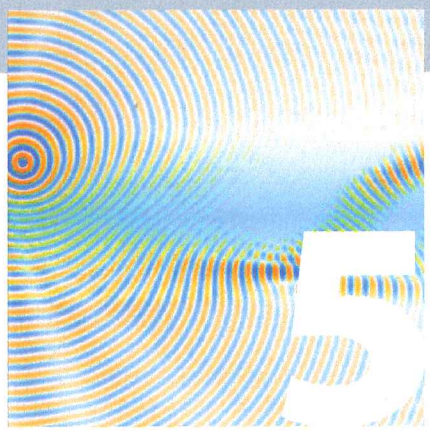

Sonograms

5.1 What Is a Sonogram?

The need for sonograms is made obvious by music or speech: we hear distinct changes in sound from moment to moment, which we would like to follow rather than lump into a single power spectrum for an entire sound sample. In other words, there is need for a kind of mixed time-frequency analysis, one that reveals how the frequency content changes with time. To capture the variations, the sound is cut into pieces to be separately analyzed. This is done with a moveable window $w(t, t_0)$ centered on time t_0 of width T. The uncertainty principle demands that the power spectrum is broadened by an amount $\Delta f = 1/T$, where T is the width of the window.

The sound signal $s(t)$ is multiplied by the window function $w(t, t_0)$ to give a windowed segment $s_w(t, t_0) = w(t, t_0)s(t)$ for which a power spectrum $p(f, t_0)$ is calculated. The window function $w(t, t_0)$ peaks at time t_0 and has a width that is under the user's control. The power spectrum $p(f, t_0)$ is thus a function of f, t_0, and the width of the window. Color or grayscale density is used to plot this data as a function of f and t_0—that is, both frequency and time. This is the *sonogram* (also sometimes called a *spectrogram*). The horizontal axis is taken to be time; the vertical axis is frequency; brightness encodes the spectral power.

A 1-second-long window imposes a 1 Hz uncertainty in the frequency, a 0.1 s window creates a 10 Hz smearing of the frequency. Most transient changes in a sound might be captured this way; 300 different power spectra can be generated in 30 seconds. To resolve a rapidly varying sound, the window length might be lowered to only 0.01 second, which smears the frequency by 100 Hz.

The sonogram fills the gulf between the extremes of the *time domain* (the sound amplitude at every instant) and the *frequency domain* (the

Figure 5.1
Violin part for Bach's Concerto in D Major for violin duet and string orchestra. Courtesy theviolinsite.com.

power spectrum for a long sample of sound). If the time window is short, the sonogram will be time-like. If the window length is relatively long, the sonogram is frequency-like. There is no universal window that is best, but rather different kinds of sounds reveal secrets best with different time windows. In many cases, it is desirable to analyze the same sound with several different window widths. A sonogram may be created in real time, on the fly, so to speak, or it may be a postanalysis of a recording.

A sonogram is a cousin of musical notation; see figure 5.1. In musical notation, the time is only loosely proportional to the horizontal distance on the page, and the notes give only the lowest, or *fundamental*, frequency to be played, whereas in sonograms the higher harmonics of a given note will also appear. (*Harmonics* are integer multiples of a given frequency—for example, 200 Hz, 300 Hz, and so on are harmonics of 100 Hz.) The frequency (vertical) axis in musical notation is logarithmic, since each octave is a factor of 2 higher in frequency, so octaves are equally spaced on the musical staff. Thus A220 and A440 are an octave apart, separated by 220 Hz and a certain distance apart on the scale, but A440 and A880 are also an octave apart and the same distance apart on the scale, but separated by twice as much in frequency, 440 Hz. Sometimes it is best to plot log(frequency) rather than frequency in sonograms too.

The sonogram on the left side of figure 5.2 is straightforward to read. A nearly pure sine wave of slowly increasing frequency was present, starting around 1100 Hz and rising to about 1250 Hz after half a second. Over the entire 1-second interval, the frequency swept from 1100 Hz to about 1500 Hz.

