Ingo R. Titze, Associate Editor

Relations Between Time and Frequency in Signal Analysis

Ingo R. Titze

Journal of Singing, March/April 2020
Volume 76, No. 4, pp. 437–439

Vocal sound is often analyzed in one of two ways, either in terms of the acoustic pressure variation over time (e.g., a microphone signal), or in terms of the frequency content in the pressure signal (e.g., a frequency spectrum). Signal analysis can be carried out equally well in the time domain or the frequency domain. A time waveform or a frequency spectrum can be displayed and processed to extract acoustic features. This dual representation is perfectly equivalent and exact for a steady sound that has no beginning and no end. Perfect mathematical transformations exist between them. A forward Fourier transform takes the time-dependent pressure signal $p(t)$ and converts it into a frequency-dependent pressure signal $P(f)$, where f is the frequency. Similarly, a backward Fourier transform takes a pressure spectrum $P(f)$ and converts it into a time-dependent pressure waveform $p(t)$.

When a sound is not constant over time, or limited with a beginning or an end, the transformations are not perfect. There is a fundamental uncertainty principle that states that the *spread in frequency multiplied by the spread in time equals a constant*. This implies that the uncertainty (or error) in frequency and the uncertainty (or error) in time cannot both be zero. If we want the frequency components to be perfectly accurate, we need an infinitely long window of time for observation. Similarly, if we want the time analysis to be perfectly accurate, we need an infinitely wide spectrum of frequencies. In practice, the limitations are handled with a compromise. We limit the waveform with a time window (a gradual start-up and a gradual decay) and we limit the frequency analysis with a finite band-width. For a window in the shape of a Gaussian distribution, the relation can be written as

$$\text{(bandwidth)* (window length)} = 1.2982804$$

Other windows are used in signal processing, such as Hanning, Hamming, or Cosine, for which the above relation differs slightly, but that is not important for this discussion.

As an example, a window length of 0.005 seconds yields a 260 Hz bandwidth, which defines the frequency resolution. For vocalization, such a bandwidth is insufficient to identify much detail in the source frequencies. Energy from more than one harmonic may be combined in this band and smeared together. On the other hand, a 0.100 second window yields a 13 Hz bandwidth, a resolution high enough to identify detail in the source frequen-

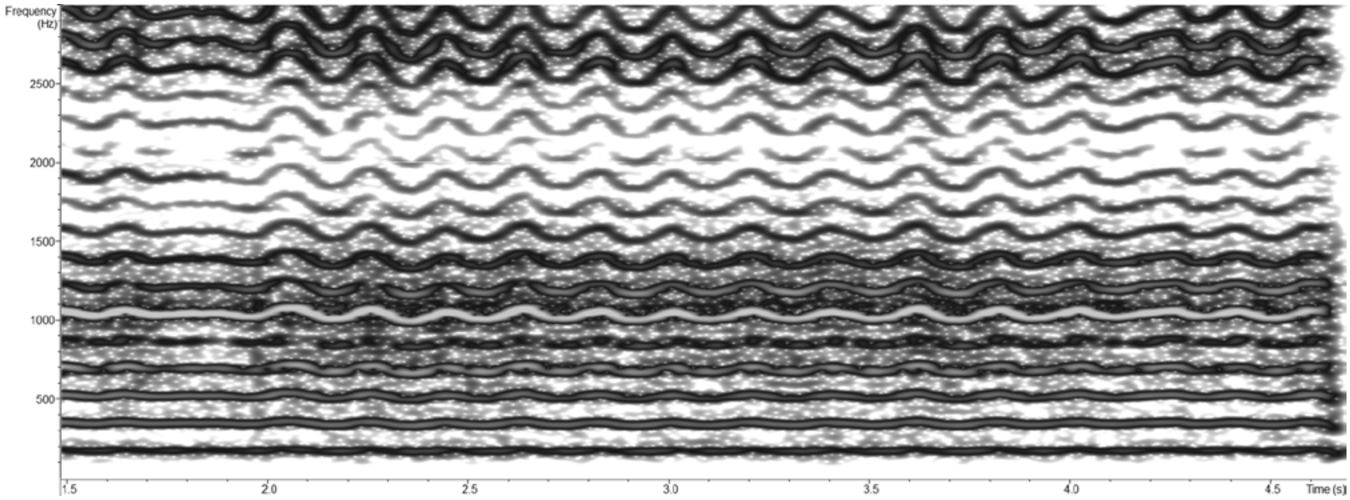

Figure 1. Narrow-band spectrogram of an /ɑ/ vowel produced by a male at 180 Hz fundamental frequency. The analysis bandwidth was 5 Hz.

cies. When digital signals are analyzed, a maximum frequency of interest is also applied, which determines the sampling rate (generally at least twice the highest frequency of interest).

