porcupine'
k'oaθ	'cloud'	bɑneɬt'u ɬɑɣɑ̃ldɛ́	'they both got killed'
k^{h}oə́n	'fire'	sɑs jadak^h	'a bear killed (someone) '
t^{h}ot'inɛ́	'English'	nɑk^{h}ɛ	'two'
k'i	'birch'	t^{h}ɑɣɛ	'three'
tθ^{h}e	'rock'	k'ɛt^{h}ɑɣɛ	'six'
tɬ'izɛ	'horse fly'	ɛdinɛk'ɑ	'I am fat'
t^{h}iɛt'ɛɣɛ	'night'	tθ'ɛn	'bone'

VI. Phonotactics

A. Possible and impossible words. Allophonic rules govern the distribution of variants (allophones) of the basic sounds of a language. But there are also phonological rules that restrict how even the basic sounds of the language can be assembled into words. Perhaps the clearest way to demonstrate the existence of such rules is by considering the **possible words** of a language. For example, among the many thousands of words of English, there happens *not* to be a word *spink* [spĩŋk]. But English speakers would generally agree that it *could* be a word. For example, one might name a new toy, or a newly discovered sub-atomic particle, a *spink*, and English speakers would easily accept and use this new word. On the other hand, something like [tftkt] could not possibly be a word of English. It's not that [tftkt] is physically unpronounceable: in fact, it's an actual word (it means 'you sprained') in Tashlhiyt Berber, a language of North Africa. Nor is the unacceptability of [tftkt] due to any of the allophonic rules of English: [t^hftkt] (with allophonic aspiration of the initial /t/) is still unacceptable. The distinct status of [spĩŋk] (a non-occurring but possible word) vs. [tftkt] demonstrates

> that there are certain phonological rules which English words conform to, above and beyond patterns of allophonic variation;
> that these rules are different from those of other languages (such as Berber); and
> that English speakers are in some sense *aware* of these rules, as reflected in consistent judgments about possible vs. impossible words.

These sorts of rules, concerning how the sounds can be sequenced to form possible words of a language, are known as **phonotactic rules** (or simply 'phonotactics,' from *phon* 'sound' + Latin *-tact-* 'touching'). Possible words, which obey the phonotactics, are **well-formed**; while the remaining sequences of sounds are **ill-formed**.

What are the phonotactic rules to which English words must conform? For starters, words must contain at least one vowel, a rule which [tftkt] obviously violates. Moreover, words cannot begin with a sequence of stops: indeed, a word can begin with no more than two voiceless consonants: either an affricate, or an [s] + stop sequence, as in [spɪŋk]. In addition, note that

> [spɪŋk] is well-formed (while *[spɪmk] and *[spɪnk] are not; the '*' indicates ill-formedness). Similarly,
> [spɪnt], [spɪmp] (vs. *[spɪmt], *[spɪŋt], *[spɪŋp], *[spɪnp]).

The generalization here is that within words, a nasal + stop sequence must have the *same place of articulation*: bilabial [mp], alveolar [nt], or velar [ŋk]. These are but a few examples of English phonotactic rules.

Compared to Tashlhiyt Berber, English phonotactics seem rather strict. But compared to Japanese, English seems quite permissive. In Japanese, words can begin with no more than one consonant; words must end in a vowel or nasal; and the only permissable word-internal consonant sequences are double (**geminate**) consonants (e.g. [tootte] 'passing', [nippoɴ] 'Japan', [gakkoo] 'school', and nasal + stop or fricative sequences with the same place of articulation (like English) (e.g. [tombo] 'dragonfly', [kande] 'teaching', [kaŋkee] 'relation', [sensee] 'teacher'). Thus, when English words are borrowed into Japanese, they are adapted to Japanese phonotactics, e.g. [sɯtoɾaikɯ] ('strike'). ([ɯ] = unrounded [u], [ɾ] = a 'flapped' [d] (i.e. very brief closure), [ɴ] = uvular nasal).

