events and thus gives poor time resolution, as with a plosive that is closely spaced to a succeeding voiced sound.

Wideband Spectrogram — For the *wideband* spectrogram, we choose a "short" window with a duration of less than a single pitch period (Figure 3.14); shortening the window widens its Fourier transform (recall the uncertainty principle). The wide Fourier transform of the window, when translated to harmonics, will overlap and add with its neighboring window transforms and smear out the harmonic line structure, roughly tracing out the spectral envelope $|\tilde{H}(\omega)|$ due to the vocal tract and glottal flow contributions. In an alternative temporal perspective, since the window length is less than a pitch period, as the window slides in time it "sees" essentially pieces of the periodically occurring sequence $\tilde{h}[n]$ (assuming tails of previous responses have died away). For the steady-state voiced sound, we can therefore express the wideband spectrogram (very) roughly (Exercise 3.9 asks the reader to complete the argument) as

$$S(\omega, \tau) \approx \beta|\tilde{H}(\omega)|^2 E[\tau] \tag{3.6}$$

where β is a constant scale factor and where $E[\tau]$ is the energy in the waveform under the sliding window, i.e., $E[\tau] = \sum_{n=-\infty}^{\infty} |x[n, \tau]|^2$, that rises and falls as the window slides across the waveform. In this case, where the window $w[n, \tau]$ is short, and less than a pitch period, the spectrogram shows the formants of the vocal tract in frequency, but also gives vertical striations in time every pitch period, rather than the harmonic horizontal striations as in the narrowband spectrogram. These vertical striations arise because the short window is sliding through fluctuating energy regions of the speech waveform.

In our description of the narrowband and wideband spectrograms, we have used the example of voiced speech. Similar reasoning can be made for fricative and plosive sounds. With regard to fricatives, the squared STFT magnitude of noise sounds is often referred to as the *periodogram*, which is characterized by random wiggles around the underlying function $|\tilde{H}(\omega)|^2$. The periodogram is developed formally in a stochastic process framework later in the text. For plosives, the spectrogram reveals the general spectral structure of the sound as the window $w[n, \tau]$ slides across the signal. For these sound classes, both the narrowband and wideband spectrograms show greater intensity at formants of the vocal tract; neither, however, typically shows horizontal or vertical pitch-related striations because periodicity is not present except when the vocal folds are vibrating simultaneously with these noise or impulsive sounds. With plosive sounds, the wideband spectrogram is often preferred because it gives better temporal resolution of the sound's components, especially when the plosive is closely surrounded by vowels.

Figure 3.15 compares the narrowband (20-ms Hamming window) and wideband (4-ms Hamming window) spectrograms for a particular utterance. The spectrograms were computed with a 512-point FFT. For the narrowband spectrogram, the 20-ms Hamming window was shifted at a 5-ms frame interval, and for the wideband spectrogram, the 4-ms Hamming window was shifted at a 1-ms frame interval. Both spectrograms reveal the speech spectral envelope $|\tilde{H}(\omega)| = |H(\omega)G(\omega)|$ consisting of the vocal tract formant and glottal contributions. Notice, however, the distinctive horizontal and vertical striations in the narrowband and wideband spectrograms, respectively. Observe, however, that occasionally the vertical striations are barely visible in the wideband spectrogram when the pitch is very high. Observe also a difference in