Both phonetics and phonology can be generally described as the study of speech sounds, but they are not the same field. Phonetics (the subject of Chapter 2) is specifically the study of how speech sounds are produced, what their physical properties are, and how they are interpreted. Phonology, on the other hand, is the study of the distribution of sounds in a language and the interactions between those different sounds. Phonologists ask the following kinds of questions: What is the organization of sounds in a given language? Of all the sounds in a language, which are predictable and which are unpredictable in given contexts? Which sounds affect the identities of words?

Contents

3.1 The Value of Sounds: Phonemes and Allophones
   Introduces the two levels of phonological representation—phonemes and allophones—and describes the three basic ways sounds can be distributed in a language.

3.2 Phonological Rules
   Describes how phonological rules map between the two levels, introduces the idea of natural classes, and introduces several types of common phonological processes.

3.3 Phonotactic Constraints and Foreign Accents
   Introduces the idea that there are language-specific limitations on how sounds can be put together, and relates this to some of the reasons that non-native speakers of a language seem to have a foreign accent.

3.4 Implicational Laws
   Describes how certain phonological patterns recur in languages, in a particular ordered hierarchy, and introduces some explanatory principles for these patterns.

3.5 How to Solve Phonology Problems
   Outlines some basic techniques and strategies for solving phonological analysis problems.

3.6 Practice
   Provides exercises, discussion questions, and activities related to phonology.
The Value of Sounds: Phonemes and Allophones

3.1.1 Predicting the Occurrence of Sounds

In both Kikamba (a Bantu language spoken in Kenya) and English, we can hear the sounds [k] and [g]. The Kikamba word [kosuugar] 'to guard' contains both phones, as does the English word [kəʊɡnet] cognate. The difference between Kikamba and English lies in the way the two sounds contribute to the identity of a word. In English, the two phones can distinguish words, as shown by words like [tæk] tack and [tæɡ] tag, where alternating between [k] and [ɡ] affects the message conveyed by the utterance. In this sense, phonologists say that the occurrence of these two sounds in English is unpredictable, since we cannot look at the rest of the word and determine which sound will occur. That is, if we know that a word in English begins with [tæ], we cannot predict whether the word will end with [k] or [ɡ] since both tack and tag are different, but possible, words.

In Kikamba, on the other hand, the sounds [k] and [ɡ] are predictable from their environment. Sounds are predictable when we expect to see one sound but not the other based upon the sounds that precede and/or follow it. In Kikamba, the only consonant that can come directly after an [ŋ] is [ɡ], and [ɡ] can only come immediately after [ŋ]. The combination [ŋk] does not occur in Kikamba (see Roberts-Kohn 2000). So, if there is a velar stop in a word in Kikamba, we can predict whether it will be a [k] or a [ɡ]: it will be a [ɡ] if it is immediately preceded by [ŋ]; otherwise, it will be a [k]. However, in English we cannot make this prediction: the sound [k] does appear after the sound [ŋ], as in [æŋkɔ] anchor, as does the sound [ɡ], as in [æŋɡɡ] anger.

To illustrate how strong this distribution is in Kikamba, consider the case where you have a word with a [k] such as katala, the base form from which conditional forms of the verb 'to count' are built. To say 'if you count,' you add an [ŋ] to the front of the word: [okatala]. But to say 'if I count,' you add an [ŋ]. Even though this word has a [k] in it, we have seen that it is a rule in this language that [k] cannot appear after [ŋ]—so the [k] appears as a [ɡ] instead: [ŋgatala] 'if I count.' This type of alternation does not happen in English, because there is no rule governing the distribution of [k] and [ɡ].

So while Kikamba and English both use the phones [k] and [ɡ], the languages differ in that in Kikamba we can predict the occurrence of one versus the other, and in English we cannot. If someone learning Kikamba were to use [k] after [ŋ], the identity of the word would not change. Instead, a native speaker of Kikamba might think that the speaker sounded funny, had an accent, or had mispronounced the word. On the other hand, if a learner of English were to make the same substitution in English, then the identity of the word is likely to change. Imagine confusing [k] and [ɡ] and saying at a dinner party "we're having [ɡæb] for dinner tonight." Your guests might feel rather uncomfortable, especially if they were expecting crab!