As we've seen from this brief glimpse at English and Japanese, phonotactic rules, like allophonic rules, refer to phonetically defined natural classes of sounds: nasals, stops, bilabials, etc., not to arbitrary collections of sounds such as [m,j,θ,e]. The rule requiring nasals to be at the same place of articulation as the following stop can be expressed as follows:

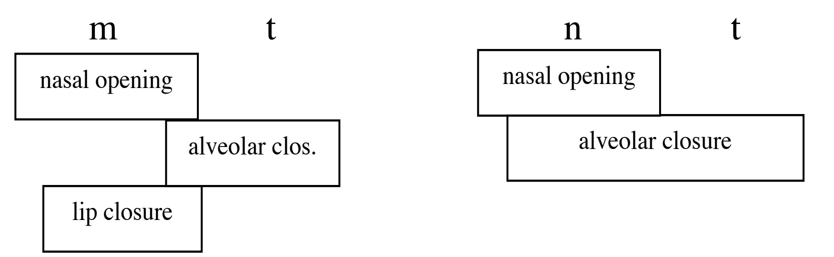
$$nasal \rightarrow place_i / \_\_ \begin{bmatrix} consonant \\ place_i \end{bmatrix}$$

(The co-indexation of the *place* variable in the two parts of the rule mean that the place of the nasal must match the place of the following consonant.)

B. Why do languages have phonotactic rules? As with allophonic rules, phonotactic rules can plausibly be viewed as set of trade-offs and strategies for satisfying the two functional constraints on language sound systems: Ease of Articulation and Ease of Perception. For example, why might Japanese and English place restrictions on consonant sequences within words? Vowels are typically the loudest part of the sound signal, and the perception of most consonants depends on, or is aided by, formant transitions in adjacent (or at least nearby) vowels. Requiring vowels to be regularly interspersed among the consonants, i.e. placing limits on consonant sequences, thus improves the consonants' perceptibility. Tashlhiyt Berber represents the extreme end of the spectrum, in terms of languages' tolerance for consonant sequences; but even in this language, *most* of the words do have vowels interspersed among the consonants. Thus, there are languages which place strict conditions on sequences of consonants (some even stricter than Japanese), there are languages such as English, which tolerate a broader range of consonant sequences, or at the extreme end, Tashlhiyt Berber. But no language *prohibits* vowels from being interspersed among consonants.

Similarly, the requirement of shared place of articulation in nasal + stop sequences, seen in both English and Japanese, can be understood as a response to Ease of Articulation. Presumably, more energy is required to produce, e.g., an [mt] sequence – with *two* closures, than an [nt], with only a single closure (see Fig. 8). There is a perceptual side to this story as well. The cues to place of articulation in a nasal are relatively weak before a stop, due to the absence of formant transitions into a following vowel. Since the place cues to the nasal in this position are weak to begin with, the phonological system 'decides' (so to speak), that maintaining a distinct place of articulation in the nasal, is not worth the extra articulatory effort it would require.

Figure 8: Closures in nasal + stop sequences



Exercises

1. The following data are from Chumash, an indigenous language of Southern California, now extinct. [q] = back (uvular) [k]. [k’,ts’,tʃ’] = *ejective* stops and affricates

(release is accompanied by a 'pop', caused by shutting and raising the larynx during closure). Identify the phonotactic rule concerning multiple fricatives within a word.

osos	'heel'	ats’is	'beard'	ʃiʃ	'gopher hole'
pʃoʃ	'snake'	ʃoqʃ	'gall'	itʃ’itʃ	'young sibling'
jasis	'poison oak'	ʃoʃo	'squirrel'	tʃ’ijuʃ	'break wind'
katskaw	'I sin'	ʃiʃk’ij	'it aches'	skinus	'I saved it for him'

2. The following data are from Russian. [x] = voiceless velar fricative; [ʲ] indicates a palatalized preceding consonant has a palatalized (j-like) release. Identify the phonotactic rule concerning voicing in word-final consonants and consonant sequences.

trut	'labour'	mox	'moss'	rof	'ditch'
krofʲ	'blood'	slavarʲ	'dictionary'	dʲenʲ	'day'
doʃtʲ	'rain'	atʲets	'father'	pədarək	'gift'
ʒivot	'belly'	rot	'mouth'	fkus	'taste'
snʲek	'snow'	stol	'table'	zup	'tooth'
vrak	'enemy'	platok	'kerchief'	kalxos	'collective farm'
garaʃ	'garage'	kavʲor	'rug'	muzʲej	'museum'

