# *Appendix B* A Biologist's Introduction to Spectrum Analysis

## About this appendix

This appendix provides some conceptual background for making and interpreting spectrograms and spectra with Canary. It introduces the short-time Fourier transform (STFT), the mathematical technique used by Canary for making spectrograms. We treat the STFT here as a black box, but one that has controls on its outside that affect its operation in important ways. One aim of this appendix is to convey enough qualitative understanding of the behavior of this box to allow intelligent use of its controls. Specific details of the controls are covered in Chapter 3. A second aim of this appendix is to explain some of the limitations and tradeoffs intrinsic to spectrum analysis of time-varying signals. More rigorous mathematical treatments of spectral analysis, at several levels of sophistication, can be found in the references listed at the end of the appendix.

Several approaches are available for explaining the fundamentals of digital spectrum analysis. The approach taken in this appendix is geared specifically to spectrum analysis with Canary; thus some of the terms and concepts used here (e.g., "analysis resolution" and "grid resolution") may not appear in other, more general discussions of spectrum analysis, such as those listed at the end of the appendix.

The discussions in this appendix assume a basic understanding of how sound is recorded and represented digitally. If you are not already acquainted with concepts such as sampling rate and amplitude resolution (sample size), read Appendix A.

## Time-domain and frequency-domain representations of sound

Any acoustic signal can be graphically or mathematically depicted in either of two forms, called the *time-domain* and *frequency-domain* representations. In the time domain, the amplitude of a signal is represented as a function of time. Figure B.1a shows the time-domain representation of the simplest type of acoustic signal, a pure tone. Such a signal is called a *sinusoid* because its amplitude is a sine function of time, characterized by some frequency, which is measured in cycles per second, or Hertz (Hz). (In terms of real-world physical quantities, the amplitude may represent a measurement such as the pressure exerted by vibrating air or water molecules, or a voltage at some point in an electric circuit.) In the frequency domain, the amplitude of a signal is represented as a function of frequency. The frequency-domain representation of a pure tone is a vertical line (Figure B.1b).

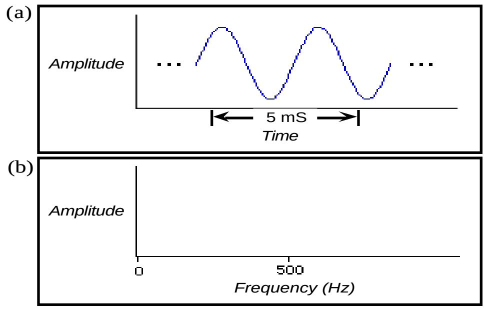


**Figure B.1.** Time-domain and frequency-domain representations of an infinitely long pure sinusoidal signal with a frequency of 500 Hz. **(a)** Time domain. **(b)** Frequency domain.

Any sound, no matter how complex, can be represented as a sum of pure tones (sinusoidal components). Each tone in the series has a particular amplitude, relative to the others, and a particular phase relationship (i.e., it may be shifted in time relative to the other components). The frequency composition of complex signals is usually not apparent from inspection of the time-domain representation. Spectrum analysis is the process of converting the time-domain representation of a signal (which is the representation directly produced by most measuring and recording devices) to a frequency-domain representation that shows how different frequency components contribute to the sound.

The complete frequency-domain representation of a signal consists of two parts. The *magnitude spectrum* (Figure B.2b) contains information about the magnitude of each frequency component in the entire signal. The *phase spectrum* (not shown) contains information about the phase or timing relationships among the frequency components, but in a form that is not easily interpreted. Since the phase spectrum is rarely of practical use in most bioacoustic work and is not provided by Canary, it is not discussed further here. Henceforth, unless otherwise noted, we use the term “spectrum” to refer to the magnitude spectrum alone.

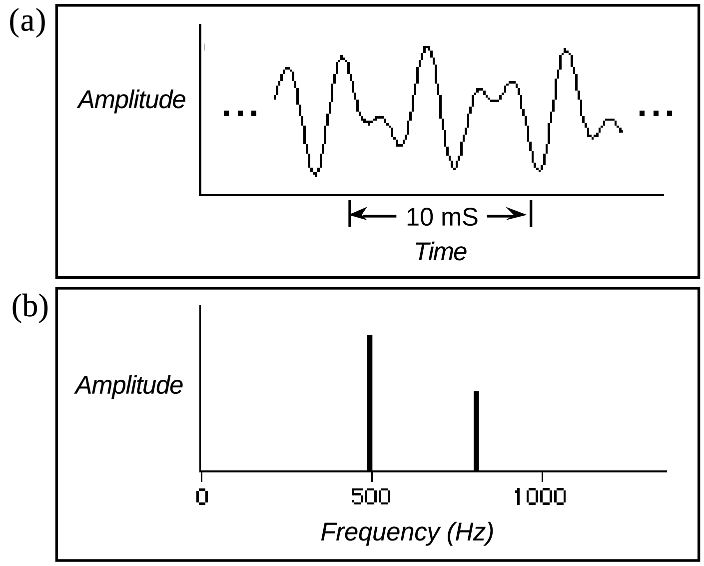


**Figure B.2.** Time-domain and frequency-domain representations of an infinitely long sound consisting of two tones, with frequencies of 490 Hz and 800 Hz. **(a)** Time domain. **(b)** Frequency domain.

The *Fourier transform* is a mathematical function that converts the time-domain form of a signal (which is the representation directly produced by most measuring and recording devices) to a frequency-domain representation, or spectrum. When the signal and spectrum are represented as a sequence of discrete digital samples, a version of the Fourier transform called the *discrete Fourier transform (DFT)* is used. The input to the DFT is a finite sequence of values— the amplitude values of the signal— sampled (digitized) at regular intervals. The output is a sequence of values specifying the amplitudes of a sequence of discrete frequency components, evenly spaced from zero Hz to half the sampling frequency (Figure B.3). Canary implements the DFT using an algorithm known as the *fast Fourier transform (FFT).*