The bottom line is that in Kikamba, the sounds [k] and [ɡ] are predictably distributed, while in English they are not.
3.1.2 Allophones and Phonemes

In every language, certain sounds pattern together as if they were simply variants of the “same” sound, instead of different sounds that can be used to distinguish words, even though they may be phonetically distinct. For example, the sounds [k] and [g] are clearly different sounds: we use them to make the contrast between different words in English, as we saw in Section 3.1.1. But as we also saw, these two sounds are completely predictable in Kikamba. In Kikamba, then, these sounds can be thought of as variants of the “same” sound, because in any given context, if there is some velar stop sound, we can predict which one ([k] or [g]) it will be.

Similarly, if you ask a native speaker of English how many different sounds are represented by the underlined letters in the words *pin*, *pin*, and *spin*, they will probably say “two,” grouping the aspirated [pʰ] of *pin* and unaspirated [p] of *spin* together. Though [pʰ] and [p] are phonetically different sounds, native English speakers often overlook this difference and may even consider them to be the “same” sound (see Section 2.6.5).

One of the goals of this file is to help you understand more clearly the distinction between “same” and “different” sounds. To do this, we will discuss the terms allophone and phoneme. Since these concepts are the crux of phonological analysis, it is important that they be clearly understood. Perhaps the best way to start to explain these terms is through examples. On a separate piece of paper, transcribe the following five words in IPA:

(1) top stop little kitten hunter

It is likely that you transcribed all of these words with a [t], like the following:

(2) [tʰɔp] [ʃɔp] [lɪtʃ] [kɪtʃ] [ʰ æntʃ]

This is good, since it probably reflects something that is psychologically real to you. But, in fact, the physical reality (the acoustic phonetic fact) is that the ‘t’ you transcribed in those five examples is pronounced slightly differently from one example to the next. To illustrate this, pronounce the five words again. Concentrate on what the ‘t’ sounds like in each example, but be sure to say them as you normally would if you were talking to a friend—that is, don’t try to enunciate them abnormally clearly.

What differences did you notice? Compare, for example, the /t/ of _top_ to that of _stop_. You should be able to detect a short burst or puff of air after the /t/ in _top_ that is absent in _stop_. That puff of air is what we have called aspiration (see Section 2.6.5), which is transcribed with a superscripted [ʰ]. So while a native speaker might think of the ‘t’ sound in _top_ and _stop_ as being the same sound, the ‘t’ is actually pronounced differently in each word. This difference can be captured in the transcription, as in [tʰɔp] and [ʃɔp], respectively.

Now say the words _little_ and _kitten_. We might say that the ‘t’ in _little_ sounds “softer” than the one in _stop_, and is clearly voiced. For most speakers of American English (but not of British English), the ‘t’ in words like _little_ is pronounced as a flap, [ɾ], much like the r in Spanish in words like _para_ ‘for’ and _toro_ ‘bull’ (see Section 2.2.5). English _kitten_, on the other hand, is pronounced with the same sound we hear in the expression _uh-oh_, a glottal stop [ʔ]. So, we could transcribe _little_ and _kitten_ as [lɾʃ] and [kʰɪʔʃ], respectively.

For some speakers of American English, in casual speech words like _hunter_ are pronounced with no ‘t’ at all, but rather as [hʌntʃ]. Try to say it this way and see if it sounds like something you’ve heard before. In any case, while you may have initially transcribed the five words above with a /t/, they may also be transcribed in a way that reflects the different pronunciations of that sound, as in the following:

(3) [tʰɔp] [ʃɔp] [lɪɾʃ] [kʰɪʔʃ] [hʌntʃ]
To a native speaker, all of the words above may seem to have a ‘t’ in them, at least at some psychological level.\(^1\) Evidence of this lies in the fact that one may transcribe them all with a ‘t,’ at least until trained in transcription. Someone who lacks linguistic training would probably not hesitate to state that all the above words have a ‘t’ and would need to be convinced that subtle differences, like aspiration, exist among them. In this sense, the above words do have a ‘t.’ On the other hand, we can observe that the ‘t’ may be pronounced in several different ways.

