Lesson 7: Using the FFT

Introduction

By now you know what spectral analysis is, such as the DFT and its FFT implementation, etc. However, there are some problems that you need to keep in mind when using them. In this lesson, I will discuss some of these problems: *power leakage* caused by sudden changes in the signal, the *tradeoff between time and frequency resolution*, and the function of *windows*. You should be able to use this knowledge to guide what you do to obtain accurate results in estimating spectral information.

sidebar begins

Required Reading:
Ch 10.

sidebar ends

Power Leakage

Power leakage is the phenomena where power at a particular frequency (called the *center frequency* or *central lobe*) "leaks" into neighboring frequencies, which results in the center frequency peak being reduced, the peak width becoming broader, and nearby frequencies (or side lobes') amplitudes increasing. This happens when the signal abruptly changes, for instance suddenly turning on or off. Let's see the case of a truncated signal.

When we compute a signal's spectrum, values of the signal for all time are required. However, in practice, we observe signals for only finite durations. Therefore, the spectrum of a signal can only be approximated from a finite data record. Let's say we have an analog signal, we sample it at a rate F_s, and we limit the duration of the signal to the time interval $T_0 = NT$, where N is the number of samples and $T = 1/F_s$ is the sample interval. We denote the original discrete signal as $\{x_t\}$ and duration limited signal as $\{y_t\}$. This is equivalent to multiplying $\{x_t\}$ by a rectangular window w_t of length N. That is,

$$y_t = x_t w_t \tag{7-1}$$

where

$$w_t = \begin{cases} 1 & 0 \leq t \leq N-1 \\ 0 & otherwise \end{cases} \tag{7-2}$$

The rectangular window w_t sharply chopped the original signal to get the finite signal y_t.

I'll use a sinusoid as an example for the signal x_t,

$$x_t = \cos \omega_0 t \tag{7-3}$$

Using Euler's formula, the signal also can be written as

$$x_t = \frac{1}{2}\left(e^{j\omega_0 t} + e^{-j\omega_0 t}\right) \tag{7-4}$$

The windowed version of this signal is

$$y_t = \frac{1}{2}\left(e^{j\omega_0 t} + e^{-j\omega_0 t}\right) w_t \tag{7-5}$$

Setting $\omega_k = k2\pi/N$ (converting frequency from Hz to radians/sample) in DFT formula and applying it,

$$\begin{aligned} Y_k \equiv Y(\omega_k) &= \sum_{t=0}^{N-1} w_t \frac{1}{2}(e^{j\omega_0 t} + e^{-j\omega_0 t})e^{-jw_k t} \\ &= \frac{1}{2}\sum_{t=0}^{N-1} w_t(e^{-j(\omega_k-\omega_0)t} + e^{-j(\omega_k+\omega_0)t}) \\ &= \frac{1}{2}(W(\omega_k-\omega_0) + W(\omega_k+\omega_0)) \end{aligned} \tag{7-6}$$

where $W(\omega)$ is the DFT of the window w_t. Actually, w_t can be viewed as the long pulse we talked about in the previous lesson (Fourier Analysis). Its DFT is

$$\begin{aligned} W_k \equiv W(\omega_k) &= \sum_{t=0}^{N-1} w_t e^{-j\omega_k t} \\ &= \sum_{t=0}^{N-1} e^{-j\omega_k t} &\text{(7-7)} \\ &= \frac{\sin(\omega_k N/2)}{\sin \omega_k/2} e^{-j\omega_k (N-1)/2} &\text{(7-8)} \end{aligned}$$

Figure 7.1 is a plot of $Y(\omega_k)$ with a window length of $N = 128$, and $\omega_0 = 0.2\pi$. The power of the original signal sequence x_t — concentrated at a single frequency $\omega_0 = 0.2\pi$ — has been spread by the rectangular window into the entire frequency range. That is called *power leakage*. The rectangular window's effect is to cut out a piece of the original "long" digital signal; this action is equivalent to turning the signal suddenly on at $t = 0$ and off at $t = N$. This causes the power leakage.

