

discuss computational models of **phonological learning**: how phonological rules can be automatically induced by machine learning algorithms.

Finally, we apply the transducer-based model of phonology to an important problem in text-to-speech systems: mapping from strings of letters to strings of phones. We first survey the issues involved in building a large pronunciation dictionary, and then show how the transducer-based lexicons and spelling rules of Chapter 3 can be augmented with pronunciations to map from orthography to pronunciation.

This chapter focuses on the non-probabilistic areas of computational linguistics and pronunciations modeling. Chapter 5 will turn to the role of probabilistic models, including such areas as probabilistic models of pronunciation variation and probabilistic methods for learning phonological rules.

4.1 SPEECH SOUNDS AND PHONETIC TRANSCRIPTION

The study of the pronunciation of words is part of the field of **phonetics**, the study of the speech sounds used in the languages of the world. We will be modeling the pronunciation of a word as a string of symbols which represent **phones** or **segments**. A phone is a speech sound; we will represent phones with phonetic symbols that bears some resemblance to a letter in an alphabetic language like English. So for example there is a phone represented by *l* that usually corresponds to the letter *l* and a phone represented by *p* that usually corresponds to the letter *p*. Actually, as we will see later, phones have much more variation than letters do. This chapter will only briefly touch on other aspects of phonetics such as **prosody**, which includes things like changes in pitch and duration.

This section surveys the different phones of English, particularly American English, showing how they are produced and how they are represented symbolically. We will be using two different alphabets for describing phones. The first is the **International Phonetic Alphabet (IPA)**. The IPA is an evolving standard originally developed by the International Phonetic Association in 1888 with the goal of transcribing the sounds of all human languages. The IPA is not just an alphabet but also a set of principles for transcription, which differ according to the needs of the transcription, so the same utterance can be transcribed in different ways all according to the principles of the IPA. In the interests of brevity in this book we will focus on the symbols that are most relevant for English; thus Figure 4.1 shows a subset of the IPA symbols for transcribing consonants, while Figure 4.2 shows a subset of the IPA

PHONETICS

PHONES

IPA

IPA Symbol	ARPAbet Symbol	Word	IPA Transcription	ARPAbet Transcription
[p]	[p]	<u>p</u> arsley	['pɑrsli]	[p aa r s l iy]
[t]	[t]	<u>t</u> arragon	['tærəɡɑn]	[t ae r ax g aa n]
[k]	[k]	<u>c</u> atnip	['kætnɪp]	[k ae t n ix p]
[b]	[b]	<u>b</u> ay	[beɪ]	[b ey]
[d]	[d]	<u>d</u> ill	[dɪl]	[d ih l]
[g]	[g]	<u>g</u> arlic	['gɑrlɪk]	[g aa r l ix k]
[m]	[m]	<u>m</u> int	[mɪnt]	[m ih n t]
[n]	[n]	<u>n</u> utmeg	['nʌtmɛɡ]	[n ah t m eh g]
[ŋ]	[ŋɡ]	<u>g</u> inseng	['dʒɪnsɪŋ]	[jh ih n s ix ŋɡ]
[f]	[f]	<u>f</u> ennel	['fɛnl]	[f eh n el]
[v]	[v]	<u>c</u> loye	[kloʊv]	[k l ow v]
[θ]	[th]	<u>t</u> histle	['θɪsl]	[th ih s el]
[ð]	[dh]	<u>h</u> eather	['hɛðə]	[h eh dh axr]
[s]	[s]	<u>s</u> age	[seɪdʒ]	[s ey jh]
[z]	[z]	<u>h</u> azelnut	['heɪzlnʌt]	[h ey z el n ah t]
[ʃ]	[sh]	<u>s</u> quash	[skwɑʃ]	[s k w a sh]
[ʒ]	[zh]	<u>a</u> mbrosia	[æm'brʊʒə]	[ae m b r ow zh ax]
[tʃ]	[ch]	<u>c</u> hicory	['tʃɪkəri]	[ch ih k axr iy]
[dʒ]	[jh]	<u>s</u> age	[seɪdʒ]	[s ey jh]
[l]	[l]	<u>l</u> icorice	['lɪkəriʃ]	[l ih k axr ix sh]
[w]	[w]	<u>k</u> iwi	['kiwi]	[k iy w iy]
[r]	[r]	<u>p</u> arsley	['pɑrsli]	[p aa r s l iy]
[j]	[y]	<u>y</u> ew	[ju]	[y uw]
[h]	[h]	<u>h</u> orseradish	['hɔrsrædɪʃ]	[h ao r s r ae d ih sh]
[ʔ]	[q]	uh-oh	[ʔʌʔoʊ]	[q ah q ow]
[ɹ]	[dx]	<u>b</u> utter	['bʌɹə]	[b ah dx axr]
[ɹ̃]	[nx]	<u>w</u> intergreen	[wɪɹ̃əɡrɪn]	[w ih nx axr g r i n]
[l]	[el]	<u>t</u> histle	['θɪsl]	[th ih s el]

Figure 4.1 IPA and ARPAbet symbols for transcription of English consonants.

symbols for transcribing vowels.¹ These tables also give the ARPAbet symbols; ARPAbet (Shoup, 1980) is another phonetic alphabet, but one that is specifically designed for American English and which uses ASCII symbols;

¹ For simplicity we use the symbol [r] for the American English “r” sound, rather than the more-standard IPA symbol [ɹ].

it can be thought of as a convenient ASCII representation of an American-English subset of the IPA. ARPAbet symbols are often used in applications where non-ASCII fonts are inconvenient, such as in on-line pronunciation dictionaries.