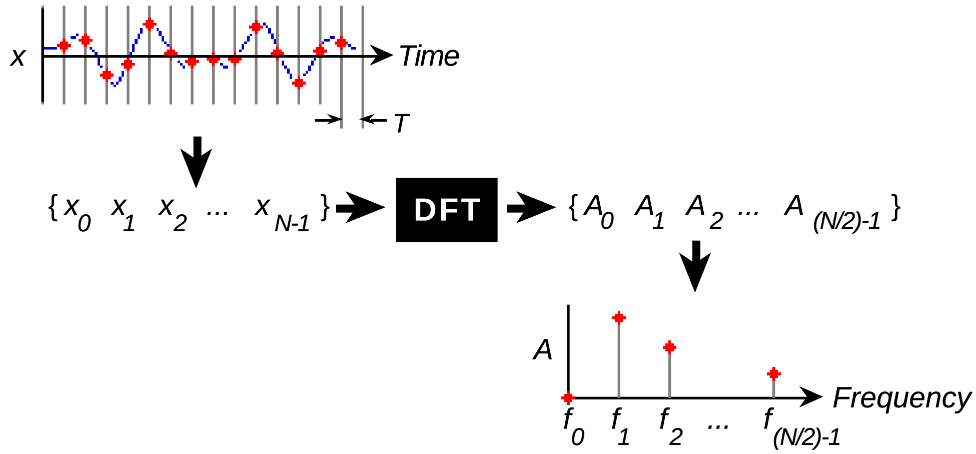


**Figure B.3.** Schematic representation of the discrete Fourier transform (DFT) as a black box. The input to the DFT is a sequence of digitized amplitude values ($x_0$, $x_1$, $x_2$, ... $x_{N-1}$) at *N* discrete points in time. The output is a sequence of amplitude values ($A_0$, $A_1$, $A_2$, ... $A_{(N/2)-1}$) at *N/2* discrete frequencies. The highest frequency, $f_{(N/2)-1}$, is equal to half the sampling rate (= $^1/_{(2T)}$, where T is the sampling period, as shown in the figure). The output can be plotted as a magnitude spectrum.

In practice, a spectrum is always made over some finite time interval. This interval may encompass the full length of a signal, or it may consist of some shorter part of a signal.

## Spectral analysis of time-varying signals: spectrograms and STFT analysis

An individual spectrum provides no information about temporal changes in frequency composition during the interval over which the spectrum is made. To see how the frequency composition of a signal changes over time, we can examine a sound *spectrogram*. The spectrograms produced by Canary plot frequency on the vertical axis versus time on the horizontal; the amplitude of a given frequency component at a given time is represented by a grayscale value between white and black (Figure B.4).[1] Spectrograms are produced by a procedure known as the *short-time Fourier transform (STFT)*.

[1]There are other ways of representing amplitude, such as by color, or by using contour lines, but grayscale spectrograms are most widely used by biologists.

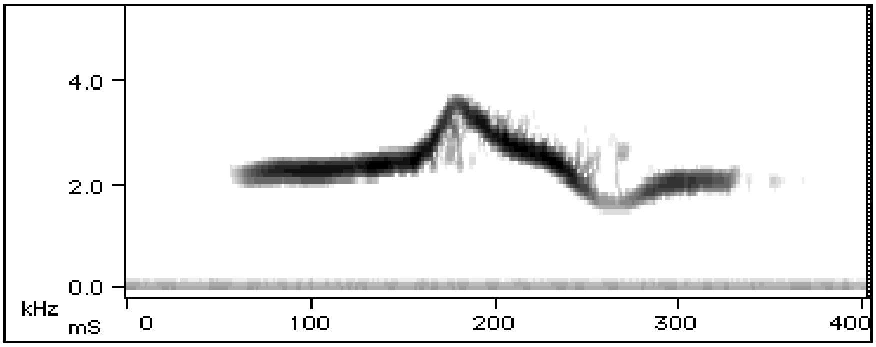


**Figure B.4.** Sound spectrogram of one syllable from song of a rose-breasted grosbeak, digitized at 22.3 kHz.

There are two convenient ways to describe the operation of the STFT. One approach is to think of the STFT as dividing the entire signal into successive short time intervals or *frames* (which may overlap each other in time). Each frame is used as the input to a DFT, generating a series of spectra (one for each frame) that approximate the "instantaneous" spectrum of the signal at successive moments in time. To display a spectrogram, the spectra of successive frames are plotted side by side with frequency running vertically and amplitude represented by grayscale values (Figure B.5a). We call this the "spectral slice" model of STFT analysis. A given STFT can be characterized by its *frame length*, usually expressed as the number of digitized amplitude samples that are processed to create each individual spectrum.

An alternative description considers the STFT as equivalent in function to a bank of bandpass filters, each centered at a slightly different analysis frequency. The output amplitude of each filter is proportional to the amplitude of the signal in a discrete frequency band or *bin*, centered on the analysis frequency of the filter. To display a spectrogram, the time-varying output amplitudes of filters at successive analysis frequencies are plotted above each other, with amplitude again represented by grayscale values (Figure B.5b). We call this the "filterbank" model of STFT analysis. A given STFT can be characterized by its *bandwidth*, the range of input frequencies around the central analysis frequency that are passed by each filter. All of the filters of a single STFT have the same bandwidth, irrespective of analysis frequency.[1]

These two descriptions of STFT analysis are related in specific ways that are discussed further below. Depending on the context, one or the other of these models may be a more convenient way to think about the STFT. Canary's controls and measurement panels are designed to facilitate considering spectrograms and spectra from either perspective. In the remainder of this appendix, we will refer to both models in discussing how Canary generates spectrograms and spectra.

[1]There are other time-frequency representations  (for example, the wavelet transform) that employ different filter bandwidths at different center frequencies.

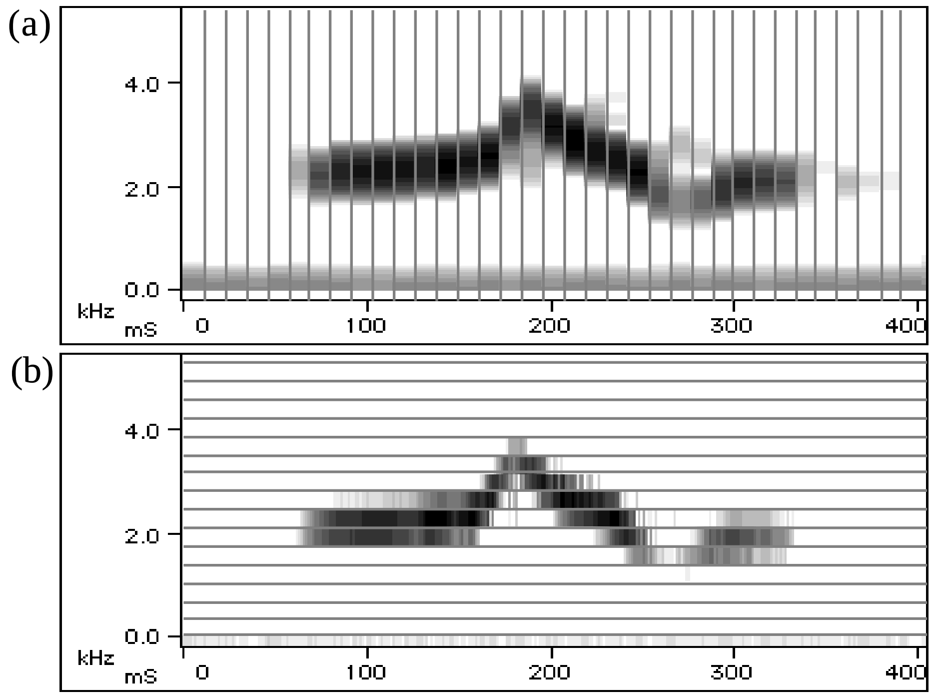


**Figure B.5.** Two ways of considering a sound spectrogram. Both spectrograms are of the same signal shown in Figure B.4, but with different horizontal and vertical resolution. **(a)** Each vertical bar represents the spectrum of a single short time interval or frame, and approximates the “instantaneous” spectrum at a point midway through the frame. **(b)** Each horizontal bar represents the amplitude of the time-varying output of one bandpass filter.

