CHAPTER

1 What is phonology?

PREVIEW

KEY TERMS

sound
symbol
transcription
grammar
continuous nature of speech
accuracy

This chapter introduces phonology, the study of the sound systems of language. Its key objective is to:

- introduce the notion of phonological rule
- explain the nature of sound as a physical phenomenon
- highlight the tradeoff between accuracy and usefulness in representing sound
- distinguish between phonetics and phonology
- contrast the continuous and discrete aspects of linguistic sounds
- introduce the notion of "sound as cognitive symbol"

Phonology is one of the core fields that composes the discipline of linguistics, which is defined as the scientific study of language structure. One way to understand what the subject matter of phonology is, is to contrast it with other fields within linguistics. A very brief explanation is that phonology is the study of sound structure in language, which is different from the study of sentence structure (syntax) or word structure (morphology), or how languages change over time (historical linguistics). This definition is very simple, and also inadequate. An important feature of the structure of a sentence is how it is pronounced – its sound structure. The pronunciation of a given word is also a fundamental part of the structure of the word. And certainly the principles of pronunciation in a language are subject to change over time. So the study of phonology eventually touches on other domains of linguistics.

An important question is how phonology differs from the closely related discipline of phonetics. Making a principled separation between phonetics and phonology is difficult – just as it is difficult to make a principled separation between physics and chemistry, or sociology and anthropology. A common characterization of the difference between phonetics and phonology is that phonetics deals with "actual" physical sounds as they are manifested in human speech, and concentrates on acoustic waveforms, formant values, measurements of duration measured in milliseconds, of amplitude and frequency, or in the physical principles underlying the production of sounds, which involves the study of resonances and the study of the muscles and other articulatory structures used to produce physical sounds. On the other hand, phonology, it is said, is an abstract cognitive system dealing with rules in a mental grammar: principles of subconscious "thought" as they relate to language sound. Yet once we look into the central questions of phonology in greater depth, we will find that the boundaries between the disciplines of phonetics and phonology are not entirely clear-cut. As research in both of these fields has progressed, it has become apparent that a better understanding of many issues in phonology requires that you bring phonetics into consideration, just as a phonological analysis is a prerequisite for any phonetic study of language.

1.1 Concerns of phonology

As a step towards understanding what phonology is, and especially how it differs from phonetics, we will consider some specific aspects of sound structure that would be part of a phonological analysis. The point which is most important to appreciate at this moment is that the "sounds" which phonology is concerned with are symbolic sounds – they are cognitive abstractions, which represent but are not the same as physical sounds.

The sounds of a language. One aspect of phonology considers what the "sounds" of a language are. We would want to take note in a description of the phonology of English that we lack a particular vowel that exists in German in words like *schön* 'beautiful,' a vowel which is also found in French (spelled *eu,* as in *jeune* 'young'), or Norwegian (*øl* 'beer'). Similarly, the consonant spelled *th* in English *thing*, *path* does exist in English (as well as in Icelandic where it is spelled with the letter *þ*, or Modern Greek where it is spelled with *θ*, or Saami where it is spelled *ŧ*), but this sound does not occur in German or French, and it is not used in Latin American Spanish, although it does occur in Continental Spanish in words such as *cerveza* 'beer,' where by the spelling conventions of Spanish, the letters *c* and *z* represent the same sound as the one spelled *θ* (in Greek) or *th* (in English).

Rules for combining sounds. Another aspect of language sound which a phonological analysis would take account of is that in any given language, certain combinations of sounds are allowed, but other combinations are systematically impossible. The fact that English has the words *brick*, *break*, *bridge*, *bread* is a clear indication that there is no restriction against having words begin with the consonant sequence *br*; besides these words, one can think of many more words beginning with *br* such as *bribe*, *brow* and so on. Similarly, there are many words which begin with *bl*, such as *blue*, *blatant*, *blast*, *blend*, *blink*, showing that there is no rule against words beginning with *bl*. It is also a fact that there is no word **blick*[1] in English, even though the similar words *blink*, *brick* do exist. The question is, why is there no word **blick* in English? The best explanation for the nonexistence of this word is simply that it is an accidental gap – not every logically possible combination of sounds which follows the rules of English phonology is found as an actual word of the language.

Native speakers of English have the intuition that while *blick* is not actually a word of English, it is a theoretically possible word of English, and such a word might easily enter the language, for example via the introduction of a new brand of detergent. Fifty years ago the English language did not have any word pronounced *bick*, but based on the existence of words like *big* and *pick*, that word would certainly have been included in the set of nonexistent but theoretically allowed words of English. Contemporary English, of course, actually does contain that word – spelled *Bic* – which is a type of pen.