IPA Symbol	ARPAbet Symbol	Word	IPA Transcription	ARPAbet Transcription
[i]	[iy]	lily	['li]	[l ih l iy]
[ɪ]	[ih]	lily	['li]	[l ih l iy]
[eɪ]	[ey]	daisy	['deɪzi]	[d ey z i]
[ɛ]	[eh]	poinsettia	[pɔɪn'sɛɹiə]	[p oy n s eh dx iy ax]
[æ]	[ae]	aster	['æstə]	[ae s t axr]
[ɑ]	[aa]	poppy	['pɑpi]	[p aa p i]
[ɔ]	[ao]	orchid	['ɔrkɪd]	[ao r k ix d]
[ʊ]	[uh]	woodruff	['wʊdrʌf]	[w uh d r ah f]
[oʊ]	[ow]	lotus	['ləʊtəs]	[l ow dx ax s]
[u]	[uw]	tulip	['tulɪp]	[t uw l ix p]
[ʌ]	[ah]	buttercup	['bʌtərkʌp]	[b ah dx axr k ah p]
[ɜ]	[er]	bird	['bɜd]	[b er d]
[aɪ]	[ay]	iris	['aɪrɪs]	[ay r ix s]
[aʊ]	[aw]	sunflower	['sʌnflaʊə]	[s ah n f l aw axr]
[ɔɪ]	[oy]	poinsettia	[pɔɪn'sɛɹiə]	[p oy n s eh dx iy ax]
[ju]	[y uw]	feverfew	['fɪvəfju]	[f iy v axr f y uw]
[ə]	[ax]	woodruff	['wʊdrəf]	[w uh d r ax f]
[i]	[ix]	tulip	['tulɪp]	[t uw l ix p]
[ɜ]	[axr]	heather	['hɛðə]	[h eh dh axr]
[ʌ]	[ux]	dude ²	[dʌd]	[d ux d]

Figure 4.2 IPA and ARPAbet symbols for transcription of English vowels.

Many of the IPA and ARPAbet symbols are equivalent to the Roman letters used in the orthography of English and many other languages. So for example the IPA and ARPAbet symbol [p] represents the consonant sound at

² The last phone, [ʌ]/[ux], is quite rare in general American English and indeed is an “extension” not present in the original ARPAbet. Labov (1994) notes that the realization of a fronted [uw] as [ux] has made it more common in (at least) Western and Northern Cities dialects of American English starting in the late 1970s. This fronting was first called to public by imitations and recordings of ‘Valley Girls’ speech by Moon Zappa (Zappa and Zappa, 1982). Nevertheless, for most speakers [uw] is still much more common than [ux] in words like *dude*.

the beginning of *platypus*, *puma*, and *pachyderm*, the middle of *leopard*, or the end of *antelope* (note that the final orthographic *e* of *antelope* does not correspond to any final vowel; the *p* is the last sound).

The mapping between the letters of English orthography and IPA symbols is rarely as simple as this, however. This is because the mapping between English orthography and pronunciation is quite opaque; a single letter can represent very different sounds in different contexts. Figure 4.3 shows that the English letter *c* is represented as IPA [k] in the word *cougar*, but IPA [s] in the word *civet*. Besides appearing as *c* and *k*, the sound marked as [k] in the IPA can appear as part of *x* (*fox*), as *ck* (*jackal*), and as *cc* (*raccoon*). Many other languages, for example Spanish, are much more transparent in their sound-orthography mapping than English.

Word	jackal	raccoon	cougar	civet
IPA	[ˈdʒæk.l]	[ræ.kun]	[ˈku.gə]	[ˈsɪ.vɪt]
ARPAbet	[j h æ k e l]	[r æ k u w n]	[k u w g a x r]	[s i h v i x t]

Figure 4.3 The mapping between IPA symbols and letters in English orthography is complicated; both IPA [k] and English orthographic [c] have many alternative realizations.

The Vocal Organs

ARTICULATORY
PHONETICS

We turn now to **articulatory phonetics**, the study of how phones are produced, as the various organs in the mouth, throat, and nose modify the airflow from the lungs.