## Frame length, filter bandwidth, and the time-frequency uncertainty principle

The frame length of a STFT determines the *time analysis resolution ($\Delta t$)* of the spectrogram. Changes in the signal that occur within one frame-length of each other (e.g., the end of one sound and the beginning of another, or changes in frequency) cannot be resolved as separate events. Thus, shorter frame lengths allow better time analysis resolution.

Similarly, the bandwidth of a STFT determines the *frequency analysis resolution ($\Delta f$)* of the spectrogram: frequency components that differ by less than one filter-bandwidth cannot be distinguished from each other in the output of the filterbank. Thus a STFT with a relatively wide filter bandwidth will have poorer frequency analysis resolution than one with a narrower bandwidth.

Ideally we might like to have very fine time *and* frequency analysis resolution in a spectrogram. These two demands are intrinsically incompatible, however: the frame length and filter bandwidth of a STFT are inversely proportional to each other, and cannot be varied independently. Although a short frame length yields a spectrogram with finer time analysis resolution, it also results in wide bandwidth filters and correspondingly poor frequency analysis resolution. Thus a tradeoff exists between how precisely a spectrogram can specify the spectral (frequency) composition of a signal and how precisely it can specify the time at which the signal exhibited that particular spectrum.[1]

The relationship between frame length and filter bandwidth applies to spectra as well as spectrograms. The spectrum of a single frame at a particular point in time can be thought of as a cross-section or vertical slice through the output of a filterbank. The bandwidth of the filters in the bank is determined by the length of the frame.

Figure B.6 illustrates the relationship between frame length and filter bandwidth. The two spectra, of a 1000 Hz pure tone digitized at 22.3 kHz, were made with different frame lengths and thus different bandwidths. Spectrum (a), with a frame length of 1024 points (46.0 mS)[2], shows a fairly sharp peak at 1000 Hz because of its relatively narrow bandwidth filter; spectrum (b), with a frame length of 256 points (11.5 mS), corresponding to a wider bandwidth filter, has much poorer frequency resolution.

---

[1]The spectrogram is one of many different types of time-frequency representations (TFRs) that show how the frequency spectrum of a signal changes over time. The TFR with the highest resolution is the Wigner distribution. Spectrograms and all other (reasonable) TFR’s are smoothed (blurred) versions of the Wigner distribution. The smoothing of the spectrogram is controlled by the length and shape of the spectrogram’s windowing function. The uncertainty principle gives a lower bound on the amount of blurring that takes place when passing from the Wigner distibution to the spectrogram. Although it might be tempting to use the Wigner distribution without smoothing, there are practical disadvantages to this. See the recent book by Cohen for further discussion.

[2]The frame length of a STFT can be expressed either in “points” (i.e., the number of digital samples in the frame), or in seconds. The time between successive points is equal to the inverse of the sampling rate (1/fs), so the frame length in seconds equals the number of points in the frame divided by the sampling frequency.

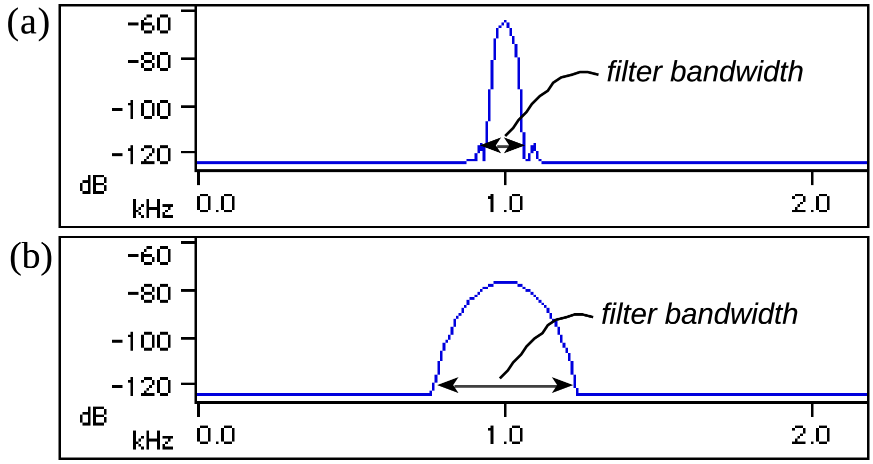


**Figure B.6.** Relationship between frame length and filter bandwidth.[1] Each spectrum is of a single frame of a 1000 Hz tone, digitized at 22.3 kHz. In both spectra, FFT size = 2048 points, window function = Blackman, clipping level = -130 dB.
**(a)** Frame length = 1024 points = 46.0 mS; filter bandwidth = 135 Hz.
**(b)** Frame length = 256 points = 11.5 mS; filter bandwidth = 540 Hz.

## Making spectrograms

A spectrogram produced by Canary is a two-dimensional grid of discrete data points on a plane in which the axes are time and frequency. At each gridpoint, an estimate of the amplitude of sound energy is plotted as a grayscale value. In a spectrogram displayed in “boxy” mode, the gridpoints are at the corners of the boxes (Figure B.7). The grayscale value in each box reflects the amplitude at its upper left corner.

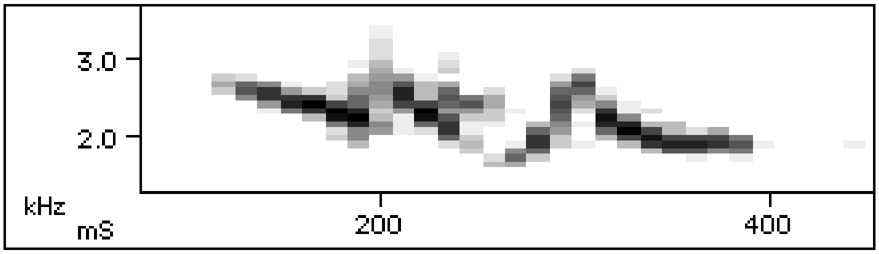


**Figure B.7.** Low-resolution boxy spectrogram of part of a song of an American robin, digitized at 22.3 kHz. The grayscale value in each box represents an estimate of the energy amplitude at the time-frequency point that is at the upper left corner of the box. Filter bandwidth = 353 Hz, frame length = 256 points (= 11.5 mS). Grid resolution = 11.5 mS x 86.9 Hz.