While the nonexistence of *blick* in English is accidental, the exclusion from English of many other imaginable but nonexistent words is based on a principled restriction of the language. While there are words that begin with *sn* like *snake*, *snip* and *snort*, there are no words beginning with *bn*, and thus **bnick*, **bnark*, **bniddle* are not words of English. There simply are no words in English which begin with *bn*. Moreover, native speakers of English have a clear intuition that hypothetical **bnick*, **bnark*, **bniddle* could not be words of English. Similarly, there are no words in English which are pronounced with *pn* at the beginning, a fact which is not only demonstrated by the systematic lack of words such as **pnark*, **pnig*, **pnilge*,

1 The asterisk is used to indicate that a given word is non-existent or wrong.

but also by the fact that the word spelled *pneumonia* which derives from Ancient Greek (a language which does allow such consonant combinations) is pronounced without *p*. A description of the phonology of English would then provide a basis for characterizing such restrictions on sequences of sounds.

Variations in pronunciation. In addition to providing an account of possible versus impossible words in a language, a phonological analysis will explain other general patterns in the pronunciation of words. For example, there is a very general rule of English phonology which dictates that the plural suffix on nouns will be pronounced as [ɨz], represented in spelling as *es*, when the preceding consonant is one of a certain set of consonants including [š] (spelled *sh*) as in bushes, [č] (spelled as *ch*) as in *churches*, and [ǰ] (spelled *j*, *ge*, *dge*) as in *cages*, *bridges*. This pattern of pronunciation is not limited to the plural, so despite the difference in spelling, the possessive suffix *s*[2] is also subject to the same rules of pronunciation: thus, plural *bushes* is pronounced the same as the possessive *bush's*, and plural *churches* is pronounced the same as possessive *church's*.

This is the sense in which phonology is about the sounds of language. From the phonological perspective, a "sound" is a specific unit which combines with other such specific units, and which represent physical sounds.

1.2 Phonetics – what is physical sound?

Phonetics, on the other hand, is about the concrete, instrumentally measurable physical properties and production of these cognitive speech sounds. That being the case, we must ask a very basic question about phonetics (one which we also raise about phonology). Given that phonetics and phonology both study "sound" in language, what *are* sounds, and how does one *represent* the sounds of languages? The question of the physical reality of an object, and how to represent the object, is central in any science. If we have no understanding of the physical reality, we have no way of talking meaningfully about it. Before deciding *how* to represent a sound, we need to first consider *what* a sound is. To answer this question, we will look at two basic aspects of speech sounds as they are studied in phonetics, namely **acoustics** which is the study of the properties of the physical sound wave that we hear, and **articulation**, which is the study of how to modify the shape of the vocal tract, thereby producing a certain acoustic output (sound).

1.2.1 Acoustics

A "sound" is a complex pattern of rapid variations in air pressure, traveling from a sound source and striking the ear, which causes a series of neural signals to be received in the brain: this is true of speech, music and random noises.

2 This is the "apostrophe *s*" suffix found in *The child's shoe*, meaning 'the shoe owned by the child.'

Waveforms. A concrete way to visually represent a sound is with an acoustic **waveform**. A number of computer programs allow one to record sound into a file and display the result on the screen. This means one can visually inspect a representation of the physical pattern of the variation in air pressure. Figure 1 gives the waveforms of a particular instance of the English words *seed* and *Sid*.