Sound is produced by the rapid movement of air. Most sounds in human spoken languages are produced by expelling air from the lungs through the windpipe (technically the **trachea**) and then out the mouth or nose. As it passes through the trachea, the air passes through the **larynx**, commonly known as the Adam’s apple or voicebox. The larynx contains two small folds of muscle, the **vocal folds** (often referred to non-technically as the **vocal cords**) which can be moved together or apart. The space between these two folds is called the **glottis**. If the folds are close together (but not tightly closed), they will vibrate as air passes through them; if they are far apart, they won’t vibrate. Sounds made with the vocal folds together and vibrating are called **voiced**; sounds made without this vocal cord vibration are called **unvoiced** or **voiceless**. Voiced sounds include [b], [d], [g], [v], [z], and all the English vowels, among others. Unvoiced sounds include [p], [t], [k], [f], [z], and others.

GLOTTIS

VOICED

UNVOICED

VOICELESS

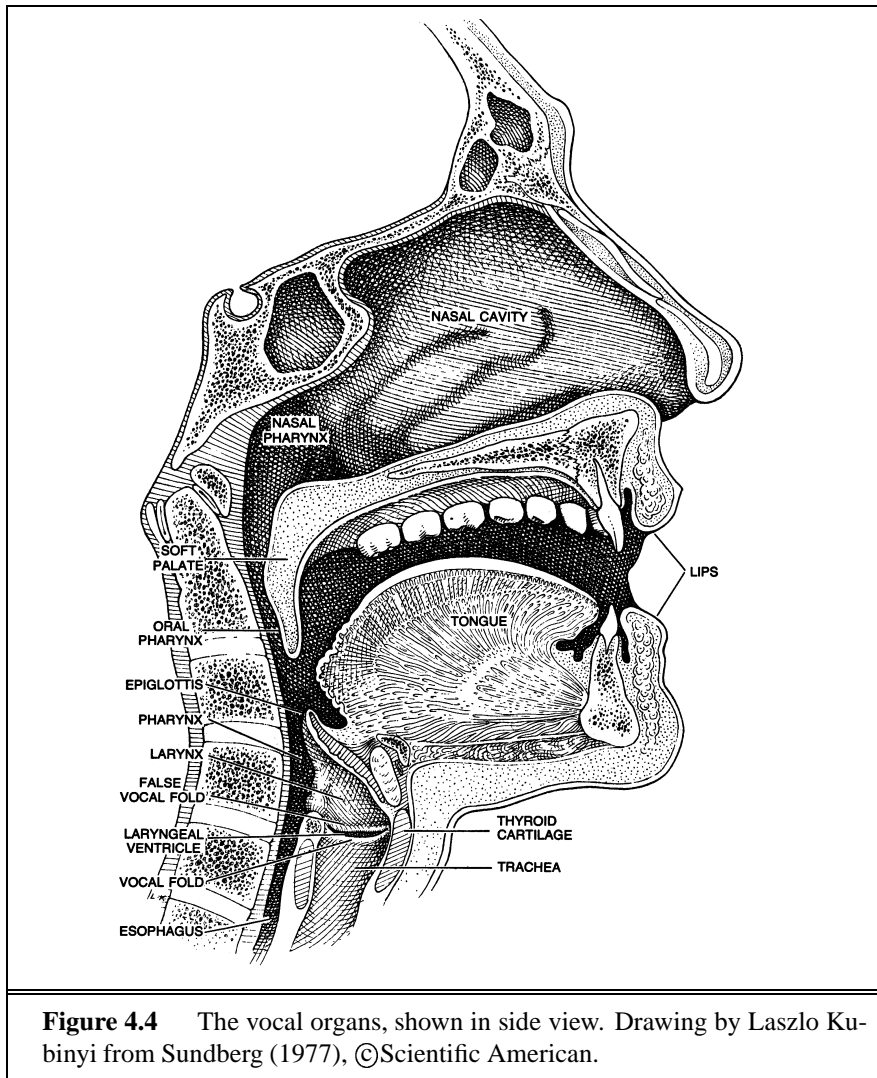


Figure 4.4 The vocal organs, shown in side view. Drawing by Laszlo Kubinyi from Sundberg (1977), ©Scientific American.

The area above the trachea is called the **vocal tract**, and consists of the **oral tract** and the **nasal tract**. After the air leaves the trachea, it can exit the body through the mouth or the nose. Most sounds are made by air passing through the mouth. Sounds made by air passing through the nose are called **nasal sounds**; nasal sounds use both the oral and nasal tracts as resonating cavities; English nasal sounds include *m*, and *n*, and *ng*.

Phones are divided into two main classes: **consonants** and **vowels**. Both kinds of sounds are formed by the motion of air through the mouth,

NASAL
SOUNDS

CONSONANTS

VOWELS

throat or nose. Consonants are made by restricting or blocking the airflow in some way, and may be voiced or unvoiced. Vowels have less obstruction, are usually voiced, and are generally louder and longer-lasting than consonants. The technical use of these terms is much like the common usage; [p], [b], [t], [d], [k], [g], [f], [v], [s], [z], [r], [l], etc., are consonants; [aa], [ae], [aw], [ao], [ih], [aw], [ow], [uw], etc., are vowels. **Semivowels** (such as [y] and [w]) have some of the properties of both; they are voiced like vowels, but they are short and less syllabic like consonants.