Unlike a speaker of English, a native speaker of Hindi does not ignore the difference between aspirated and unaspirated sounds when speaking or hearing Hindi. To a speaker of Hindi, the aspirated sound [pʰ] is as different from unaspirated [p] as [pʰ] is from [b] to our ears. The difference between aspirated and unaspirated stops must be noticed by Hindi speakers because their language contains many words that are pronounced in nearly the same way, except that one word will have an aspirated stop where the other has an unaspirated stop. The data in (4) illustrate this.

<table>
<thead>
<tr>
<th>Hindi</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>[pʰəl]</td>
<td>‘fruit’</td>
</tr>
<tr>
<td>[pəl]</td>
<td>‘moment’</td>
</tr>
<tr>
<td>[bəl]</td>
<td>‘strength’</td>
</tr>
</tbody>
</table>

A native speaker of English may not be aware of the difference between aspirated and unaspirated stops because aspiration will never make a difference in the meanings of English words. If we hear someone say [məp] and [məpʰ], we may recognize them as different pronunciations of the same word *map*, but not as different words. Because of the different ways in which [p] and [pʰ] affect meaning distinctions in English and Hindi, these sounds have different values in the phonological systems of the two languages. We say that these two sounds are noncontrastive in English, since interchanging the two does not result in a change of meaning. In Hindi, on the other hand, [p] and [pʰ] are contrastive, since replacing one sound with the other in a word can change the word’s meaning. We will have more to say about this terminological distinction below.

Linguists attempt to characterize these different relations between sounds in language by grouping the sounds in a language’s sound inventory into classes. Each class contains all of the sounds that a native speaker considers as the “same” sound. For example, [t] and [tʰ] in English would be members of the same class. But English [t] and [d] are members of different classes because they are contrastive. That is, if you interchange one for the other in a word, you can cause a change in the word’s meaning, e.g. [tʰaum] *time* versus [daum] *dime*. On the other hand, speakers of Hindi would not classify [t] and [tʰ] as members of the same class because they perceive them as different. That is, they are contrastive in Hindi.

A class of speech sounds that seem to be variants of the same sound is called a phoneme. Each member of a particular phoneme class is called an allophone, which corresponds to an actual phonetic segment produced by a speaker. That is, the various ways that a phoneme is pronounced are called allophones.

In this view, we can say that the ‘t’ sounds in words like *stop, top, little,* and *kitten* all belong to a single class, which we will label by the symbol /t/, characterizing this particular phoneme. By saying that *stop* and *top*, for example, each have the phoneme /t/, we are saying that the sounds [t] and [tʰ] are related.

\(^1\)The reasons for this may be manifold, including phonetic similarities, phonological patterning, different pronunciations across language varieties, or spelling.
In (5) we see how the phoneme /t/ is related to its allophones in English and how the Hindi phonemes /t/ and /tʰ/ are related to their allophones. In English, [t], [tʰ], [ɾ], and [ʔ] are allophones of the same phoneme, which we can label /t/. In this way, we can say that in English the phoneme /t/ has the allophones [t] as in [stap], [tʰ] as in [tʰap], [ɾ] as in [lɾ], and [ʔ] as in [kʔq]. On the other hand, in Hindi, [t] and [tʰ] are allophones of different phonemes. Note that symbols representing phonemes are written between slashes; this distinguishes them from symbols representing (allo)phones, which are written between square brackets.