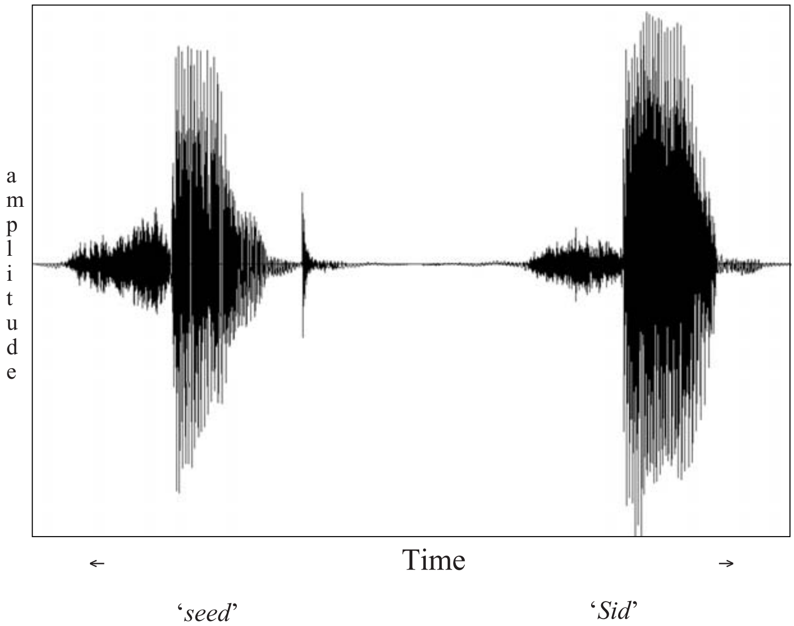

FIGURE 1
Waveforms of speech

The horizontal axis represents time, with the beginning of each word at the left and the end of the word at the right. The vertical axis represents displacement of air particles and correlates with the variations in atmospheric pressure that affect the ear. Positions with little variation from the vertical center of the graph represent smaller displacements of air particles, such as the portion that almost seems to be a straight horizontal line at the right side of each graph. Such minimal displacements from the center correspond to lower amplitude sounds. The portion in the middle where there is much greater vertical movement in the graph indicates that the sound at that point in time has higher amplitude. While such a direct representation of sounds is extremely accurate, it is also fairly uninformative.

The difference between these words lies in their vowels (*ee* versus *i*), which is the part in the middle where the fluctuations in the graph are greatest. It is difficult to see a consistent difference just looking at these pictures – though since these two vowels *are* systematically distinguished in English, it cannot be impossible. It is also very difficult to see similarities looking at actual waveforms. Consider figure 2 which gives different repetitions of these same words by the same speaker.

Absolute accuracy is impossible, both in recording and measurement. Scientific instruments discard information: microphones have limits on what they can capture, as do recording or digitizing devices. Any representation of a sound is a measurement, which is an idealization about an actual physical event.

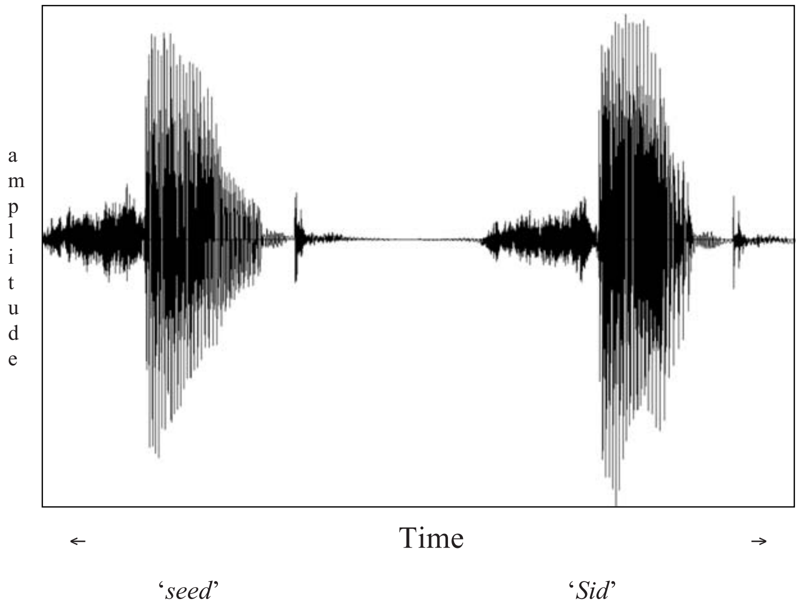

FIGURE 2
Different repetitions of words

Visual inspection gives you no reason to think that these sets of graphs are the same words said on different occasions. The problem is that while a physical waveform is a very accurate representation of a word, it provides so much information that we cannot tell what is important and what is not.