Let's see what happens if we let the window grow infinity long (or $N \to \infty$), that is

$$\begin{aligned} W_k \equiv W(\omega_k) &= \sum_{t=0}^{\infty} w_t e^{-j\omega_k t} \\ &= \sum_{t=0}^{\infty} e^{-j\omega_k t} \end{aligned} \tag{7-9}$$

This is an infinite geometric series. Its common ratio is $e^{-j\omega_k}$, so

$$\begin{aligned} W_k &= \frac{1}{1-e^{-j\omega_k}} \\ &= \frac{1}{2je^{-j\omega_k/2}\sin(\omega_k/2)} \\ &= \frac{-je^{j\omega_k/2}}{2\sin(\omega_k/2)} \end{aligned} \tag{7-10}$$

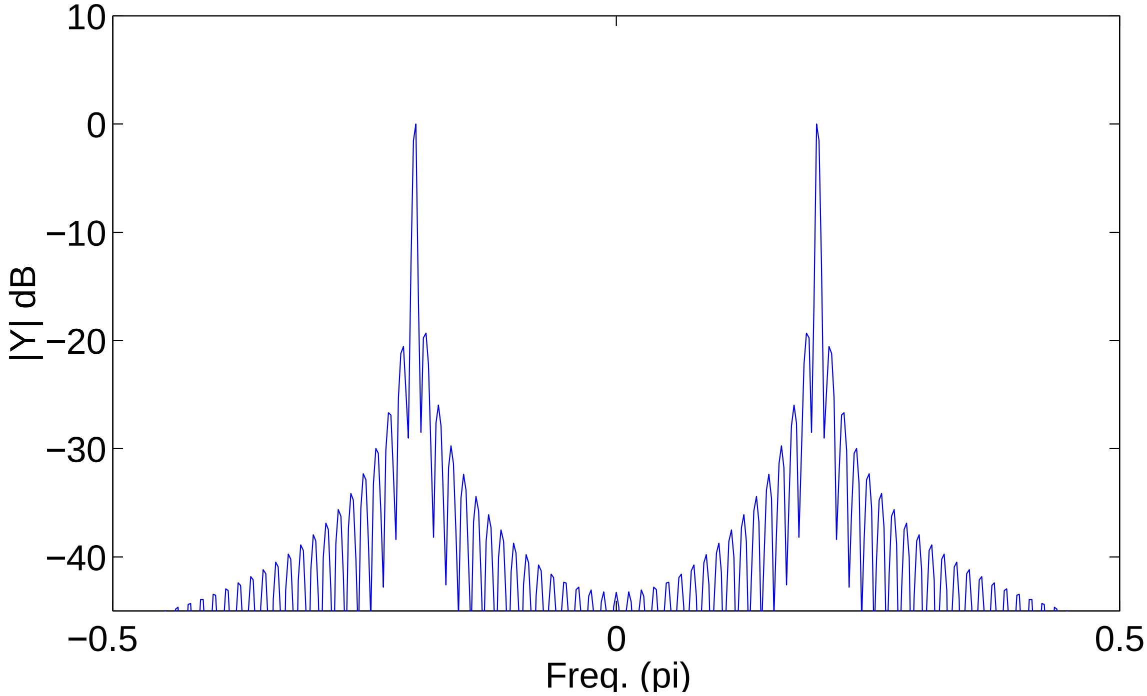

Figure 7.1: Magnitude spectrum of windowed signal $x_t = \cos\omega_0 t$, $\omega_0 = 0.2\pi$ and the rectangular window length is $N = 128$. This Illustrates the occurrence of power leakage.