[1]Filter bandwidths are often measured as the width of the band between the frequencies where the amplitude of the filter’s output is 3 dB below the peak output frequency. The arrows indicating the filter bandwidths in this figure are placed at a lower amplitude for clarity of illustration.

Canary lets you independently specify the spacing between gridpoints in the horizontal and vertical directions, and thus the width and height of the boxes in a boxy spectrogram (Figure B.8). These spacing values are called, respectively, the *time grid resolution* and *frequency grid resolution* of the spectrogram. Canary's Spectrogram Options dialog box lets you specify time and frequency grid resolution directly, or indirectly by specifying the amount of overlap between successive frames, and the FFT size, respectively. The relationships between time grid resolution and frame overlap, and between frequency grid resolution and FFT size are discussed below. See Chapter 3 for a detailed discussion of how to control these parameters in Canary.

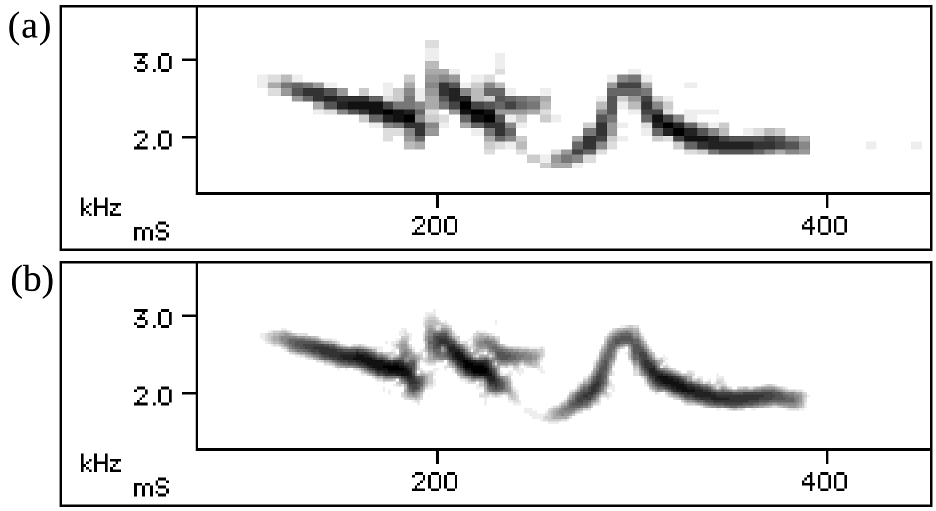


**Figure B.8.** Boxy spectrograms of the same signal as in Figure B.7, with the same analysis resolution (filter bandwidth and frame length), but different grid resolutions. **(a)** Grid resolution = 5.8 mS x 86.9 Hz. **(b)** Grid resolution = 1.4 mS x 10.9 Hz.

Grid resolution should not be confused with *analysis resolution*. Analysis resolution for time and frequency are determined by the frame length and filter bandwidth of a STFT, respectively. Analysis resolution describes the amount of smearing or blurring of temporal and frequency structure at each point on the grid, irrespective of the spacing between these points. The following sections seek to clarify the concepts of analysis resolution and grid resolution by showing examples of spectrograms that illustrate the difference between the two.

### *Analysis resolution and the time-frequency uncertainty principle*

At each point on the spectrogram grid, the tradeoff between time and frequency analysis resolution is determined by the relationship between frame length and filter bandwidth, as discussed above. According to the uncertainty principle, a spectrogram can never have extremely fine analysis resolution in both the frequency and time dimensions.

For example, Figure B.9 shows two spectrograms of the same signal that differ only in frame length and filter bandwidth. The signal consists of two repetitions of a sequence of four tones. Each tone is 20 mS long and has a frequency of 1, 2, 3, or 4 kHz. In spectrogram (a), with a frame length of 64 points (= 2.9 mS; filter bandwidth = 1412 Hz), the beginning and end of each tone can be clearly distinguished and are well-aligned with the corresponding features of the waveform. However, the frequency analysis resolution is poor: each tone appears as a bar that is nearly 800 Hz in thickness. In spectrogram (b), the frame length is 512 points, or 23 mS (filter bandwidth = 176 Hz), or about as long as each tone in the signal. Most of the frames therefore span more than one tone, in some cases including a tone and a silent interval, in other cases including two tones and an interval. The result is poor time resolution: the beginning and end of the bars representing the tones are fuzzy and poorly aligned with the actual features of the waveform (compare, for example, the beginning time of the first pulse in the waveform with the corresponding bar in the spectrogram).

(a)

(b)

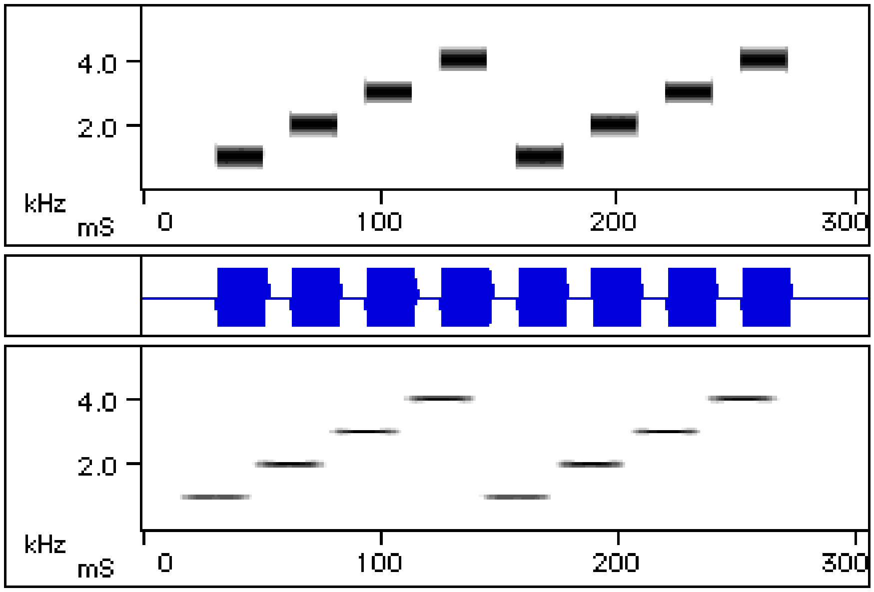


**Figure B.9.** Effect of frame length and filter bandwidth on time and frequency resolution. The signal consists of two repetitions of a sequence of four tones with frequencies of 1, 2, 3, and 4 kHz. Each tone is 20 mS in duration. The interval between tones is 10 mS. Both spectrograms have the same clipping level, time grid resolution = 1.4 mS, frequency grid resolution = 43.5 Hz (FFT size = 512 points), and window function = Hamming.