On the right, the sonogram for a rising complex tone is seen. The sound is a voice singing "ah," as in "Oprah," with a rising pitch. At any given time, a set of equally spaced partials is present. The vocal tract was held fixed, and only the pitch produced by the vocal folds was changed. Note that there is a wide frequency range, from about 600 to 1000 Hz, where partials are stronger as they pass through. This is our first glimpse of a *formant*, a whole zone of frequencies enhanced by the resonances of the vocal tract shaped for this vowel. The sonogram has already led to a discovery!

The higher partials in figure 5.2 are rising faster than the lower partials, since, for example, a factor of 2 rise in pitch means a 100 Hz rise for the lowest 100 Hz partial, but a 500 Hz rise for the fifth partial starting at 500 Hz.

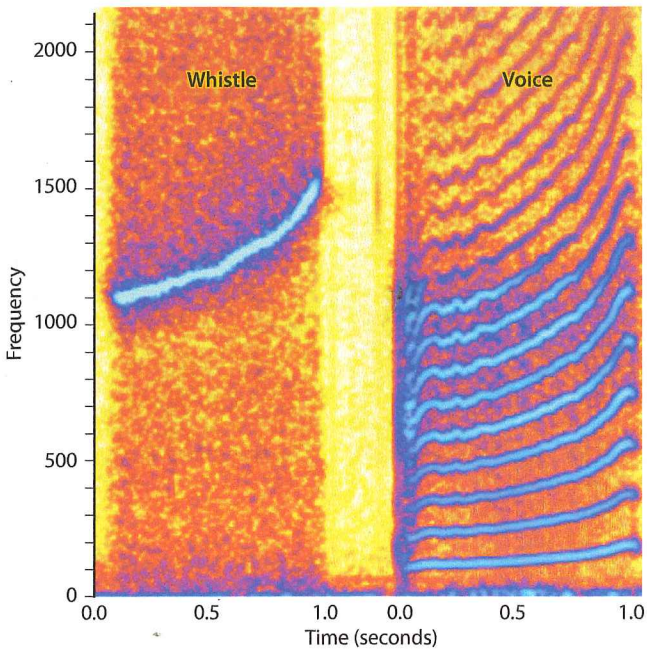

Figure 5.2

On the left, a sonogram of a rising whistle (starting at about 1100 Hz and ending at 1500 Hz) is shown. On the right, a rising voice singing "ah" (starting at about 100 Hz and ending at about 200 Hz) is shown. The whistle is very nearly a pure sine wave; the voice has many partials (overtones) above the first partial. A formant characteristic of the "ah" sound is seen as a zone of higher intensity between about 600 and 1000 Hz. Partials grow louder as they enter this zone, softer as they leave it.

5.2 Choosing Sonogram Parameters

When making a sonogram, we encounter the time-frequency uncertainty principle twice: once for the source, and once for the analysis. The source will typically have its own characteristic time or times over which it changes in significant ways. The sonogram analysis performs separate frequency analysis at different times, but it must choose the length of the window over which it does each analysis. (The time axis on the sonogram is really the time at the center of an interval or "window" of time over which the sound is sampled and analyzed.)

The frequency smearing in a sonogram can be due to the sound itself or the window length. Once the window function w_t is chosen, the algorithm performs a power spectrum analysis with the window centered at many different times. No matter how long the sound lasts, the computer examines the data in slices, as if the sound came in bursts.

There is no universal best setting for the window width T. Often a good choice for a width is the typical time for the source to change its spectral content significantly. Figure 5.3 shows sonograms for a short section of the soundtrack for the movie *The Good, the Bad, and the Ugly*, in which a whistle oscillates rapidly between two pitches. If the time window is taken to be too long, the rapid variation is washed out, smeared in the time direction. If, on the other hand, the time resolution is too short, the frequency difference between the two notes is washed out. The window on