Consonants: Place of Articulation

PLACE

Because consonants are made by restricting the airflow in some way, consonants can be distinguished by where this restriction is made: the point of maximum restriction is called the **place of articulation** of a consonant. Places of articulation, shown in Figure 4.5, are often used in automatic speech recognition as a useful way of grouping phones together into equivalence classes:

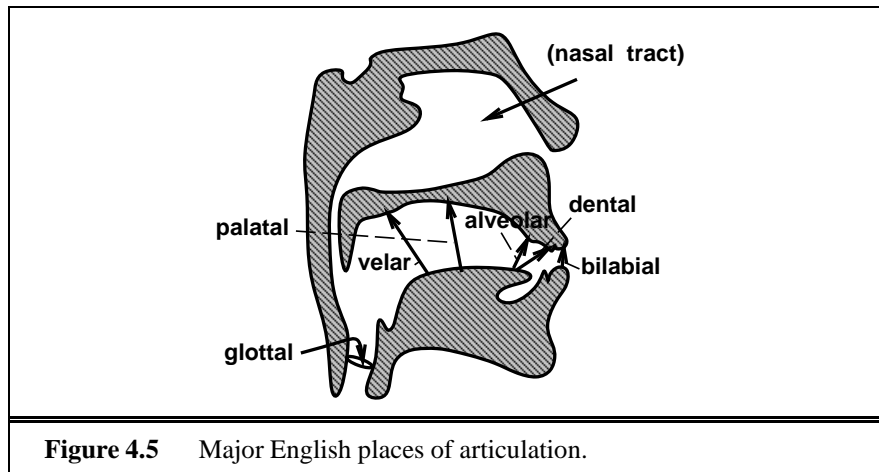


Figure 4.5 Major English places of articulation.

LABIAL

- **labial:** Consonants whose main restriction is formed by the two lips coming together have a **bilabial** place of articulation. In English these include [p] as in *possum*, [b] as in *bear*, and [m] as in *marmot*. The English **labiodental** consonants [v] and [f] are made by pressing the bottom lip against the upper row of teeth and letting the air flow through the space in the upper teeth.

DENTAL

- **dental:** Sounds that are made by placing the tongue against the teeth

are dentals. The main dentals in English are the [θ] of *thing* or the [ð] of *though*, which are made by placing the tongue behind the teeth with the tip slightly between the teeth.

- **alveolar:** The alveolar ridge is the portion of the roof of the mouth just behind the upper teeth. Most speakers of American English make the phones [s], [z], [t], and [d] by placing the tip of the tongue against the alveolar ridge. ALVEOLAR
- **palatal:** The roof of the mouth (the **palate**) rises sharply from the back of the alveolar ridge. The **palato-alveolar** sounds [ʃ] (*shrimp*), [tʃ] (*chinchilla*), [ʒ] (*Asian*), and [dʒ] (*jaguar*) are made with the blade of the tongue against this rising back of the alveolar ridge. The palatal sound [y] of *yak* is made by placing the front of the tongue up close to the palate. PALATAL
PALATE
- **velar:** The **velum** or soft palate is a movable muscular flap at the very back of the roof of the mouth. The sounds [k] (*cuckoo*), [g] (*goose*), and [ŋ] (*kingfisher*) are made by pressing the back of the tongue up against the velum. VELAR
VELUM
- **glottal:** The glottal stop [ʔ] is made by closing the glottis (by bringing the vocal folds together). GLOTTAL

Consonants: Manner of Articulation

Consonants are also distinguished by *how* the restriction in airflow is made, for example whether there is a complete stoppage of air, or only a partial blockage, etc. This feature is called the **manner of articulation** of a consonant. The combination of place and manner of articulation is usually sufficient to uniquely identify a consonant. Here are the major manners of articulation for English consonants:

- **stop:** A stop is a consonant in which airflow is completely blocked for a short time. This blockage is followed by an explosive sound as the air is released. The period of blockage is called the **closure** and the explosion is called the **release**. English has voiced stops like [b], [d], and [g] as well as unvoiced stops like [p], [t], and [k]. Stops are also called **plosives**. It is possible to use a more narrow (detailed) transcription style to distinctly represent the closure and release parts of a stop, both in ARPAbet and IPA-style transcriptions. For example the closure of a [p], [t], or [k] would be represented as [p̚], [t̚], or [k̚] (respectively) in the ARPAbet, and [p̚], [t̚], or [k̚] (respectively) STOP