VI. Alternations

Up till now, we have been been concerned with distributional patterns – statements about what sounds can occur in what contexts. We see phonological rules applying more 'actively,' however, in **alternations**. These concern changes to a particular word's pronunciation depending on the phonetic context in which it occurs. For example, in North American English, the final consonant in *beat* changes from [t] to [ɾ] (an alveolar flap, cf. Japanese examples in sec. 4.2), when the *-ing* **suffix** (ending) is added: [biɾĩŋ]. In fact, as a result of this alternation, *beat* becomes indistinguishable, in most dialects, from *bead* when *-ing* is added, for the [d] also changes to [ɾ] in this context.[9] These alternations are the result of a general rule of North American English, whereby alveolar stops ([t,d]) are 'flapped' (voiced and shortened) when they occur between two vowels, and the first vowel is stressed. The rule can be expressed as follows:

$$\begin{bmatrix} \textit{alveolar} \\ \textit{stop} \end{bmatrix} \rightarrow \textit{flap} \Big/ \begin{bmatrix} \textit{stressed} \\ \textit{vowel} \end{bmatrix} \_\_ \textit{vowel} \qquad \begin{bmatrix} \textit{alveolar} \\ \textit{stop} \end{bmatrix} \rightarrow \begin{bmatrix} \textit{not} \\ \textit{flap} \end{bmatrix} \Big/ \textit{elsewhere}$$

That is, in other contexts ('elsewhere'), the stop remains either a [t] or [d] (or another allophone thereof, such as [tʰ]). Thus [’f**oʊ**ɾəgɹæf] ('photograph'), but [fə’tʰ**ɑ**gɹəfi] ('photography'); [’ɹɛɾi] ('ready'), but [ɹi’d**i**m] ('redeem').

Of further interest is the finding that English speakers readily apply this rule to words that they've never heard before. For example, let's introduce another possible word, [klɛt]; assume that it means 'to smell mouldy or rotten'. Example: *Jeez Tom, you* [klɛt] *like a dead vulture!* What's the *-ing* form of this verb? If you're like most native speakers of North American English, you would say that Tom is [klɛɾĩŋ] rather than [klɛtĩŋ]. While it's quite possible that you've heard the words *beating, beading, photography,* etc.

[9] If a distinction does remain, it's probably in the preceding vowel (slightly longer in *beading*, cf. the observation in sec. V that vowels are shorter before voiceless consonants), rather than in the consonant itself.

before, and therefore learned their flapped pronunciations by direct imitation, it is quite impossible that you've ever heard [klɛɾɪ̃ŋ] before. So how did you know that it's [klɛɾɪ̃ŋ] rather than [klɛtɪ̃ŋ]? (It's not that some external authority prescribes that [klɛɾɪ̃ŋ] is the 'correct' pronunciation; it's that speakers of North American English would overwhelmingly converge on this same pronunciation.) This result demonstrates the psychological reality of the flapping rule: North American English speakers actively (albeit unconsciously) apply this rule to the words that they come up with in the course of speaking. Whereas distributional patterns show the *effects* of the phonological system on the words of the language, alternations catch the phonological system red-handed, so to speak, in the very act of applying to new words.

Finally, note that alternations are not a different *kind* of rule from the phonotactic and allophonic rules discussed in previous sections. Alternations are a *result* of phonotactic and/or allophonic rules. Indeed, the flapping rule above is allophonic, in that it governs the allophonic distinction between [t] and [ɾ] (and also between [d] and [ɾ]).[10] Rather, alternations provide an additional source of data, and an additional analytic technique, for discovering the rules of a language's phonological system. The technique is as follows:

> For a given set of related words, i.e. words containing some identifiable, meaningful common subpart (e.g. {*cat, cats, catty*}; {*photograph, photography, photographed*}; {*reread, replay, resettle*}, etc.), identify a **basic form** of the **stem** (the main part of the word), and of the suffixes (or prefixes, e.g. *re-* in *reread*).
>
> ⇒ For present purposes, we can equate the basic form of the stem with its pronunciation in the absence of suffixes or prefixes. The stem in *hitting*, for example, is [hɪɾ], but its basic form is [hɪt].
>
> ⇒ The basic form of a suffix or prefix can be equated with its pronunciation in the broadest range of contexts in which it occurs. For example, the basic form of the prefix seen in {*indiscreet, inherent, inactive, imprecise, imbalance, incredible, ingratitude*} is [ɪ̃n], which occurs in all contexts except before bilabial stops [p,b] (where we get [ɪ̃m]) and velar stops [k,g] (where we get [ɪ̃ŋ]).
>
> Whenever the resulting word (e.g. [hɪɾɪ̃ŋ]) differs from the basic form of the stem and the basic form of any suffix or prefix therein ([hɪt]+[ɪ̃ŋ]), identify a phonotactic or allophonic rule (or set of rules) to account for the alternation(s). The alternations in the *in-* prefix, for example, can be attributed to the phonotactic rule identified in sec. VI, requiring nasal + stop clusters to have the same place of articulation.