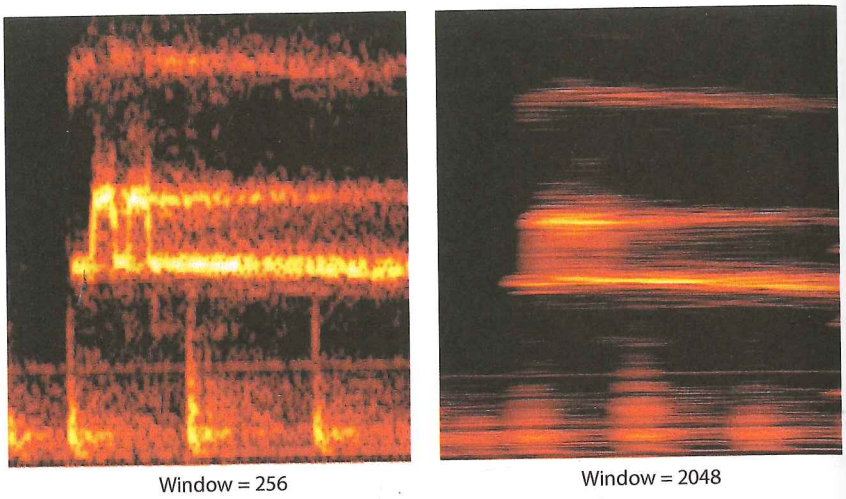

Figure 5.3

Two sonograms of the famous whistling in the theme song for the movie *The Good, the Bad, and the Ugly*. The whistler rapidly switches between two frequencies differing by the interval of a fourth, two notes low and two high. If the sample window is too long, the rapid transitions between notes are obliterated (right).

the left is a good compromise between the time and frequency uncertainties for this passage.

In figure 5.4, we examine a sonogram of a short pulse centered at 220 Hz, with three different window sizes. The shortest window is 6 ms long, corresponding to a frequency uncertainty of about 166 Hz. The longest is ten times as along, 60 ms, corresponding to a frequency uncertainty of 17 Hz. A third window is intermediate between these two. Note the corresponding time and frequency uncertainties on the plots. The signal is of duration approximately 16 ms, corresponding to an intrinsic frequency uncertainty of 63 Hz. Even if a much longer widow is used (left), the frequency uncertainty in the sonogram does not go below 63 Hz. It can grow larger than that if a shorter window is used, as at the right.

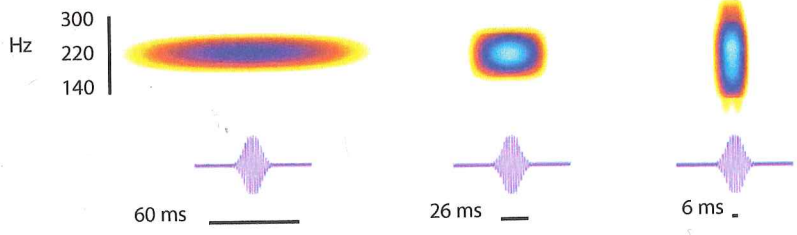

Figure 5.4

A pulse of duration about 16 ms analyzed with three different window functions, of duration 60, 26, and 6 ms.

Capturing and Re-creating Sound

6.1 Galileo—The First Recording?

Galileo Galilei (1564–1642) may have been the first person to record sound and have a perfectly clear understanding that it was indeed a permanent record. We quote from the dialog in *The Two New Sciences*, 1638, which exudes a deeper understanding of sound and hearing than his contemporaries were prepared to accept. Galileo speaks first of water waves excited when a wet finger is run around the rim of a full glass and then notes that if the tone jumps an octave, which it sometimes does, so too the waves reduce their wavelength by half. Galileo writes about the possibility of a permanent sound recording, speaking though his imaginary interlocutors Salviati and Sagredo:

> *Salviati*: This is a beautiful experiment enabling us to distinguish individually the waves which are produced by the vibrations of a sonorous body, which spread through the air, bringing to the tympanum of the ear a stimulus which the mind translates into sound. But since these waves in the water last only so long as the friction of the finger continues and are, even then, not constant but are always forming and disappearing, would it not be a fine thing if one had the ability to produce waves which would persist for a long while, even months and years, so as to easily measure and count them?
>
> *Sagredo*: Such an invention would, I assure you, command my admiration.
>
> *Salviati*: The device is one which I hit upon by accident; my part consists merely in the observation of it and in the appreciation of its value as a confirmation of something to which I had given profound consideration; and yet the device is, in itself, rather common. As I