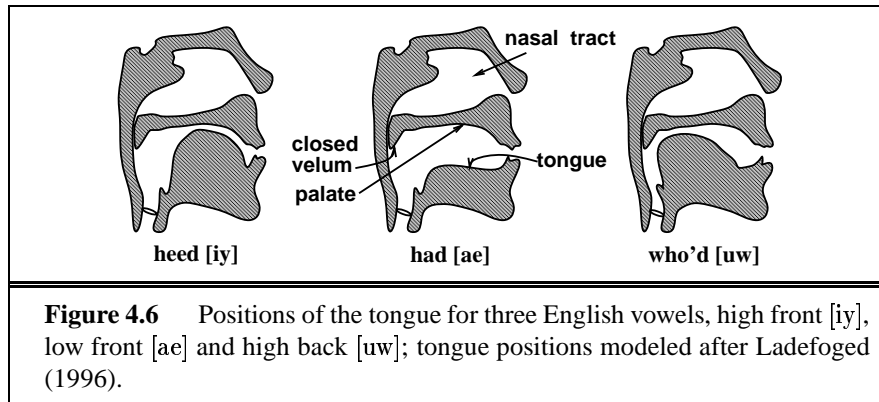
in IPA style. When this form of narrow transcription is used, the unmarked ARPABET symbols [p], [t], and [k] indicate purely the release of the consonant. We will not be using this narrow transcription style in this chapter.

- NASALS
 - **nasals:** The nasal sounds [n], [m], and [ŋ] are made by lowering the velum and allowing air to pass into the nasal cavity.
- FRICATIVE
 - **fricative:** In fricatives, airflow is constricted but not cut off completely. The turbulent airflow that results from the constriction produces a characteristic “hissing” sound. The English labiodental fricatives [f] and [v] are produced by pressing the lower lip against the upper teeth, allowing a restricted airflow between the upper teeth. The dental fricatives [θ] and [ð] allow air to flow around the tongue between the teeth. The alveolar fricatives [s] and [z] are produced with the tongue against the alveolar ridge, forcing air over the edge of the teeth. In the palatoalveolar fricatives [ʃ] and [ʒ] the tongue is at the back of the alveolar ridge forcing air through a groove formed in the tongue. The higher-pitched fricatives (in English [s], [z], [ʃ] and [ʒ]) are called **sibilants**. Stops that are followed immediately by fricatives are called **affricates**; these include English [tʃ] (*chicken*) and [dʒ] (*giraffe*).
- SIBILANTS
 - **approximant:** In approximants, the two articulators are close together but not close enough to cause turbulent airflow. In English [y] (*yellow*), the tongue moves close to the roof of the mouth but not close enough to cause the turbulence that would characterize a fricative. In English [w] (*wormwood*), the back of the tongue comes close to the velum. American [r] can be formed in at least two ways; with just the tip of the tongue extended and close to the palate or with the whole tongue bunched up near the palate. [l] is formed with the tip of the tongue up against the alveolar ridge or the teeth, with one or both sides of the tongue lowered to allow air to flow over it. [l] is called a **lateral** sound because of the drop in the sides of the tongue.
- APPROXIMANT
 - **tap:** A tap or **flap** [ɾ] is a quick motion of the tongue against the alveolar ridge. The consonant in the middle of the word *lotus* ([ləʊtəs]) is a tap in most dialects of American English; speakers of many British dialects would use a [t] instead of a tap in this word.
- TAP
- FLAP

Vowels

Like consonants, vowels can be characterized by the position of the articulators as they are made. The two most relevant parameters for vowels are

what is called vowel **height**, which correlates roughly with the location of the highest part of the tongue, and the shape of the lips (rounded or not). Figure 4.6 shows the position of the tongue for different vowels.



In the vowel [i], for example, the highest point of the tongue is toward the front of the mouth. In the vowel [u], by contrast, the high-point of the tongue is located toward the back of the mouth. Vowels in which the tongue is raised toward the front are called **front vowels**; those in which the tongue is raised toward the back are called **back vowels**. Note that while both [i] and [ε] are front vowels, the tongue is higher for [i] than for [ε]. Vowels in which the highest point of the tongue is comparatively high are called **high vowels**; vowels with mid or low values of maximum tongue height are called **mid vowels** or **low vowels**, respectively.