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th>Hindi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonemes:</td>
<td>/t/</td>
<td>/t/</td>
</tr>
<tr>
<td>Allophones</td>
<td>[t]</td>
<td>[ɾ]</td>
</tr>
<tr>
<td></td>
<td>[tʰ]</td>
<td>[ʔ]</td>
</tr>
</tbody>
</table>

By providing a description like this, linguists attempt to show that the phonological system of a language has two levels. The more concrete level involves the physical reality of phonetic segments, the allophones, whereas phonemes are something more abstract. In fact, linguists sometimes describe phonemes as the form in which we store sounds in our minds. So, phonemes are abstract psychological concepts, and they are not directly observable in a stream of speech; only the allophones of a phoneme are.

The phoneme is a unit of linguistic structure that is just as significant to the native speaker as the word or the sentence. Native speakers reveal their knowledge of phonemes in a number of ways. When an English speaker makes a slip of the tongue and says [tʃ] kem] for key chain, reversing [tʃ] and [k], he or she has demonstrated that [tʃ] functions mentally as a single unit of sound, just as [k] does. Recall from File 2.2 that [tʃ] is phonetically complex, consisting of [t] followed immediately by [ʃ]. Yet, since [ʃ] represents the pronunciation of a single phoneme /ʃ/ in English, no native speaker would make an error that would involve splitting up its phonetic components; you will never hear [tʃ kʃen] as a slip of the tongue (see File 9.3).

Knowledge of phonemes is also revealed in alphabetic spelling systems (see File 13.4). For example, English does not have separate letters for [pʰ] and [p]; they are both spelled with the letter p. Examples like this show that the English spelling system ignores differences in pronunciation that don’t result in meaning distinctions. For the most part, the English spelling system attempts to provide symbols for phonemes, not phonetic segments. In general, alphabetic writing systems tend to be phonemic rather than phonetic, though they achieve this goal with varying degrees of success. As noted in File 2.1, of course, there are multiple ways to represent the same sound (e.g., the [k] sound is written with a <k> in the word kitten but with a <c> in the word cool). What’s crucial here, though, is that both of these spellings represent /k/, and not, say, the difference between [k] and [kʰ].

### 3.1.3 Identifying Phonemes and Allophones: The Distribution of Speech Sounds

In order to determine whether sounds in a given language are allophones of a single phoneme or allophones of separate phonemes, we need to consider the distribution of the sounds involved. The distribution of a phone is the set of phonetic environments in which it occurs. For example, nasalized vowels in English occur only in the environment of a nasal consonant. More precisely, a linguist would describe the distribution of English [ɪ], [ʊ], and so on, by stating that the nasalized vowels always and only occur immediately preceding a
nasal consonant. In this book we will mainly be concerned with two types of distribution—contrastive distribution and complementary distribution—though a third distribution, free variation, will also be introduced in the following section.

Let us consider contrastive distribution first. Recall from above that a pair of phones is contrastive if interchanging the two can change the meaning of a word. This means that the sounds can occur in the same phonetic environment. It also means that the sounds are allophones of different phonemes. Two sounds are noncontrastive if replacing one phone with another does not result in a change of meaning.

Our earlier discussion of the patterning of [p] and [ph] in Hindi and English provides a good example of this difference. Recall that we said that in Hindi these two sounds could affect the meaning of a word based on examples like [pal] moment and [phal] fruit, where the two meanings are distinguished by the occurrence of [p] or [ph]. This means that the two sounds [p] and [ph] are contrastive in Hindi. In English, on the other hand, simply replacing [p] for [ph], or vice versa, will never effect a change in the meaning of a word; the sounds are noncontrastive in English.

We just determined whether or not [p] or [ph] are contrastive in Hindi and English by taking into account the distribution of sounds in each individual language. We did this by identifying a minimal pair. A minimal pair is defined as a pair of words whose pronunciations differ by exactly one sound and that have different meanings. When you find a minimal pair, you know that the two sounds that differ are contrastive and, thus, the sounds involved are allophones of different phonemes. If you try, you can think of many minimal pairs in English, or any other language you know well. For example, the minimal pair [təim] team and [təim] teen shows that [n] and [m] are allophones of separate phonemes (that is, they are contrastive) in English since they can be used to contrast meaning. In Hindi, the words [phal] ‘fruit’ and [bal] ‘strength’ constitute a minimal pair, showing [ph] and [b] to be allophones of separate phonemes; [phal] fruit and [pal] moment also form a minimal pair in Hindi. But notice that there are no minimal pairs involving [p] and [ph] in English; these two sounds are never contrastive with respect to one another. Instead, they are allophones of the same phoneme, /p/.