As a further example, consider the following data, from Dutch. The diminutive suffix indicates an attitude of endearment toward the noun, similar to English *-y* as in *birdy* or *sonny*. [c] = a voiceless palatal stop, somewhere between a [k] and a [t] with a [j]-like release. [ø] = rounded [e].

[10] Its status as an allophonic rule is complicated somewhat by the fact [t] and [d] both have [ɾ] as an allophone; thus the flapping rule neutralizes the distinction between [t] and [d] in the /stressed vowel__vowel context.

(3)

Noun	Diminutive		Noun	Diminutive	
zoːn	zoːncə	‘son’	døːr	døːrcə	‘door’
leːpəl	leːpəlcə	‘spoon’	deːkən	deːkəncə	‘blanket’
zak	zakjə	‘bag’	buːk	buːkjə	‘book’
briːf	briːfjə	‘letter’	sxip	sxipjə	‘ship’
boːt	boːcə	‘boat’	kat	kacə	‘cat’

We can equate the base form of the noun stem in the diminutive with the bare noun (the first column). Looking down the second column, however, we see two forms of the diminutive suffix: [cə] and [jə]. A context chart would show that [cə] occurs after base forms ending in [n,l,t,r]; while [j] occurs after base forms ending in [k,f,p]. Although there are more *examples* with [cə] than with [jə], the *contexts* for [jə] includes velars, bilabials, and labiodentals. In contrast, the context for [cə] boils down to a single natural class: alveolars. Since [jə] occurs in a broader range of contexts, it is the basic form of the suffix. We can now state a rule:

palatal → *stop* / *alveolar* __ *palatal* → *approximant* / *elsewhere*

This accounts for the alternations in the nouns ending in [n,l,r] (the alveolar nasal and approximants). However, it incorrectly predicts, e.g., [kat]+[jə] → [katcə], whereas the actual word is [kacə]. This problem can be addressed with a further rule:

$$\begin{bmatrix} \textit{alveolar} \\ \textit{stop} \end{bmatrix} \rightarrow \varnothing \Big/ \_ \begin{bmatrix} \textit{palatal} \\ \textit{stop} \end{bmatrix}$$

That is, [t] deletes (i.e. it alternates with *nil*) when it precedes a palatal stop.

Exercises

1. The following data are from English. State a rule to ccount for the alternations.

Noun	Plural		Noun	Plural	
bæk	bæks	'back'	fli	fliz	'flea'
flæŋk	flæŋks	'flank'	deɪ	deɪz	'day'
hɔɚd	hɔɚdz	'hoard'	kʰlu	kʰluz	'clue'
dɔg	dɔgz	'dog'	hɛ̃m	hɛ̃mz	'hem'
ɹəɪt	ɹəɪts	'right'	wɪg	wɪgz	'wig'
veɪɫ	veɪɫz	'veil'	sɔŋ	sɔŋz	'song'
slip	slips	'sleep'	fɚ	fɚz	'fur'
bɹaɪb	bɹaɪbz	'bribe'	stɪk	stɪks	'stick'

2. The following data are from Karok (slightly simplified), an indigenous language of Central California. [ʔ] represents a glottal stop (the consonant in the middle of English interjection *uh-oh*). Identify the basic forms of the stems and prefixes, and state rules to account for any alternations. Are the rules phonotactic or allophonic?

Imperative	I-*form*	You-*form*	
pasip	nipasip	ʔupasip	'shoot'
siːtva	niʃiːtva	ʔusiːtva	'steal'
kifnuk	nikifnuk	ʔukifnuk	'stoop'
suprih	niʃuprih	ʔusuprih	'measure'
ʔifik	niʔifik	ʔuʔifik	'pick up'
ʔaktuv	niʔaktuv	ʔuʔaktuv	'pluck at'
axyar	nixyar	ʔuxyar	'fill'
iʃkak	niʃkak	ʔuskak	'jump'
iʃriv	niʃriv	ʔusriv	'shoot at a target'

Further reading

Kenstowicz, Michael and Charles Kisseberth (1979) Generative Phonology. Academic Press.