Since we are interested in the part which makes these two words sound different, we might get a clearer picture of the physical difference by expanding the scale and looking just at a part of the vowel. Vowels are **periodic**, which means that the pattern of their waveform repeats over time. The display in figure 3 gives a portion of the

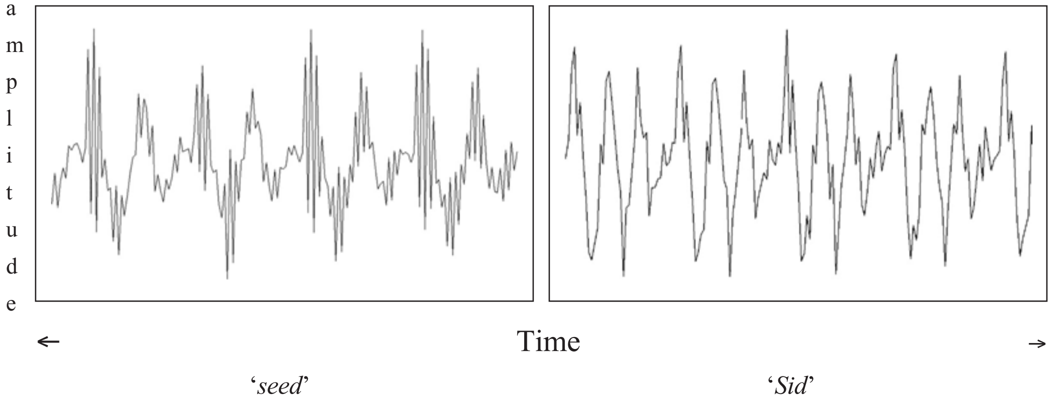

FIGURE 3
Closeup waveform of vowels of *seed*, *Sid*

vowels from the middle of the words *seed* and *Sid*, involving around 30 milliseconds (ms) of each of the words (the entire word in each of these two examples actually lasts approximately 600 ms, so this is a small part of the entire word). We can indeed see that there is a pattern which is repeated (although successive repetitions are not perfect reproductions).

Though there are visible differences between the waveforms, the basis for distinguishing these vowels remains unclear.

Sound spectra. We need a better analytical technique than just looking at raw sound, to be able to talk precisely about properties of these sounds. We therefore need to understand some basic properties of physical sounds. All sound waves are definable in terms of three properties that characterize a **sine wave** familiar from trigonometry, namely **frequency** measured in cycles per second also known as Hertz (Hz), **amplitude** measured in decibels (dB), and **phase** measured in the angular measure radians. These characteristics suffice to define any sine wave, which is the analytic basis of sounds. The property phase, which describes how far into the infinite cycle of repetition a particular sine wave is, turns out to be unimportant for the study of speech sounds, so it can be ignored. Simple sine waves (termed "pure tones" when speaking of sounds) made up of a single frequency are not commonly encountered in the real world, but can be created by a tuning fork or by electronic equipment.

Speech sounds (indeed all sounds) are complex waveforms which are virtually impossible to describe with intuitive descriptions of what they "look like." Fortunately, a complex waveform can be mathematically related to a series of simple waves which have different amplitudes at different frequencies, so that we can say that a complex waveform is "built from" a set of simple waves. Figure 4 shows a complex wave on the left which is constructed mathematically by just adding together the three simple waveforms of different frequencies and amplitudes that you see on the right.

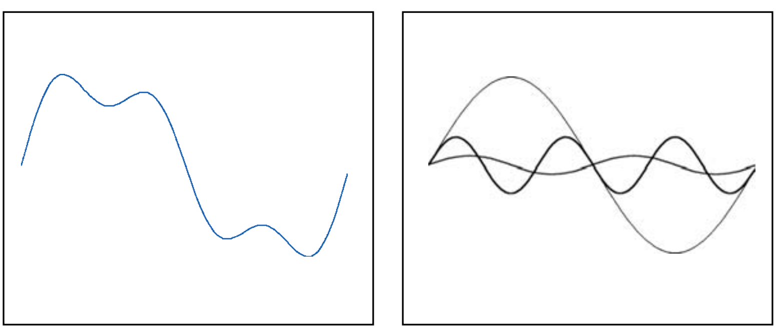

FIGURE 4
Complex wave and the component simple waves defining it

The complex wave on the left is made from simple sine waves at 100, 200, and 300Hz, and the individual components defining the complex wave are graphed on the right. The most prominent component (the one with the highest amplitude) is the one at 100 Hz, the thinnest line which makes one cycle in the chart: it has an amplitude of 60 dB. By comparison, the component at 200 Hz (graphed with a medium-weight line, which makes two cycles in the chart on the right) has the lowest amplitude, 40 dB. The 300 Hz component, graphed with the thickest line, has an intermediate amplitude of 50 dB. It is the amplitudes of the individual components which determine the overall shape of the resulting complex wave.

Now we will see what happens when we change this artificial sound to make the 200 Hz component be the most prominent component and the 100 Hz one be less prominent – if we simply switch the amplitudes of the 100 Hz and 200 Hz components, we get the wave shown in figure 5.