Consider another example in which two languages make different distinctions using the same set of sounds. In English, [l] and [r] are contrastive phonemes, as can be seen from the existence of minimal pairs such as leaf [lif] versus reef [rif], alive [aliv] versus arrive [ərəiv], or feel [fıl] versus fear [fər]. In Korean, on the other hand, [l] and [r] are never contrastive.\(^2\) Consider the data in (6).

<table>
<thead>
<tr>
<th>Citation Form</th>
<th>Nominative Case</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>[pul]</td>
<td>[puri]</td>
<td>‘fire’</td>
</tr>
<tr>
<td>[mal]</td>
<td>[mar]</td>
<td>‘language, speech’</td>
</tr>
<tr>
<td>[tal]</td>
<td>[tari]</td>
<td>‘moon’</td>
</tr>
<tr>
<td>[kʰal]</td>
<td>[kʰari]</td>
<td>‘knife’</td>
</tr>
<tr>
<td>[pal]</td>
<td>[pari]</td>
<td>‘foot’</td>
</tr>
<tr>
<td>[sal]</td>
<td>[sari]</td>
<td>‘flesh’</td>
</tr>
</tbody>
</table>

\(^2\)You will notice that the two “r” sounds in English and Korean are not the same phonetically—in English, it is a voiced alveolar retroflex liquid [ɾ], while in Korean, it is a voiced alveolar flap [ɾ] (see Section 2.2.5). Similarly, the exact articulation of /l/ in the two languages is also not identical, though we use the same symbol for both. These phonetic differences are not particularly important here, however, because we are concerned only with the distribution of the two sounds in each language (a phonological question) rather than with the quality of the two sounds (a phonetic question).
Notice that in each example, only one English translation is given. For example, the word for 'fire' has two forms: one when it is used by itself, the citation form (such as in the answer to the question "What is the Korean word for 'fire'?"); and one when it is used as the subject of a sentence, the nominative form (such as in "The fire burned brightly"). This use of different forms for different grammatical roles is similar to the use in English of, say, \textit{atom} as the noun for the smallest unit of an element ("An atom of helium has 2 protons") as opposed to \textit{atomic} as the adjectival form of the same word ("The atomic structure of helium is simple"). In the English case, it's clear that the ending -\textit{i}c at the end of \textit{atomic} is a marker of an adjective. Similarly, in Korean, the final -\textit{i} in the second column of words in \cite{6} indicates that these words are in the nominative case. (For more on these kinds of markers, see Chapter \ref{4}.) If we remove the special ending of the word \textit{atomic}, you can see that the base of the word is still \textit{atom}. But notice that the base \textit{atom} in the word \textit{atomic} is not pronounced the same as it is in isolation. The word \textit{atom} by itself is pronounced [\textipa{æ,təm}], but in \textit{atomic}, it's pronounced [\textipa{æ,təm}]. Given that these two forms represent the same base word, we might expect that they would be pronounced the same way. Instead, the pronunciation of the base word alternates depending on the phonetic context it appears in. An \textit{alternation} is simply a difference between two (or more) phonetic forms that you might otherwise expect to be related. Identifying alternations relies on the assumption that, all else being equal, the same word should be expressed by the same sounds—when we find different pronunciations of the same word that are systematically linked to particular grammatical contexts, we have an alternation. So in English, the base word \textit{atom} is expressed alternately by the sounds [\textipa{æ,təm}] and [\textipa{æ,təm}]. In Korean, the base word \textit{fire} is expressed alternately by the sounds [\textipa{pul}] and [\textipa{pul}].