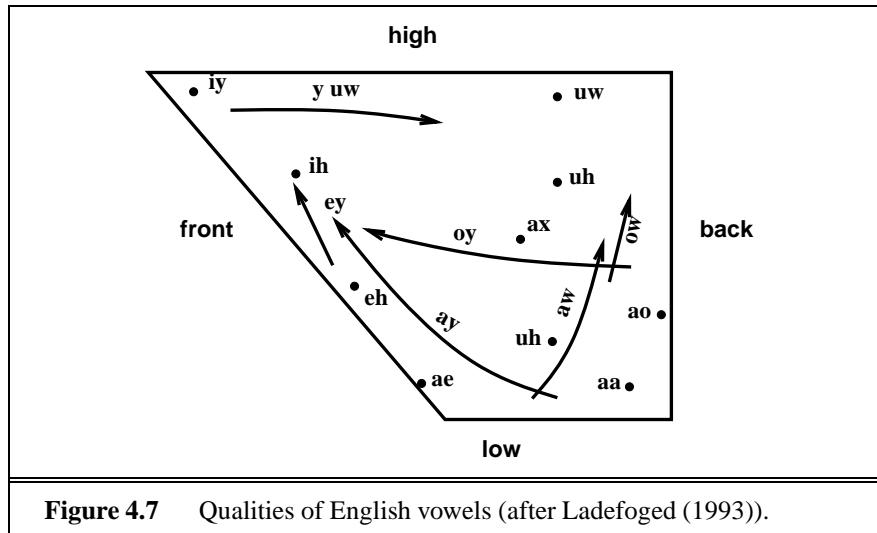
FRONT
BACK
HIGH

Figure 4.7 shows a schematic characterization of the vowel height of different vowels. It is schematic because the abstract property **height** only correlates roughly with actual tongue positions; it is in fact a more accurate reflection of acoustic facts. Note that the chart has two kinds of vowels: those in which tongue height is represented as a point and those in which it is represented as a vector. A vowels in which the tongue position changes markedly during the production of the vowel is **diphthong**. English is particularly rich in diphthongs; many are written with two symbols in the IPA (for example the [eɪ] of *hake* or the [oʊ] of *cobra*).

DIPHTHONG

The second important articulatory dimension for vowels is the shape of the lips. Certain vowels are pronounced with the lips rounded (the same lip shape used for whistling). These **rounded** vowels include [u], [ɔ], and the diphthong [oʊ].

ROUNDED



Syllables

SYLLABLE Consonants and vowels combine to make a **syllable**. There is no completely agreed-upon definition of a syllable; roughly speaking a syllable is a vowel-like sound together with some of the surrounding consonants that are most closely associated with it. The IPA period symbol [.] is used to separate syllables, so *parsley* and *catnip* have two syllables ([ˈpɑr.sli] and [ˈkæt.nɪp] respectively), *tarragon* has three [ˈtæ.rə.gən], and *dill* has one ([dɪl]). A syllable is usually described as having an optional initial consonant or set of consonants called the **onset**, followed by a vowel or vowels, followed by a final consonant or sequence of consonants called the **coda**. Thus d is the onset of [dɪl], while l is the coda. The task of breaking up a word into syllables is called **syllabification**. Although automatic syllabification algorithms exist, the problem is hard, partly because there is no agreed-upon definition of syllable boundaries. Furthermore, although it is usually clear how many syllables are in a word, Ladefoged (1993) points out there are some words (*meal*, *teal*, *seal*, *hire*, *fire*, *hour*) that can be viewed either as having one syllable or two.

ACCENTED In a natural sentence of American English, certain syllables are more **prominent** than others. These are called **accented** syllables. Accented syllables may be prominent because they are louder, they are longer, they are associated with a pitch movement, or any combination of the above. Since accent plays important roles in meaning, understanding exactly why a speaker

chooses to accent a particular syllable is very complex. But one important factor in accent is often represented in pronunciation dictionaries. This factor is called **lexical stress**. The syllable that has lexical stress is the one that will be louder or longer if the word is accented. For example the word *parsley* is stressed in its first syllable, not its second. Thus if the word *parsley* is accented in a sentence, it is the first syllable that will be stronger. We write the symbol [ˈ] before a syllable to indicate that it has lexical stress (e.g. [ˈpɑr.sli]). This difference in lexical stress can affect the meaning of a word. For example the word *content* can be a noun or an adjective. When pronounced in isolation the two senses are pronounced differently since they have different stressed syllables (the noun is pronounced [ˈkɑn.tɛnt]) and the adjective [kən.ˈtɛnt]. Other pairs like this include *object* (noun [ˈɒb.dʒɛkt] and verb [əb.ˈdʒɛkt]); see Cutler (1986) for more examples. Automatic disambiguation of such **homographs** is discussed in Chapter 17. The role of prosody is taken up again in Section 4.7.

LEXICAL
STRESS

HOMOGRAPHS

4.2 THE PHONEME AND PHONOLOGICAL RULES

'Scuse me, while I kiss the sky
 Jimi Hendrix, *Purple Haze*
 'Scuse me, while I kiss this guy
 Common mis-hearing of same lyrics

All [t]s are not created equally. That is, phones are often produced differently in different contexts. For example, consider the different pronunciations of [t] in the words *tunafish* and *starfish*. The [t] of *tunafish* is **aspirated**. Aspiration is a period of voicelessness after a stop closure and before the onset of voicing of the following vowel. Since the vocal cords are not vibrating, aspiration sounds like a puff of air after the [t] and before the vowel. By contrast, a [t] following an initial [s] is **unaspirated**; thus the [t] in *starfish* ([stɑrfɪʃ]) has no period of voicelessness after the [t] closure. This variation in the realization of [t] is predictable: whenever a [t] begins a word or unreduced syllable in English, it is aspirated. The same variation occurs for [k]; the [k] of *sky* is often mis-heard as [g] in Jimi Hendrix's lyrics because [k] and [g] are both unaspirated. In a very detailed transcription system we could use the symbol for aspiration [h] after any [t] (or [k] or [p]) which begins a word or unreduced syllable. The word *tunafish* would be transcribed [t^hunəfɪʃ] (the ARPAbet does not have a way of marking aspiration).