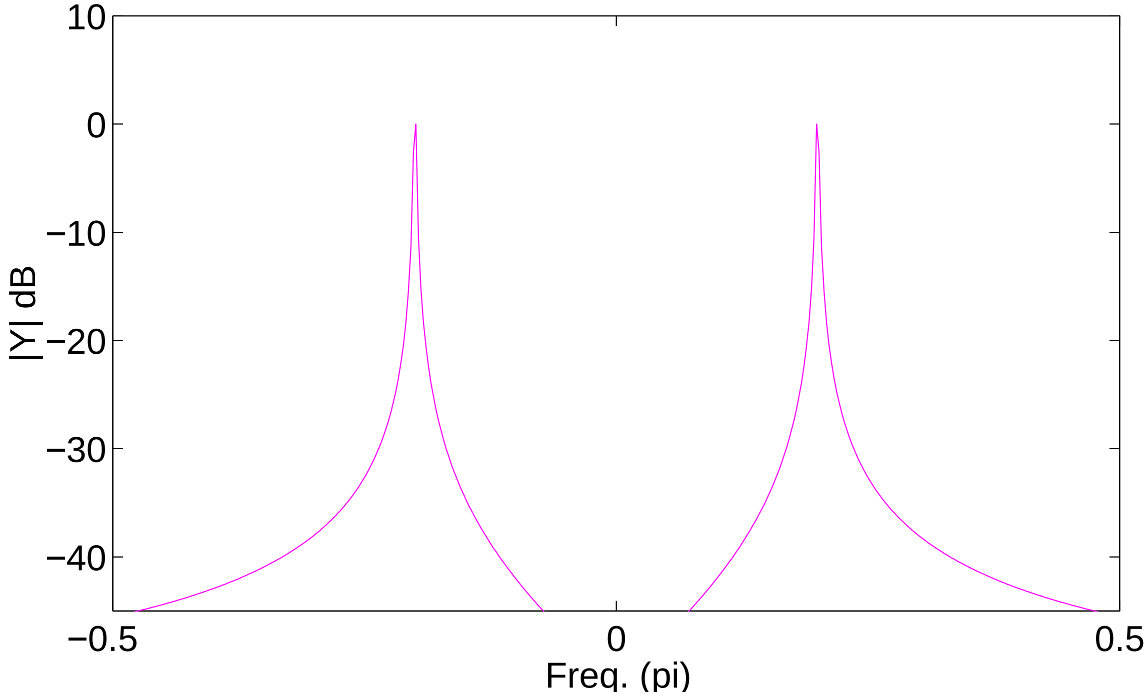

Figure 7.2: Magnitude spectrum of a one-sided signal $x_t = \cos\omega_0 t$ with $\omega_0 = 0.2\pi$.

Substituting this result back into (7-6), we obtain the spectrum of a one-sided signal x_t. This is plotted in figure 7.2.

Comparing (7-8) and (7-10), the magnitude of the former has N zeros equally spaced on the frequency axis, except at the peak frequency ω_0, where there is a $0/0$ situation, which becomes 1. All these zeros contribute to the spectrum's oscillation. The magnitude of the latter one does not have zero.

Self-Test Exercise

1. Fill in the steps leading from (7-7) to (7-8) (*Popup Answer: (7-7) is a geometric series with common ratio $e^{-j\omega_k}$, so*

$$W_k = \frac{1 - e^{-j\omega_k N}}{1 - e^{-j\omega_k}}$$

Using Euler's formula, the numerator becomes

$$1 - e^{-j\omega_k N} = e^{-j\omega_k N/2}(e^{j\omega_k N/2} - e^{-j\omega_k N/2}) = 2je^{-j\omega_k N/2}\sin(\omega_k N/2)$$

Similar, the denominator is $2je^{-j\omega_k/2}\sin(\omega_k/2)$, so

$$W_k = \frac{\sin(\omega_k N/2)}{\sin(\omega_k/2)} e^{-j\omega_k(N-1)/2}$$

which is (7-8))

Tradeoff Between Time and Frequency Resolution

Windowing not only distorts the spectral estimate due to leakage effects, it also reduces spectral resolution. There exists a tradeoff between time and frequency resolution. Wider windows produce finer frequency resolutions, which means you have the ability to distinguish between nearby spectral peaks. However, it yields worse time resolution, meaning that, if you deal with a time-varying signal (a signal with spectrum varying along time), the longer the time window is, the more information is combined within it and the more it misrepresents *when* the spectral components occurred.