Why might these words have alternating pronunciations depending on what other elements they appear with? The answer to this question takes us back to looking at the distribution of the sounds [\textipa{i}] and [\textipa{ɾ}] in Korean. In the words listed in \cite{6}, we see that [\textipa{i}] and [\textipa{ɾ}] do not occur in the same phonetic environment. Even though we are looking at the "same" word (for example, the word for 'fire'), the phonetic quality of the third sound is not always the same—it alternates between [\textipa{i}] and [\textipa{ɾ}]. Specifically, as you can see from the data in \cite{6}, you use an [\textipa{i}] when the sound is the last sound in the word (is in "word-final" position), but you use [\textipa{ɾ}] when the sound is between two vowels (is in "intervocalic" position). In fact, if you were to look at all of Korean, you would find that [\textipa{ɾ}] appears only between two vowels, while [\textipa{i}] never appears in that position. Meanwhile, [\textipa{i}] can appear at the ends of words, but [\textipa{ɾ}] never does. These two observations mean that if someone gave you the frame [\textipa{to_i_ye}] in Korean, you could tell them whether an [\textipa{i}] or an [\textipa{ɾ}] goes in the blank. In Korean, it must be an [\textipa{i}]! Notice that you cannot do the same thing in English: either an [\textipa{i}] or an [\textipa{ɾ}] could go in the blank to form a possible English word.

The difference that has been illustrated here between English and Korean is that in English, [\textipa{i}] and [\textipa{ɾ}] are in contrastive distribution, while in Korean, [\textipa{i}] and [\textipa{ɾ}] are in \textit{complementary distribution}. Sounds showing this type of distribution are considered to be allophones of the same phoneme. To understand better what we mean by complementary distribution, think about what the term complementary means: two complementary parts of something make up a whole. For example, the set of people in your class at any given moment can be divided into the set of people who are under 5'5" tall and the set of people who are 5'5" tall or taller. These two sets of people complement each other. They are mutually exclusive (one person can’t simultaneously be both shorter and taller than 5'5"), but together they make up the whole class.

The Korean sounds [\textipa{i}] and [\textipa{ɾ}] are in complementary distribution, because they appear in different sets of environments: [\textipa{ɾ}] occurs between vowels, and [\textipa{i}] occurs word-finally.

\footnotesize{In some modern words that have come into Korean from other languages, [\textipa{i}] can also appear at the beginning of words. But it never appears as the only segment between two vowels.}
Given our assumptions about alternations—that the same word should be expressed by the same sounds—we can hypothesize that even though [I] and [R] are phonetically different phones in Korean, they are allophones of a single phoneme. We can represent that single phoneme as /I/. At this point, it probably seems like an arbitrary choice as to why /I/ should be the phoneme and not /R/; we will talk more about how to make this choice in File 3.5.

If two sounds are in complementary distribution in a language, there will never be a minimal pair that uses them to distinguish two words. Furthermore, the appearance of one allophone or the other will always be predictable, as we saw above with the frame [to__]. You can predict that [R] but not [I] will appear between vowels in any word in Korean, and that [I] but not [R] will appear word-finally—even if you have never studied Korean. This kind of prediction is a powerful tool in helping phonologists understand the structure of languages.

What's particularly interesting about this (and all other) phonological distributions is that it represents actual knowledge that native speakers have. For example, if you give a native speaker of Korean the new (nonsense) word moladam and ask them to say it out loud, they will say it with an [R] between the two vowels, and not an [I]! Of course, this is not something that anyone has explicitly taught them (especially since they have never seen this word before), but the distribution of sounds is one of the things that you know when you know a language (see File 1.2).

Consider another linguistic example, namely, the distribution of the English sounds [p] and [pʰ] shown in (7).

