On wide-band and narrow-band spectrograms

Fort Worth: Harcourt, pp 190-192.

The spectrograms that have been used to illustrate this chapter so far are called wide-band spectrograms. They are very accurate in the time dimension. They show each vibration of the vocal folds as a separate vertical line and indicate the precise moment of a stop burst with a vertical spike. But they are less accurate in the frequency dimension. There are usually several component frequencies present in a single formant, all of them being lumped together in one wide band on the spectrogram.

It is a fact of physics that one can know either fairly precisely when a sound occurred or, to a comparable degree of accuracy, what its frequency is. This should be intuitively clear when you recall that knowing the frequency of a sound involves observing the variations in air pressure over a period of time. This period of time has to be long enough to ensure observations of a number of repetitions of the variations of air pressure. You can either know that a pulse from the vocal folds has happened (producing the vertical voicing striation in all the spectrograms we have considered so far), or, if the piece of the sound wave being analyzed contains two or three pulses of the vocal folds, we can tell how far apart they are and hence know the frequency.

Spectrograms that are more accurate in the frequency dimension (at the expense of accuracy in the time dimension) are called narrow-band spectrograms. Figure 8.20 shows both wide- and narrow-band spectrograms of the question “Is Pat sad, or mad?” In the wide-band spectrogram, there are sharp spikes at the release of each stop, for example, for the /d/ at the end of the utterance. The spikes are smeared in the time dimension in the narrow-band spectrogram. But the frequencies that compose each formant are visible.

When the vocal folds vibrate, they produce what are called harmonics of their fundamental frequency of vibration. Harmonics are vibrations at whole-number multiples of the fundamental frequency. Thus when the vocal folds are vibrating...
at 100 Hz, they produce harmonics at 200, 300, 400 Hz, and so on. In a given vowel, the particular harmonics that are evident are those that correspond to the resonances of the vocal tract shape occurring in that vowel. I have put two small white squares in the middle of the fifth, tenth, and fifteenth harmonics in the middle of the vowels in “sad” and “mad.” The vocal folds are vibrating at about 118 Hz in “sad,” so the fifth harmonic has a frequency of $5 \times 118 = 590$ Hz, the tenth harmonic a frequency of 1,180 Hz, and the fifteenth harmonic a frequency of 1,770 Hz. The first formant is formed by the fifth and sixth harmonics, and the main components of the second formant are the fourteenth and fifteenth harmonics. Compare this with the vowel in “mad,” which has very similar formants, both being examples of the /æ/ phoneme. Near the beginning of the last word the third harmonic is the main component of the first formant and the eighth harmonic the main component of the second formant. As we have noted, the quality of a vowel sound depends on the frequencies of the formants. But the pitch depends on the fundamental frequency, which is determined by the rate of vibration of the vocal folds.

In women’s voices, which usually have a higher pitch, the formants are sometimes more difficult to locate precisely. Figure 8.21 show spectrograms of a female speaker of American English saying the same set of vowels as those of the male speaker in Figure 8.7. Even though these spectrograms have been made with considerable care, choosing the most appropriate degree of narrowness of the spectrogram to best show the formant frequencies, the harmonics still interfere with the display of the formants. Notice, for example, the change in vowel quality in the vowel [u], which appears as a series of steps as different harmonics become available to make up the formant. In a narrow band spectrogram it is even more difficult to locate the centers of the formants when the fundamental frequency is high.

Narrow-band spectrograms are obviously useful for determining the intonation—or tone—of an utterance. One can do this by looking at the fundamental frequency itself, but when this goes from, say, 100 to 120 Hz, the frequency of the tenth harmonic will go from 1,000 to 1,200 Hz, which is much easier to see. The actual pitch—or, to be more exact, the fundamental frequency—at any moment will be one-tenth that of the tenth harmonic. As we saw earlier in this chapter (and also in Chapter 5), computers can analyze speech to give a good record of the fundamental frequency (the pitch). But most fundamental frequency routines make occasional errors when the pitch is too low or when the vocal folds are not vibrating regularly. In these cases, a narrow-band spectrographic analysis can be very useful.