To illustrate the concept of frequency resolution, let's use a digital signal consisting of three frequency components,

$$x_t = \cos\omega_0 t + \cos\omega_1 t + \cos\omega_2 t \tag{7-11}$$

When x_t is truncated to N samples in the range $0 \le t \le N-1$, the windowed spectrum is

$$Y(\omega_k) = \frac{1}{2}[W(\omega-\omega_0)+W(\omega-\omega_1)+W(\omega-\omega_2)+W(\omega+\omega_0)+W(\omega+\omega_1)+W(\omega+\omega_2)] \tag{7-12}$$

Examining (7-8) again, the first zero crossing on either side of the center frequency is the ω_k that satisfies

$$\frac{\omega_k N}{2} = \pi \tag{7-13}$$

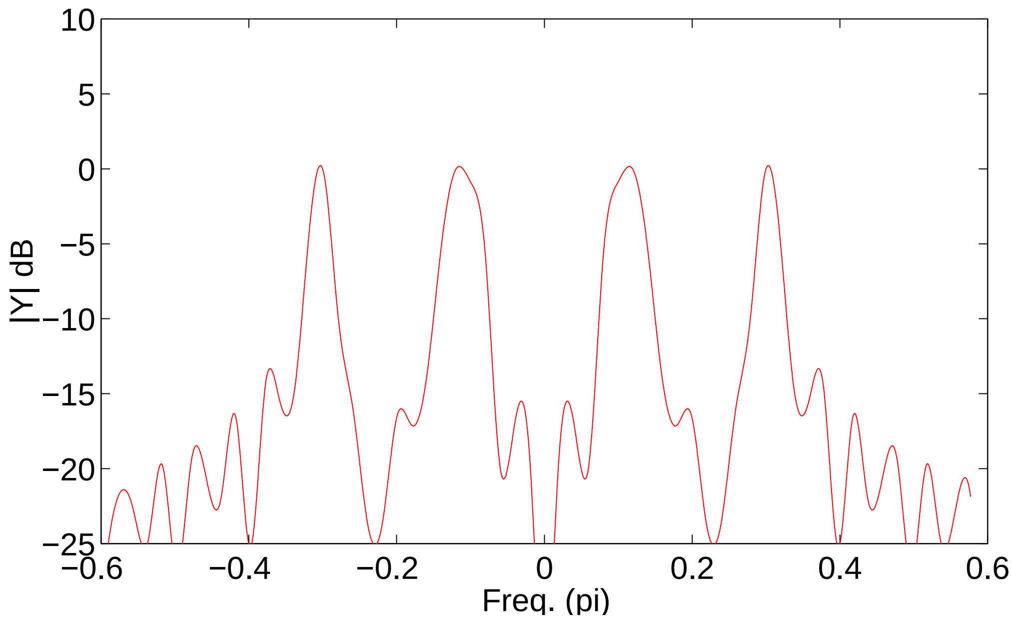

Figure 7.3: Magnitude spectrum of the signal (7-11) made up of three sinusoids. $\omega_0 = 0.1\pi$, $\omega_1 = 0.12\pi$, $\omega_2 = 0.3\pi$, and the rectangular window length was $N = 16$. Because of the small window size, ω_0 and ω_1 are indistinguishable.