**(a)** Wide-band spectrogram: frame length = 64 points ( = 2.9 mS), filter bandwidth = 1412 Hz.

**(b)** Narrow-band spectrogram: frame length = 512 points ( = 23.0 mS), filter bandwidth = 176 Hz.

The waveform between the spectrograms shows the timing of the pulses.

What is the “best” analysis resolution to choose? The answer depends on how rapidly the signal’s frequency spectrum changes, and on what type of information is most important to show in the spectrogram, given your particular application. For many applications, it may be best to start with an intermediate frame length (e.g., 256 or 512 points) and filter bandwidth. If you need to observe very short events or rapid changes in the signal, a shorter frame may be better[1]; if precise frequency representation is more important, a longer frame may be better. If you need better time *and* frequency resolution than you can achieve in one spectrogram, you may need to make two spectrograms: a wide-band spectrogram with a small frame for making precise time

[1]If the features that you’re interested in are distinguishable in the waveform (e.g., the beginning or end of a sound, or some other rapid change in amplitude), you’ll achieve the best precision and accuracy by making time measurements on the waveform rather than the spectrogram.

measurements, and a narrow-band spectrogram with a larger frame for precise frequency measurements.

### *Time grid resolution and frame overlap*

Time grid resolution is the time between the beginnings of successive frames. In a boxy spectrogram, this interval is visible as the width of the individual boxes (Figures B.7 and B.8). Successive frames that are analyzed may be overlapping (positive overlap), contiguous (zero overlap), or discontiguous (negative overlap). Overlap between frames is usually expressed as a percentage of the frame length.

Figure B.10 illustrates the different effects of changes to frame length and time grid resolution. Each pulse in the signal is a frequency-modulated tone that sweeps upward in frequency over a range of 380 Hz centered at 1, 2, 3, or 4 kHz. Spectrograms (a) and (b) both have a frame length of 512 points (= 23.0 mS; filter bandwidth = 176 Hz). (a) was made with 0% overlap (grid resolution = 23.0 mS), whereas (b) was made with an overlap of 93.8% (grid resolution = 1.4 mS). In the low-resolution spectrogram (a), each box is as wide as a frame, which in turn is about the same size as each pulse in the signal. The result is a spectrogram that gives an extremely misleading picture of the signal. Spectrogram (b), with a greater frame overlap, is much “smoother” than the one with less overlap, and it reveals the frequency modulation of each pulse in the signal. It still provides poor time analysis resolution, however, because of its large frame length— notice the fuzzy beginning and end to each bar on the spectrogram and the poor alignment with the corresponding features in the waveform. Comparison of the spectrograms in Figure B.10 demonstrates that improved time grid resolution is not a substitute for finer time analysis resolution, which can be obtained only by using a shorter frame (Figure B.10c).

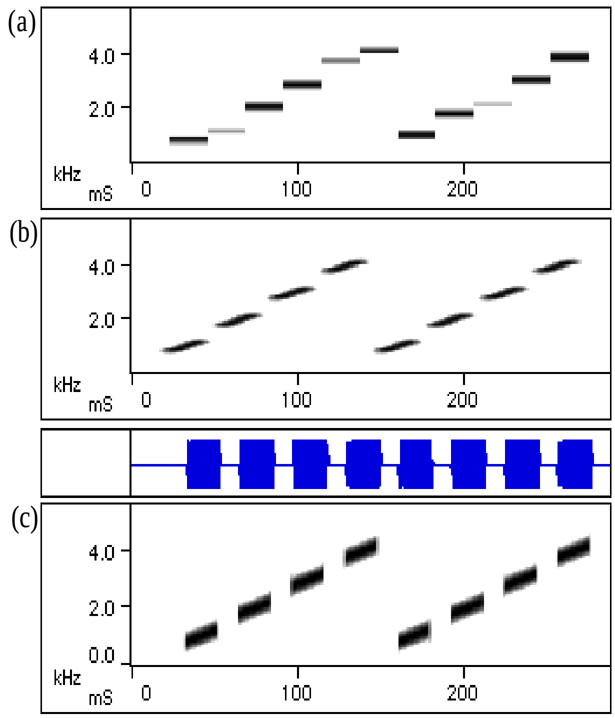


**Figure B.10.** Different effects on spectrograms of changing frame length (time analysis resolution) and time grid resolution. The signal is two repetitions of a series of four frequency-modulated tones, each 20 mS long, with 10 mS between tones. Each tone sweeps upward in frequency through a range of about 380 Hz centered around 1, 2, 3, or 4 kHz. Spectrograms (a) and (b) have the same frame length, but (b) has finer time grid resolution. (b) and (c) have the same grid resolution, but (c) has a shorter frame length (finer time analysis resolution). In both spectrograms, filter bandwidth = 176 Hz (frame length = 512 points = 23.0 mS), frequency grid resolution = 43.5 Hz (FFT size = 512 points).
**(a)** Frame length = 512 points = 23.0 mS (filter bandwidth = 176 Hz); Time grid resolution = 23.0 mS (overlap = 0%).
**(b)** Frame length = 512 points = 23.0 mS (filter bandwidth = 176 Hz); Time grid resolution = 1.4 mS (overlap = 93.8%).
**(c)** Frame length = 64 points = 2.9 mS (filter bandwidth = 1412 Hz); Time grid resolution = 1.4 mS (overlap = 50%).
The waveform between the spectrograms shows the timing of the pulses.

## *Frequency grid resolution and FFT size*

Frequency grid resolution is the difference (in Hz) between the central analysis frequencies of adjacent filters in the filterbank modeled by a STFT, and thus the size of the frequency bins. In a boxy spectrogram, this spacing is visible as the height of the individual boxes (Figures B.7 and B.8). Frequency grid resolution depends on the sampling rate (which is fixed for a given digitized signal) and a parameter of the FFT algorithm called *FFT size*. The relationship is

*frequency grid resolution = (sampling frequency) / FFT size*

where frequency grid resolution and sampling frequency are measured in Hz, and FFT size is measured in points.[1] Thus a larger FFT size draws the spectrogram on a grid with finer frequency

[1]A *point* is a single digital sample.

resolution (smaller frequency bins, vertically smaller boxes). The number of frequency bins in a spectrogram or spectrum is half the FFT size.[1]

Figure B.11 illustrates the different effects of changes to filter bandwidth and frequency grid resolution. Spectrograms (a) and (b) both have a filter bandwidth of 1412 Hz (frame length = 64 points = 2.9 mS). However, the frequency grid resolution in (a) is 348 Hz, whereas in (b) it is 43.5 Hz. Spectrogram (b), with finer grid resolution, is “smoother” than (a), but it still provides poor frequency analysis resolution because of its wide bandwidth— the bars representing the pulses in the signal are still quite thick in the vertical dimension. Only by using a narrower bandwidth (Figure B.11c) can we get finer analysis resolution.