A spectrogram is a good example of the trade off between frequency and time resolution because frequency is on one axis and time is on the other axis. A so-called narrow-band spectrogram (Figure 1, black on white background) reduces the uncertainty in frequency, making individual frequency components visible and clear. Here, a male singer produced a steady /ɑ/ vowel at about 180 Hz fundamental frequency with a small amount of vibrato. Harmonic lines are stacked in the vertical direction, well separated from each other because the analysis bandwidth was chosen to be narrow, 5 Hz. On the horizontal time axis, however, the cycle to cycle pressure variation is not visible except for the vibrato. For the fundamental frequency, energy from one cycle spreads over adjacent cycles because the window is wide, here about 0.26 seconds. The fundamental period is 0.0055 seconds. Time variation of the vibrato cycle is visible because this variation is relatively slow. It has a period of about 0.18 seconds, on the order of the analysis window.

In a similar but opposite way, a wide-band spectrogram (Figure 2, white on black background for slightly better clarity) smears individual frequencies together (in the vertical direction), but preserves the cycle to cycle pressure-pulse variations of the fundamental frequency on the time axis. In the two figures, the same signal was analyzed. Only the analysis window (and therewith the bandwidth) was changed, from 5 Hz to 172 Hz. The vertical striations represent individual time periods (cycles) of vibration, but harmonics are not visible because multiple frequencies bleed energy into the same 172 Hz band. In this spectrogram, the fundamental frequency can be determined by counting the number of periods over a unit of time. That will agree with the frequency distance between horizontal lines in Figure 1. The wide-band spectrogram is generally used for formant analysis because formant bandwidths are on the order of 100 Hz. They do not require the frequency discrimination that source frequencies do.

In summary, vocologists who analyze vocal sounds need to understand and not be frustrated with uncertainties in frequency and time resolution in spectrograms. It is simply part of nature. If there are rapid formant transitions or rapid fundamental frequency changes in an utterance, the uncertainties become even more restrictive. In a rapid pitch glide, for example, that changes on the order of an octave in 0.5 seconds, a 0.05 second window may be chosen to track the glide with 10 points. The frequency resolution is then on the order of 26 Hz for each point, which may or may not be enough to identify a smooth glide, depending

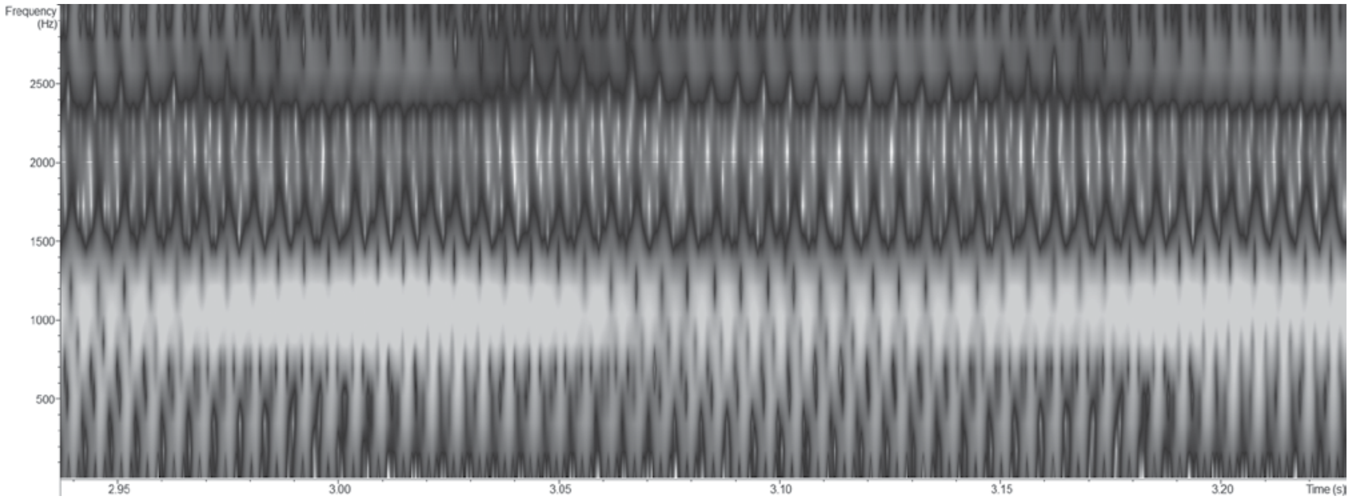

Figure 2. **Wide-band spectrogram of the same /ɑ/ of Figure 1, but analyzed with a 172 Hz bandwidth.**

on the absolute frequency range in the octave. It may require repeated analysis with different bandwidths (or time windows). The good news is that signal processing engineers are developing ever more accurate time-frequency analysis tools that are making their way into cost effective applications on small devices.