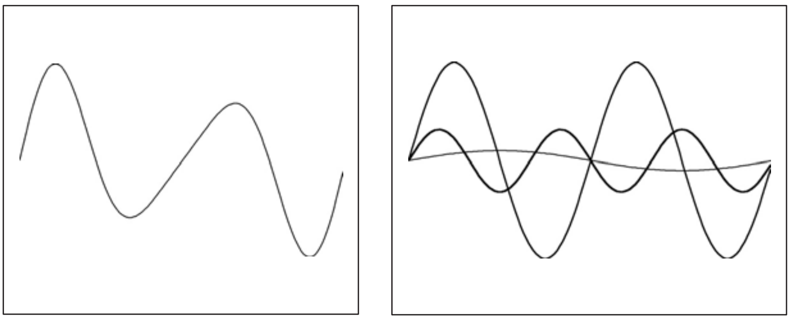

FIGURE 5
Effect of changing component amplitude

Changing the amplitude of one such component changes the overall character of the waveform. A complex wave is mathematically equivalent to a corresponding series of sine wave components, so describing a complex wave directly is equivalent to describing the individual components. If we see two differently shaped complex waves and we can't describe their differences directly in terms of the complex waves, we can instead focus on the equivalent series of sine wave components, and describe the differences in terms of very simple information about component frequency and amplitude.

Just as a single complex waveform can be constructed from a series of simple waves at different frequencies and amplitudes, a single complex waveform can also be mathematically broken down into a series of components which have different frequencies and amplitudes. Rather than graph the full shape of each specific sine wave component – which becomes very hard to understand if there are more than a handful of components – we can simply graph the two important values for each of the component sine waves, the amplitude and frequency. This is known as a **spectrum**: it is the defining frequency and amplitude components of a complex waveform, over a fixed period of time. The spectrum of the waveform in figure 4 is plotted in figure 6, where the horizontal axis corresponds to frequency from 0 to 7,000 Hz and the vertical axis corresponds to amplitude from 0 to 60 dB. Note that in this display, time is not represented: the spectrum simply describes amplitude and frequency, and information about how long a particular complex waveform lasts would have to be represented somewhere else.

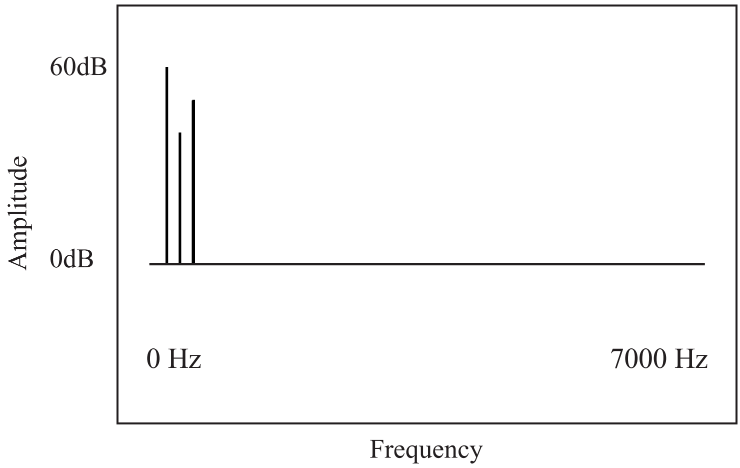

FIGURE 6
Spectrum

This is a very simple spectrum, representing an artificially constructed sound containing only three components. Naturally occurring sounds have many more components than this.

Since complex sounds can be mathematically broken down into a series of simple components, we can use this very useful tool to look at the vowel sounds of *seed* and *Sid*: we look at the physical properties of the component frequencies that define the two vowels that we were interested in. Figure 7 provides the spectrum of the portion from the middle of the vowels of *Sid* and *seed* which we looked at in figure 3. The horizontal axis again represents frequency, ranging from 0 to 7000 Hz, and the vertical axis represents amplitude in decibels. Here, the spectrum is represented as a continuous set of amplitude values for all frequencies in this frequency range, and not just three discrete frequencies as seen in the constructed sound of figure 6.