Ladefoged, Peter (2001) A Course in Phonetics (4th Ed.), Harcourt Brace.

Ladefoged, Peter (1996) Elements of Acoustic Phonetics (2nd Ed.). University of Chicago Press.

Glossary

Acoustic: Pertaining to the properties of soundwaves.

Affricates: A **stop+fricative** sequence, made with the same **articulator**, sometimes treated as a single consonant.

Allophonic: Different in pronunciation, but not indicating a difference in meaning (cf. **phonemic**).

Alternations: Changes to a word's pronunciation depending on the phonetic context in which it occurs.

Alveolar ridge: The ridge of gum-covered bone behind the upper teeth.

Amplitude: Loudness.

Aperiodic: Without any repeating pattern, characteristic of the waveforms of **fricatives** and **stop** releases.

Approximant: A **consonant** produced with less constriction than that of a **fricative**, but more than a **vowel.**

Articulator: An organ of the **vocal tract** used in speech production.

Basic form: The form of a **stem** (or **prefix/suffix**) prior to undergoing any **alternations**.

Bilabial: Produced with the two lips.

Broad transcription: A less detailed phonetic transcription, reflecting only **phonemic** distinctions, enclosed in /slashes/ (cf. **Narrow transcription**).

Central: Produced with the **tongue body** neither forward nor retracted.

Close: With greater constriction, opposite of **open**.

Coarticulation: Overlapping movements of the **articulators**.

Complementary distribution: When one sound occurs in one context only, and another sound never occurs in that context, only occurring elsewhere.

Complex waveform: A waveform containing a number of component **frequencies**.

Consonant: A speech sound involving significant obstruction of airflow.

Cues: Properties of the **acoustic** signal, used in recognizing speech.

Cycles per second: A measure of **frequency** of a soundwave, also called **Hertz.**

Decibels: A measure of loudness (abreviated dB).

Dental: Produced with the tongue tip and the upper teeth.

Diacritic: A supplementary phonetic symbol, usually appearing above or below the main symbol.

Dialect: A regional variant of a language.

Distribution: Where things are found: specifically, the phonetic contexts in which a given set of speech sounds occur.

Formants: Peaks in the **spectrum** of a **complex waveform**.

Fourier analysis: A mathematical technique of breaking complex waveforms down into their component frequencies, used in spectrograms.

Free variation: A kind of **allophonic** variation, where either allophone can be used in a given context without affecting the meaning of the word.

Frequency: The number of cycles of a periodic wave, heard as pitch.

Fricative: A consonant produced with a **close** but incomplete constriction, resulting in a hissing noise.

Fundamental frequency: The lowest **frequency** component of a **complex waveform**, heard as the basic pitch of the speakers voice, also called **F0.**

Geminate: A consonant maintained for roughly twice the normal duration of the corresponding single consonant.

Glottis: The space between the folds of the **larynx**.

Harmonics: Higher-**frequency** components of a **complex waveform** (cf. **fundamental frequency**).

Hertz: Cycles per second, a measure of frequency (abbreviated Hz).

Ill-formed: Violating the phonological rules of a language.

International Phonetic Alphabet (IPA): A convention for phonetic transcription, widely used by linguists.

Labiodental: Produced with the lower lip and upper teeth.

Larynx: The valve at the top of the **trachea**, the source of **voicing**.

Lateral: Produced with lowering of the side(s) of the tongue.

Manner of articulation: The degree of obstruction of airflow involved in a given **consonant**.

Mid: Produced with the **tongue body** neither high nor low.

Minimal pair: Two words with different meanings, which are identical except for the phonetic distinction in question, used to establish the **phonemic** status of the phonetic distinction.

Narrow transcription: A fully detailed phonetic transcription, reflecting **allophonic** variation, enclosed in [square brackets].

Nasalized: Produced with lowering of the velum, allowing air to flow through the nasal passages.

Natural class of sounds: A set of **sounds** within a given language which can be defined in terms of some shared phonetic property or properties.

Obstruction: Blockage, specifically blockage of airflow in the vocal tract.

Open With less constriction, opposite of **close**.

Palatal: Produced by the tongue body in the region of the **palate.**

Palate: The roof of the mouth, commonly called the 'hard palate.'