UNASPIRATED

There are other contextual variants of [t]. For example, when [t] occurs between two vowels, particularly when the first is stressed, it is pronounced as a tap. Recall that a tap is a voiced sound in which the top of the tongue is curled up and back and struck quickly against the alveolar ridge. Thus the word *buttercup* is usually pronounced [bʌɾəʔkʌp]/[b ah dx axr k ah p] rather than [bʌtəʔkʌp]/[b ah t axr k ah p].

Another variant of [t] occurs before the dental consonant [θ]. Here the [t] becomes dentalized ([t̪]). That is, instead of the tongue forming a closure against the alveolar ridge, the tongue touches the back of the teeth.

How do we represent this relation between a [t] and its different realizations in different contexts? We generally capture this kind of pronunciation variation by positing an abstract class called the **phoneme**, which is realized as different **allophones** in different contexts. We traditionally write phonemes inside slashes. So in the above examples, /t/ is a phoneme whose allophones include [t^h], [ɾ], and [t̪]. A phoneme is thus a kind of generalization or abstraction over different phonetic realizations. Often we equate the phonemic and the lexical levels, thinking of the lexicon as containing transcriptions expressed in terms of phonemes. When we are transcribing the pronunciations of words we can choose to represent them at this broad phonemic level; such a **broad transcription** leaves out a lot of predictable phonetic detail. We can also choose to use a **narrow transcription** that includes more detail, including allophonic variation, and uses the various diacritics. Figure 4.8 summarizes a number of allophones of /t/; Figure 4.9 shows a few of the most commonly used IPA diacritics.

PHONEME
ALLOPHONES

NARROW
TRANSCRIPTION

Phone	Environment	Example	IPA
[t ^h]	in initial position	<i>toucan</i>	[t ^h uk ^h æɪn]
[t]	after [s] or in reduced syllables	<i>starfish</i>	[stɑɾfɪʃ]
[ʔ]	word-finally or after vowel before [n]	<i>kitten</i>	[k ^h ɪʔn]
[ʔt]	sometimes word-finally	<i>cat</i>	[k ^h æʔt]
[ɾ]	between vowels	<i>buttercup</i>	[bʌɾəʔk ^h ʌp]
[t̪]	before consonants or word-finally	<i>fruitcake</i>	[frut̪k ^h eɪk]
[t̪]	before dental consonants ([θ])	<i>eighth</i>	[eɪt̪θ]
[̚]	sometimes word-finally	<i>past</i>	[pæ̚s]

Figure 4.8 Some allophones of /t/ in General American English.

The relationship between a phoneme and its allophones is often captured by writing a **phonological rule**. Here is the phonological rule for dentalization in the traditional notation of Chomsky and Halle (1968):

$$/t/ \rightarrow [t̪] / __ \theta \quad (4.1)$$

In this notation, the surface allophone appears to the right of the arrow, and the phonetic environment is indicated by the symbols surrounding the underbar (___). These rules resemble the rules of two-level morphology of Chapter 3 but since they don't use multiple types of rewrite arrows, this rule is ambiguous between an obligatory or optional rule. Here is a version of the flapping rule:

$$/ \left\{ \begin{array}{c} t \\ d \end{array} \right\} / \rightarrow [r] / \acute{V} __ V \quad (4.2)$$

Diacritics			Suprasegmentals		
◌ ^h	Voiceless	[a̟]	ˈ	Primary stress	[ˈpu.mə]
	Aspirated	[p ^h]	ˌ	Secondary stress	[ˌfɔʊrə.græf]
◌̩	Syllabic	[l̩]	ː	Long	[aː]
◌̃	Nasalized	[æ̃]	ˑ	Half long	[aˑ]
◌̚	Unreleased	[t̚]	ˑ	Syllable break	[pu.mə]
◌̪	Dental	[t̪]			

Figure 4.9 Some of the IPA diacritics and symbols for suprasegmentals.

4.3 PHONOLOGICAL RULES AND TRANSDUCERS

Chapter 3 showed that spelling rules can be implemented by transducers. Phonological rules can be implemented as transducers in the same way; indeed the original work by Johnson (1972) and Kaplan and Kay (1981) on finite-state models was based on phonological rules rather than spelling rules. There are a number of different models of **computational phonology** that use finite automata in various ways to realize phonological rules. We will describe the **two-level morphology** of Koskeniemi (1983) used in Chapter 3, but the interested reader should be aware of other recent models.³ While Chapter 3 gave examples of two-level rules, it did not talk about the

³ One example is Bird and Ellison's (1994) model of the multi-tier representations of autosegmental phonology in which each phonological tier is represented by a finite-state automaton, and autosegmental association by the synchronization of two automata.

motivation for these rules, and the differences between traditional ordered rules and two-level rules. We will begin with this comparison.