(7) spat [spæt]  pat [pʰæt]
     spool [spul]  pool [pʰul]
     speak [spik] peek [pʰik]

As you can see in the English words in (7), [p] and [pʰ] do not occur in the same phonetic environment. As a result, there are no minimal pairs involving a [p]-[pʰ] contrast. In fact, the phones are in complementary distribution: [p] occurs after [s] but never word-initially, while [pʰ] occurs word-initially but never after [s]. Since these sounds appear in different phonetic environments, there can be no pair of words composed of identical strings of sounds except that one has [p] and the other has [pʰ]. As stated above, phones that are in complementary distribution are allophones of a single phoneme. In this case, [p] and [pʰ] are both allophones of the phoneme we can represent as /p/. Furthermore, the appearance of one allophone or another in a given context is predictable. For example, we can predict that the allophone [pʰ] (but never [p]) will appear in word-initial position. So even in words not listed in (7), we know that it will be [pʰ], rather than [p], that will occur at the beginning of a word. Similarly, we can predict that [p] (but never [pʰ]) will follow [s] in other words.

### 3.1.4 Free Variation

Most phonological distributions can be described as either contrastive or complementary. Remember that the hallmark of a contrastive distribution is that you can't predict which of two (or more) sounds belongs in a certain context, because each will produce a different but meaningful word; the hallmark of a complementary distribution is that you can predict which of two sounds belongs in any given context. In some contexts, however, more than one pronunciation of a given sound may be possible. In these cases, you may not be able to predict exactly which sound will occur, but the choice does not affect the meaning of the

---

4 In point of fact, this is true not just at the beginning of a word but at the beginning of any stressed syllable. That is, in English, [pʰ] but not [p] can appear as the first consonant of a stressed syllable.
word. Consider, for example, the pronunciations of some English words in (8) (remember that [p'] represents an unreleased voiceless bilabial stop).

(8) leap [lip] leap [lip']
    soap [soup] soap [soup']
    troop [trup] troop [trup']
    happy [hæpi] — [hæp'i]

These words show that [p] and [p'] both share some of the same phonetic environments; specifically, they can both appear at the ends of words. Unlike the case of English [b] versus [p], or [m] versus [n], however, there are no minimal pairs involving these sounds in the language. Why not? Although there are pairs of words in (8) that differ in only one sound, none of these words contrast in meaning. Thus, the choice between [p] and [p'] in leap, soap, and troop does not make a difference in meaning; that is, the sounds are noncontrastive. Rather, they are interchangeable in word-final position. Sounds with this type of patterning are considered to be in free variation. To a native speaker, sounds like [p] and [p'] that are in free variation are perceived as being the "same" sound. We can conclude that they are allophones of the same phoneme, because they are perceived as the same and do not serve to distinguish the meanings of words.

Another term that you may encounter to describe the different types of phonological distributions in language is overlapping distribution. If two sounds are in overlapping distribution, they can occur in the same environment. Both sounds that are in contrastive distribution and sounds that are in free variation are therefore considered to have an overlapping distribution; only sounds that are in complementary distribution do not overlap. For example, in English, the sounds [d] and [l] are in overlapping distribution because they can contrast: the words lid and lit form a minimal pair, and both [d] and [l] can occur after [t]—that is, the environment [t] is one where [d] and [l] overlap. Similarly, [t] and [t'] have an overlapping distribution because they can also both occur after [l], as two different pronunciations of the word lit. The difference between [d] and [t] on the one hand, and [t] and [t'] on the other, is that interchanging [d] and [t] changes the identity of the words, while interchanging [t] and [t'] does not.

3.1.5 Summary

To summarize, a phone's distribution is the collection of phonetic environments in which the phone may appear; when linguists describe a phone's distribution, they describe this collection. Relative to each other, two (or more) phones will be in contrastive distribution, in complementary distribution, or in free variation. Phones in contrastive distribution may appear in minimal pairs and are allophones of different phonemes. Phones in free variation may, like phones in contrastive distribution, also appear in the same phonetic environments, but never cause a contrast in meaning; they are allophones of the same phoneme. In either of these two types of distribution, given a particular phonetic environment, one cannot predict which of the phones will occur. If the phones are in complementary distribution, their appearance in particular phonetic environments is predictable; they never appear in minimal pairs, and they are allophones of the same phoneme.