Periodic: Characterized by a regular, repeating pattern. Periodic waveforms have a 'humming' sound.

Pharynx: The back of the throat.

Phonation: The state of the **larynx** during a speech **sound**.

Phonemic: A distinction between two **sounds** (or sequences of sounds) which corresponds to a difference in the *meaning* of words, either by itself, or as the primary distinction among a set of cues.

Phonotactic rules: Rules restricting how **sounds** can be combined to form words within a given language.

Place of articulation: The location of a consonant's **obstruction** in the vocal tract.

Possible word: A nonsense word which satisfies the phonological rules of the language.

Post-alveolar: Produced with the tongue tip in the region behind the **alveolar ridge**.

Pure tone: Sound energy characterized by a simple sine wave, approximated by the sound of a tuning fork.

Robust: As applied to phonemic distinctions, unlikely to be misperceived, due to strong cues.

Rounding: Drawing together of the corners of the lips, as in **rounded vowels.**

Sound: A portion of the speech signal during which the sound energy (and the configuration of the mouth to produce that sound energy) remains relatively stable.

Spectrogram: A visual display of sound energy, showing how the **spectrum** changes over time.

Spectrum: The **amplitude** profile of the **harmonics** of a **complex waveform**.

Stops: Consonants produced with complete closure of the **vocal tract**.

Stress: Greater loudness, duration and pitch of particular vowels within words.

Striation: A stripe-like pattern.

Suffix: A word 'ending' with a recognizable meaning, such as the *-s* at the end of *cats*.

Syllable: 'Mini-words' into which longer words can be broken down, each consisting of a single vowel (or diphthong), together with any consonants that can be grouped with it.

Trachea: The 'windpipe,' lead from the through down to the lungs.

Uvula: The fleshy appendage at the back of the **velum**.

Velo-pharyngeal port: The space between the **velum** and the **pharynx**, leading into the nasal passages.

Velum: The 'soft palate'.

Vocal tract: The lungs, throat, mouth and nose, particularly as used in speech.

Voicing: Pulsing of air in the **glottis** as it passes through the vibrating **larynx**.

Vowel: Sounds produced without significant **obstruction** of airflow in the **vocal tract**.

Well-formed: Obeying the phonological rules of the language.

THE INTERNATIONAL PHONETIC ALPHABET (revised to 1993)

CONSONANTS (PULMONIC)

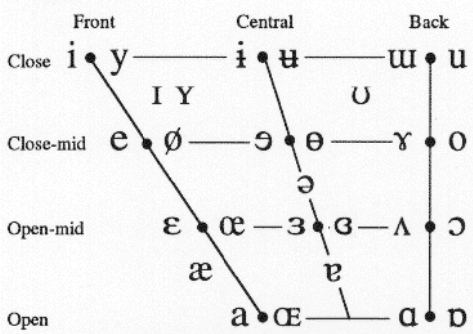
	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap				ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant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

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

CONSONANTS (NON-PULMONIC)

Clicks	Voiced implosives	Ejectives
		ʼ as in:
ʘ Bilabial	ɓ Bilabial	pʼ Bilabial
ǀ Dental	ɗ Dental/alveolar	tʼ Dental/alveolar
ǃ (Post)alveolar	ʄ Palatal	kʼ Velar
ǂ Palatoalveolar	ɠ Velar	sʼ Alveolar fricative
ǁ Alveolar lateral	ʛ Uvular	

VOWELS



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

OTHER SYMBOLS

ʍ Voiceless labial-velar fricative

w Voiced labial-velar approximant

ɥ Voiced labial-palatal approximant

ʜ Voiceless epiglottal fricative

ʢ Voiced epiglottal fricative

ʡ Epiglottal plosive

ɕ ʑ Alveolo-palatal fricatives

ɺ Alveolar lateral flap

ɧ Simultaneous ʃ and x

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

k͡p t͡s

SUPRASEGMENTALS

ˈ Primary stress

ˌ Secondary stress ˌfoʊnəˈtɪʃən

ː Long eː

ˑ Half-long eˑ

˘ Extra-short ĕ

. Syllable break ɹi.ækt

ǀ Minor (foot) group

ǁ Major (intonation) group

‿ Linking (absence of a break)

TONES & WORD ACCENTS

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

DIACRITICS

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

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

This chart is reprinted here by permission of the International Phonetic Association.