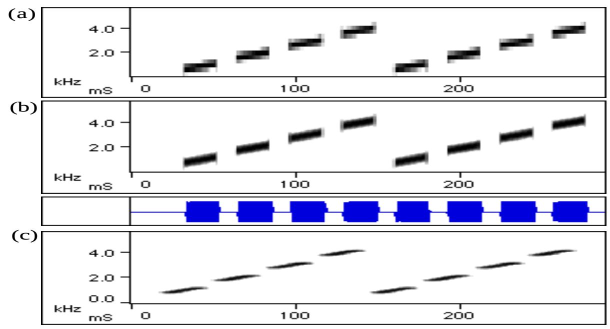


**Figure B.11.** Different effects on spectrograms of changing filter bandwidth (frequency analysis resolution) and frequency grid resolution (FFT size). The signal is the sequence of frequency-modulated tones shown in Figure B.10. Spectrograms (a) and (b) have the same filter bandwidth, but (b) has finer frequency grid resolution. (b) and (c) have the same grid resolution, but (c) has a narrower bandwidth. In both spectrograms, filter bandwidth = 1412 Hz (frame length = 64 points = 2.9 mS), time grid resolution = 1.4 mS.
**(a)** Filter bandwidth = 1412 Hz (frame length = 64 points = 2.9 mS); Frequency grid resolution = 348 Hz (FFT size = 64 points).
**(b)** Filter bandwidth = 1412 Hz (frame length = 64 points = 2.9 mS); Frequency grid resolution = 43.5 Hz (FFT size = 512 points).
**(c)** Filter bandwidth = 176 Hz (frame length = 512 points = 23.0 mS); Frequency grid resolution = 43.5 Hz (FFT size = 512 points).
The waveform between the spectrograms shows the timing of the pulses.

### *Spectral smearing and sidelobes*

The spectrogram (or a single-frame spectral slice) produced by a STFT is “imperfect” in several respects. First, as discussed above, each filter simulated by the STFT has a finite band of frequencies to which it responds; the filter is unable to discriminate different frequencies within this band. According to the uncertainty principle, the filter bandwidth can be reduced— thus improving frequency resolution— only by analyzing a longer frame, which reduces temporal resolution.

Second, the passbands of adjacent filters overlap in frequency, so that some frequencies are passed (though partially attenuated) by more than one filter (Figure B.12). Consequently, when a

[1]Ordinarily, the FFT size of a discrete Fourier transform equals the frame size. Canary allows you to specify a larger FFT to obtain finer grid resolution. This is achieved by zero-padding the selected frame length up to a frame whose length is equal to the FFT size.

spectrum or spectrogram is constructed by plotting the output of all of the filters, a signal consisting of a pure tone becomes "smeared" in frequency (Figure B.12c).

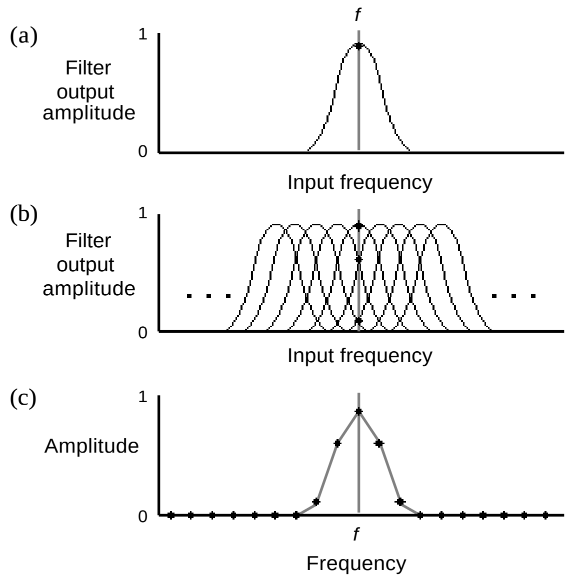


**Figure B.12.** Spectral smearing resulting from overlapping bandpass filters.

**(a)** A single hypothetical bandpass filter centered at frequency *f*. For clarity of illustration, sidelobes to the main passband are not shown (see text and Figure B.13).

**(b)** A set of overlapping filters. Each curve shows the filter function of one filter in a bank simulated by a STFT. Frequency *f* falls within the passbands of the filter centered at *f*, and of two filters on either side.

**(c)** Spectrum of a pure tone signal of frequency *f* produced by the filterbank shown in (b). The spectrum consists of one amplitude value from each filter. Because the filters overlap, the spectrum is smeared, showing energy at frequencies adjacent to *f*.

Third, each filter does not completely block the passage of all frequencies outside of its nominal passband. For each filter there is an infinite series of diminishing "ripples" in the filter's response to frequencies above and below the passband (Figure B.13a). These ripples arise because of the onset and termination of the portion of the signal that appears in a single frame. Since a spectrum of a pure tone made by passing the tone through a set of bandpass filters resembles the frequency response of a single filter (Figure B.12), a STFT spectrum of any signal (even a pure tone) contains frequency ripples. In a logarithmic spectrum, these ripples show up as "sidelobes" (Figure B.13b).

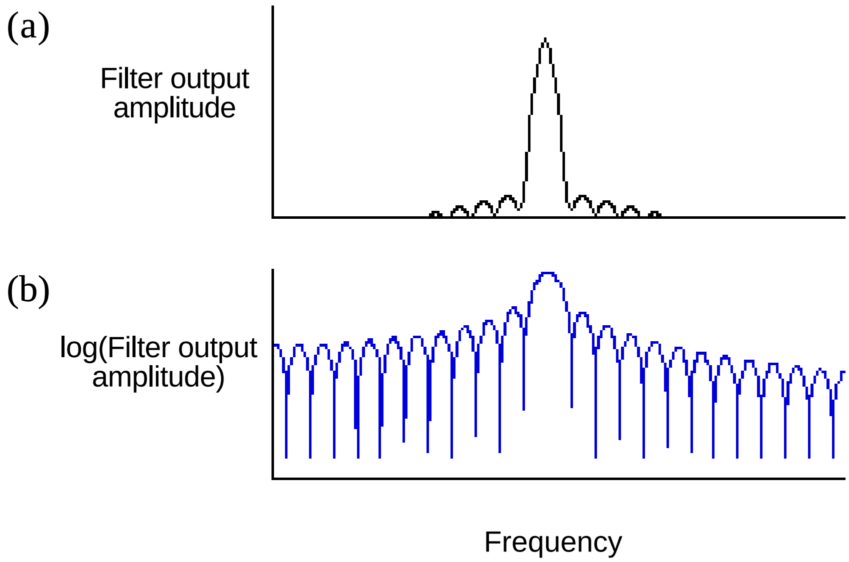


**Figure B.13.** Frequency response of a hypothetical bandpass filter from a set of filters simulated by a short-time Fourier transform, showing ripples or sidelobes above and below the central lobe, or passband. The magnitude of the sidelobes relative to the central lobe can be reduced by use of a window function (see text). Note that a spectrum produced by passing a pure tone through a set of overlapping filters is shaped like the filter frequency response (see Figure B.12). **(a)** Linear plot. **(b)** Logarithmic plot.