As a first example, Figure 4.10 shows a transducer which models the application of the simplified flapping rule in (4.3):

$$/t/ \rightarrow [r] / \acute{V} _ V \quad (4.3)$$

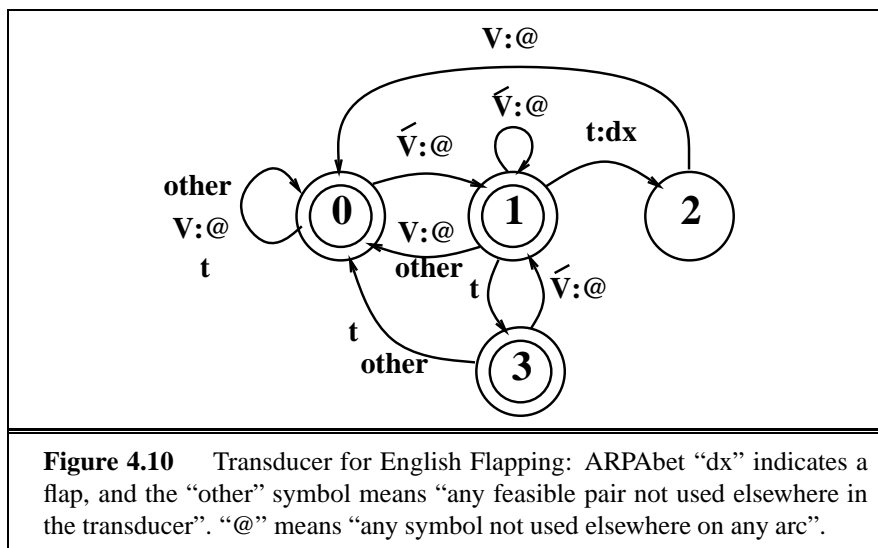


Figure 4.10 Transducer for English Flapping: ARPAbet “dx” indicates a flap, and the “other” symbol means “any feasible pair not used elsewhere in the transducer”. “@” means “any symbol not used elsewhere on any arc”.

The transducer in Figure 4.10 accepts any string in which flaps occur in the correct places (after a stressed vowel, before an unstressed vowel), and rejects strings in which flapping doesn’t occur, or in which flapping occurs in the wrong environment. Of course the factors that flapping are actually a good deal more complicated, as we will see in Section 5.7.

In a traditional phonological system, many different phonological rules apply between the lexical form and the surface form. Sometimes these rules interact; the output from one rule affects the input to another rule. One way to implement rule-interaction in a transducer system is to run transducers in a *cascade*. Consider, for example, the rules that are needed to deal with the phonological behavior of the English noun plural suffix *-s*. This suffix is pronounced [ɪz] after the phones [s], [ʃ], [z], or [ʒ] (so *peaches* is pronounced [pitʃɪz], and *faxes* is pronounced [fæksɪz]), [z] after voiced sounds (*pigs* is pronounced [pɪgz]), and [s] after unvoiced sounds (*cats* is pronounced [kæts]). We model this variation by writing phonological rules for the realization of the morpheme in different contexts. We first need to choose one of these three forms (s, z, and ɪz) as the “lexical” pronunciation of the suffix; we

chose z only because it turns out to simplify rule writing. Next we write two phonological rules. One, similar to the E-insertion spelling rule of page 77, inserts a $[i]$ after a morpheme-final sibilant and before the plural morpheme $[z]$. The other makes sure that the $-s$ suffix is properly realized as $[s]$ after unvoiced consonants.

$$\epsilon \rightarrow i / [+sibilant] \text{ } \hat{\text{ _ }} z \# \quad (4.4)$$

$$z \rightarrow s / [-voice] \text{ } \hat{\text{ _ }} \# \quad (4.5)$$

These two rules must be *ordered*; rule (4.4) must apply before (4.5). This is because the environment of (4.4) includes z , and the rule (4.5) changes z . Consider running both rules on the lexical form *fox* concatenated with the plural $-s$:

Lexical form: faks[^]z
(4.4) applies: faks[^]iz
(4.5) doesn't apply: faks[^]iz

If the devoicing rule (4.5) was ordered first, we would get the wrong result (what would this incorrect result be?). This situation, in which one rule destroys the environment for another, is called **bleeding**:⁴

Lexical form: faks[^]z
(4.5) applies: faks[^]s
(4.4) doesn't apply: faks[^]s

As was suggested in Chapter 3, each of these rules can be represented by a transducer. Since the rules are ordered, the transducers would also need to be ordered. For example if they are placed in a **cascade**, the output of the first transducer would feed the input of the second transducer.

Many rules can be cascaded together this way. As Chapter 3 discussed, running a cascade, particularly one with many levels, can be unwieldy, and so transducer cascades are usually replaced with a single more complex transducer by **composing** the individual transducers.

Koskenniemi's method of **two-level morphology** that was sketchily introduced in Chapter 3 is another way to solve the problem of rule ordering. Koskenniemi (1983) observed that most phonological rules in a grammar are independent of one another; that feeding and bleeding relations between

⁴ If we had chosen to represent the lexical pronunciation of $-s$ as $[s]$ rather than $[z]$, we would have written the rule inversely to voice the $-s$ after voiced sounds, but the rules would still need to be ordered; the ordering would simply flip.