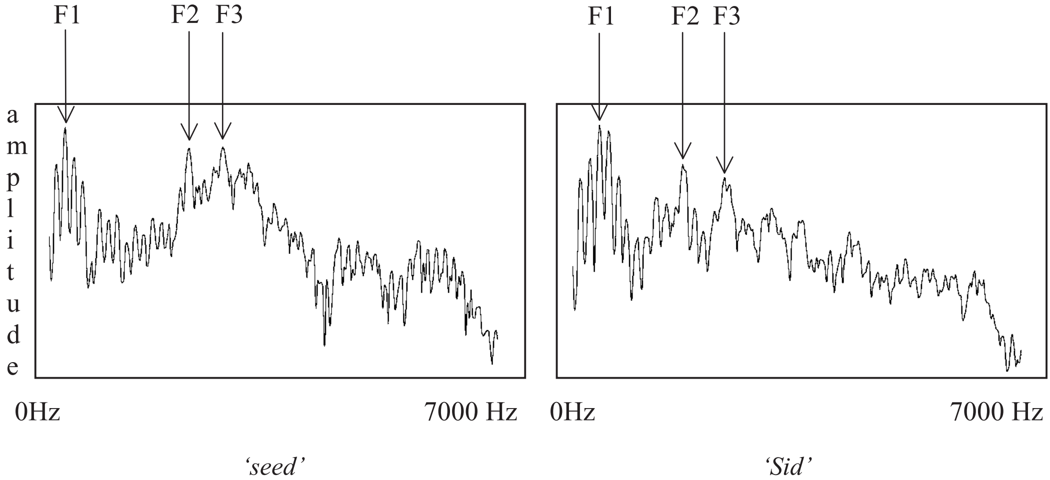

FIGURE 7
Spectrum of the vowels of *seed*, *Sid*

Inaccuracy in spectral data has three main sources. Half of the information in the original signal, phase, has been discarded. Frequency information is only approximate and is related to how much speech is analyzed. Finally, a spectrum assumes that sound properties are constant during the period being analyzed. If too large a piece of speech is taken for analysis, a misrepresentative blending of a continuously changing signal results.

In these spectra, certain frequency regions are more prominent than others, due to **resonances** in the vocal tract. Resonances are frequency regions where sound amplitude is enhanced. These frequencies are perceptually more prominent than other lower-amplitude frequencies. The frequencies at which these resonances occur are related to the length of various parts of the vocal tract (ultimately related to the position of the tongue and lips as specific sounds are made). The relation between size and frequency is simple and familiar: a large bottle has a low-resonance frequency and a small bottle has a higher-resonance frequency. The first three of these prominent frequency regions, called **formants**, are indicated with pointed vertical lines in the graphs. You can see that in the spectrum for *seed* on the left, the first formant (F1) occurs at a lower frequency than the first formant of the vowel in *Sid*. However, the second and third formants (F2, F3) of *seed* occur at somewhat higher frequencies than F2 and F3 of *Sid*. By comparing the frequencies at which these formants occur, one can begin to systematically describe the physical properties of the vowels in *seed* and *Sid*. One of the most important properties which allows a listener to distinguish speech sounds, such as the vowels of *seed* versus *Sid*, is the frequencies of these formants.

Viewing the waveform versus the spectrum of a sound involves a tradeoff between accuracy and usefulness. While the spectrum is more informative since it allows us to focus on certain specific properties (formant frequencies), it is a less accurate representation of reality than the original waveform. Another very significant limitation of this type of spectral display is that it only characterizes a single brief moment in the utterance: speech is made up of more than just little 30 millisecond bits of steady sound. We need to include information about changes over time in a sound.

A spectrogram can be made by a mechanical spectrograph, which uses an adjustable filter to select different frequency ranges and display the changes in amplitude at each frequency range; or, it can be created by a computer program, which uses Fourier analysis to determine these component amplitudes.

Spectrograms. Another display, the **spectrogram**, shows both frequency and amplitude properties as they change over time, by adding a third dimension of information to the display. Figure 8 provides spectrograms of the entirety of the two words *seed* and *Sid*. In this display, the horizontal axis represents the time dimension: the utterance begins at the left and ends at the right. The vertical axis represents frequency information, lower frequencies appearing at the bottom and higher frequencies at the top. Amplitude is represented as darkness: higher amplitudes are darker and lower amplitudes are lighter.

The initial portion of the spectrogram between the arrows represents the consonant *s*, and the second portion with the series of minute vertical striations represents the vowel (the consonant *d* is visible as the light horizontal band at the bottom of the graph, beginning at around 500 ms). The formants which characterize the vowels of *seed* and *Sid* are represented as dark bands, the first formant being the darker lower band and the second and third formants being the two somewhat lighter bands appearing approximately one-third of the way up the display.

Looking at these spectrograms, we learn two other things about these vowels that we would not have suspected from looking at the spectrum in