### *Window functions*

The magnitude of the sidelobes (relative to the magnitude of the central lobe) in a spectrogram or spectrum of a pure tone is related to how abruptly the signal's amplitude changes at the beginning and end of a frame. A sinusoidal tone that instantly rises to its full amplitude at the beginning of a frame, and then instantly falls to zero at the end, has higher sidelobes than a tone that rises and falls smoothly in amplitude (Figure B.14).

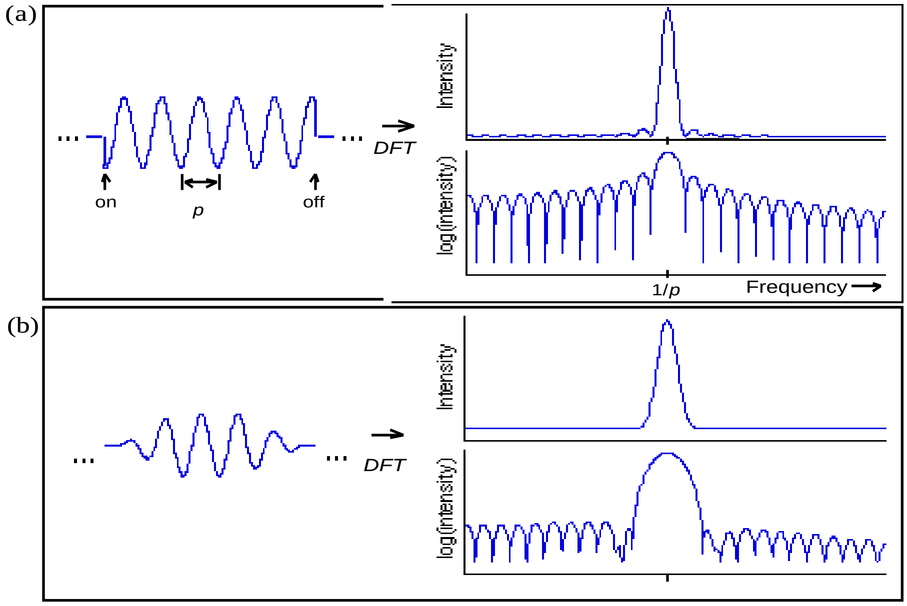


**Figure B.14.** Relationship between abruptness of onset and termination of signal in one frame and spectral sidelobes. Each panel shows a signal on the left, and its spectrum on the right.

**(a)** A single frame of an untapered sinusoidal signal has a spectrum that contains a band of energy around the central frequency, flanked by frequency ripples, as if the signal had been passed through a bank of bandpass filters like the one shown in Figure B.13; the ripples appear as sidelobes in the logarithmic (lower) spectrum.

**(c)** A single frame of a sinusoidal signal multiplied by a “taper” or window function, has smaller sidelobes; the ripples are too small to be visible in the linear (upper) spectrum.

The magnitude of the sidelobes in a spectrum or spectrogram can be reduced by multiplying the frame by a *window function* that tapers the waveform as shown in Figure B.14.[1] Tapering the waveform in the frame is equivalent to changing the shape of the analysis filter (in particular, lowering it sidelobes). Canary supplies five window functions to choose from. Figure B.15 shows spectra of a pure tone made with each of the available window functions. These are also the shapes of the resulting analysis filters. Each window function reduces the height of the highest sidelobe to some particular proportion of the height of the central peak; this reduction in sidelobe magnitude is termed the sidelobe *rejection*, and is expressed in decibels. Given a particular frame length, the choice of window function thus determines the sidelobe rejection, and also the width of the center lobe. The width of the center lobe in the spectrum of a pure tone is the filter bandwidth. For example, the rectangular window function has a narrower filter bandwidth for a given frame length than the Hamming window function, but the Hamming window has lower sidelobes. Figure B.16 shows filter bandwidths corresponding to various frame lengths for each of the five window functions, in order of increasing sidelobe rejection.

[1]Window functions are also sometimes called “tapers”.

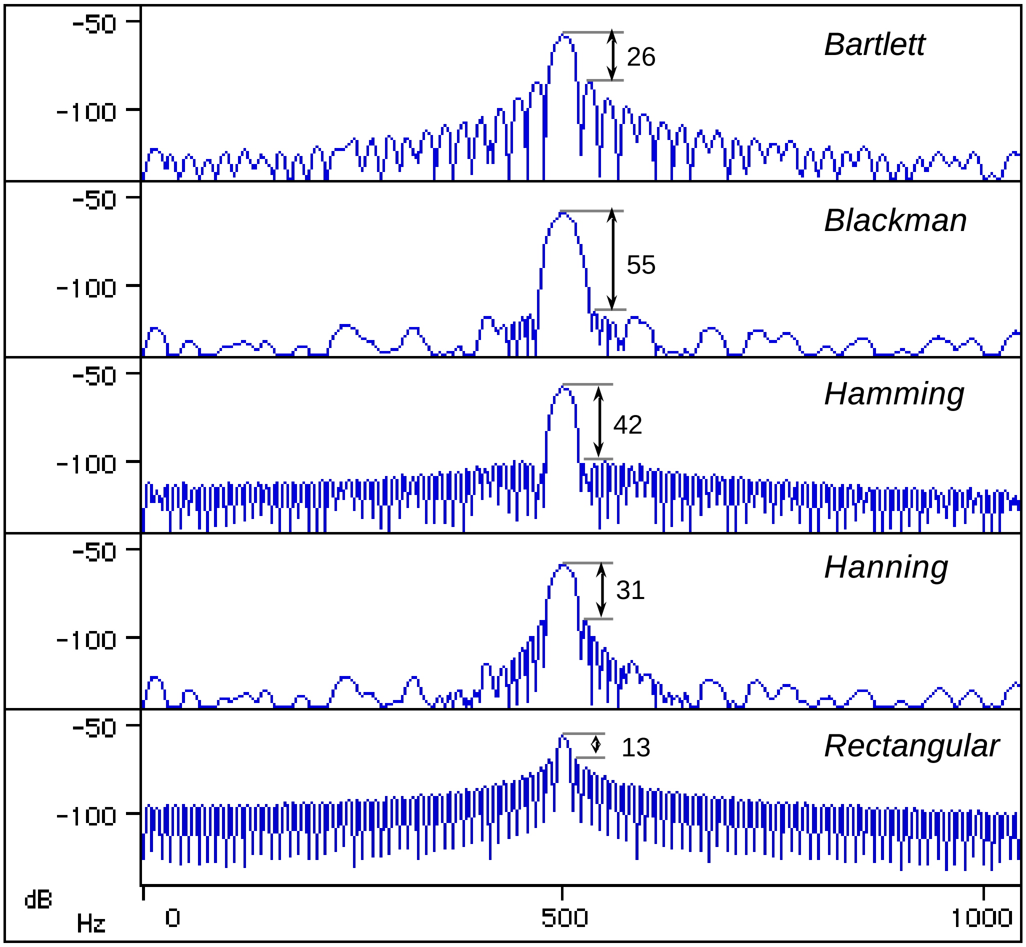


**Figure B.15.** Single-frame spectra of a 500 Hz tone made with five different window functions. Frame length = 2048 points, FFT size = 8192 points for all spectra. The vertical arrows indicate the sidelobe rejection in dB for each window function.

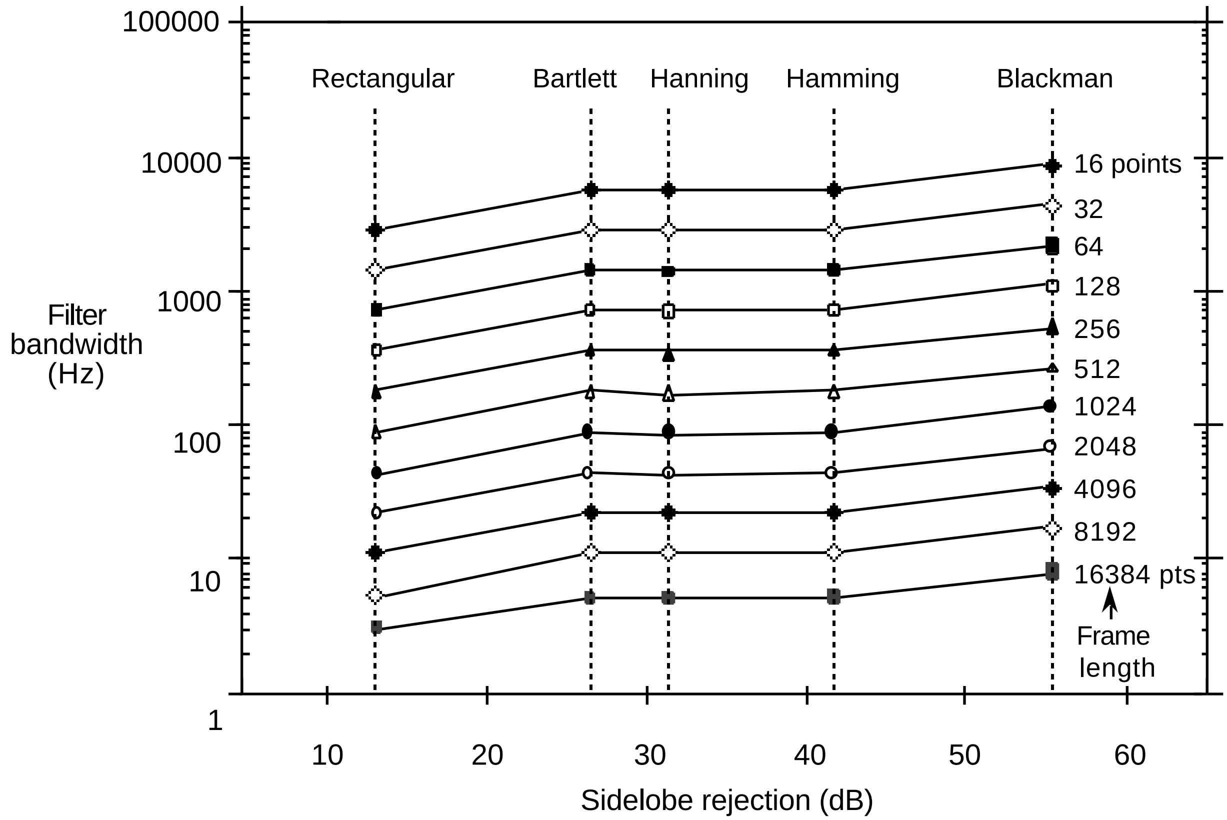


**Figure B.16.** Filter bandwidths corresponding to different frame lengths for each of Canary's five window functions, in order of the windows' increasing sidelobe rejection, given a sample rate of 22.3 kHz.

## For further reading

The books and articles listed below can provide entry at several levels into the vast literature on spectrum analysis and digital signal processing.


Beecher, M. D. 1988. Spectrographic analysis of animal vocalizations: Implications of the "uncertainty principle." *Bioacoustics* 1:(1): 187-207.

Includes a discussion of choosing an "optimum" filter bandwidth for the analysis of frequency-modulated bioacoustic signals.

Cohen, L. 1995. *Time-frequency analysis.* Prentice-Hall, Englewood Cliffs, NJ.

Hlawatsch, F. and G.F. Boudreaux-Bartels. 1992. Linear and quadratic time-requency signal representations. *IEEE Signal Processing Magazine*, 9(2): 21-67.

A technical overview and comparison of the properties of a variety of time-frequency representations (including spectrograms), written for engineers.

Jaffe, D. A. 1987. Spectrum analysis tutorial. Part 1: The Discrete Fourier Transform; Part 2: Properties and applications of the Discrete Fourier Transform. *Computer Music Journal*, 11(3): 9-35.

An excellent introduction to the foundations of digital spectrum analysis. These tutorials assume no mathematics beyond high school algebra, trigonometry, and geometry. More advanced mathematical tools (e.g., vector and complex number manipulations) are developed as needed in these articles.

Marler, P. 1969. Tonal quality of bird sounds. In: *Bird Vocalizations: Their Relation to Current Problems in Biology and Psychology* (ed. R. A. Hinde), pp. 5-18. Cambridge University Press.

Includes an excellent qualitative discussion of how the time and frequency analysis resolution of a spectrum analyzer interact with signal characteristics to affect the "appearance" of a sound either as a spectrogram or as an acoustic sensation.

Oppenheim, A.V. and Schafer, R.W. 1975. *Digital Signal Processing.* Prentice-Hall, Englewood Cliffs, NJ. xiv + 585 p.

A classic reference, written principally for engineers.

Rabiner, L.R. and Gold, B. 1975. *Theory and Application of Digital Signal Processing.* Prentice-Hall, Englewood Cliffs, NJ. xv + 762 p.

Another classic engineering reference.

Yost, W.A. and Nielsen, D.W. 1985. *Fundamentals of Hearing: An Introduction.* 2d ed. Holt, Rinehart and Winston, New York. x + 269 p.

A good general text on human hearing that includes some discussion of the elementary physics of sound and an appendix that introduces basic concepts of Fourier analysis.