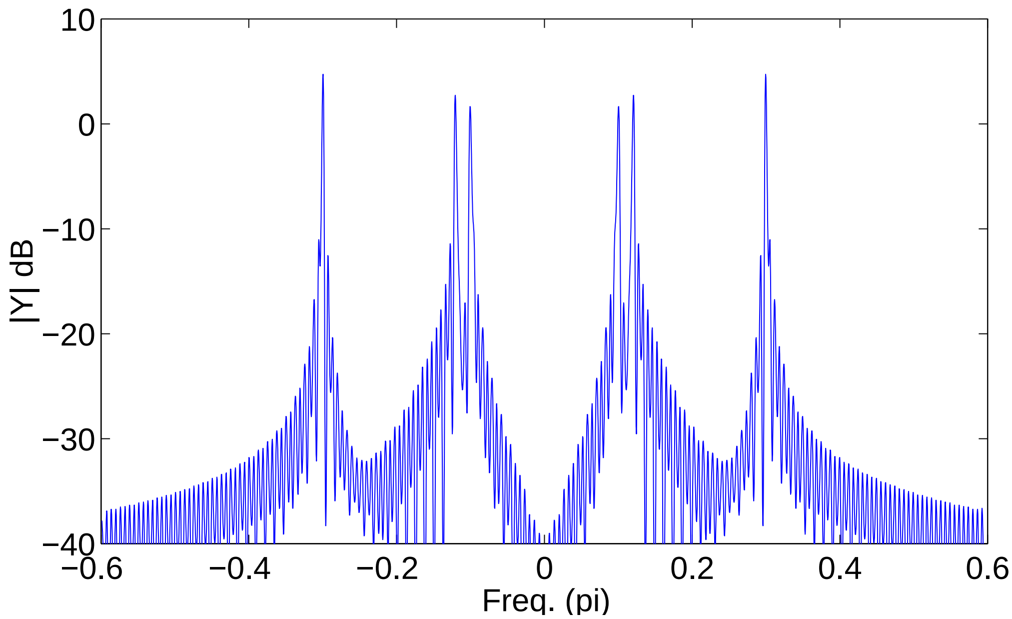

Figure 7.4: The same figure as 7.3, but the rectangular window length $N = 128$ allows separation of the ω_0 and ω_1 frequency components.

which is the first frequency at which $\sin(\omega_k N/2) = 0$, or $\omega_k = 2\pi/N$. The width of the center lobe is twice this, because there is no zero at $\omega_k = 0$. The center lobes of the two window functions $W(\omega - \omega_i)$ and $W(\omega - \omega_j)$ — corresponding to two frequency components ω_i and ω_j ($i \neq j$) — will overlap if $|\omega_i - \omega_j| < 4\pi/N$. As a result, the two spectral lines (peaks) of x_t are not distinguishable. Only if $|\omega_i - \omega_j| \geq 4\pi/N$ can we see two separate lobes in the spectrum $Y(\omega_k)$. Therefore, our ability to resolve spectral lines of different frequencies is limited by the window main lobe width, that is $4\pi/N$. As window length N increases, the spacing between two resolvable frequencies decreases as $1/N$, and so we get better frequency resolution. Figures 7.3 and 7.4 illustrate this conclusion. In these two figures, signal has the three components $\omega_0 = 0.1\pi$, $\omega_1 = 0.12\pi$ and $\omega_2 = 0.3\pi$. The length of the window for 7.3 is $N = 16$, where the two close components ω_0 and ω_1 are indistinguishable. The window length for 7.4 is $N = 128$, and now we can see two peaks around the 0.1π area because the main lobes have become narrow. At the same time, the side lobes becomes lower, which means that less central frequency power leaks into neighboring frequencies. This is why the center frequency peak also becomes higher.

You might then ask if it is always good to extend the window length to get better frequency resolution. The answer is no. Depending on the signal, a longer segment might provide worse spectral information. This is the case when the signal is time-varying, that is the signal's spectrum changes along time (in every instant, the signal has different spectrum or frequency components). When we estimate the spectrum of a windowed signal using the DFT, we treat the segment as though it were a periodic signal (even though it is not), which repeats with period equal to the segment length. You can see that, for a signal that changes with time, this approach mixes together time domain information from the entire segment. The actual spectrum we get from a windowed signal is the *average* spectrum in that window. This means that, for a time-varying signal, a long window will misrepresent what the actual frequencies are (or, rather it will smear the frequencies together). The frequency information is all there, but we don't know *when* they occurred — we only know that it was sometime during the window. This suggests that the best approach would be to decrease the window size to get better time resolution and to get a more accurate time estimate. Unfortunately, as we just learned, short windows result in bad frequency resolution. This is the so-called time/frequency tradeoff. You must do your best to choose a window size that meets your both time and frequency resolution criteria.

A good way to understand this tradeoff is using a *spectrogram*, or short-time Fourier transform. This is a method to deal with time-varying signals to get a compromise result for both the time and frequency domains. A spectrogram is a sequence of FFT results produced by a sliding window. Sliding window processing starts at some time with a fixed window length, computes an FFT, slides the window by a fixed increment, recomputes the FFT, and repeats this sliding and FFT computation for the duration of interest. The result is usually plotted as a surface in a plot with the X axis being time, the Y axis frequency, and surface height being power. Text book chapter 10 figure 7.1 presents a spectrogram where the window length is 1024 and the increment is 200 points (so neighboring windows overlap by 824 samples). The resulting spectrogram is a two-dimensional function with variables time and frequency. The magnitude of the spectrogram gives the spectrum within windows at particular times.

An example of a bird call and its spectrogram is shown in figures 7.5 and 7.6. For each, the top plot is a section of the bird call waveform and the bottom is a spectrogram with two variables time and frequency. Spectrogram color shows the magnitude of the spectrogram with red high and blue

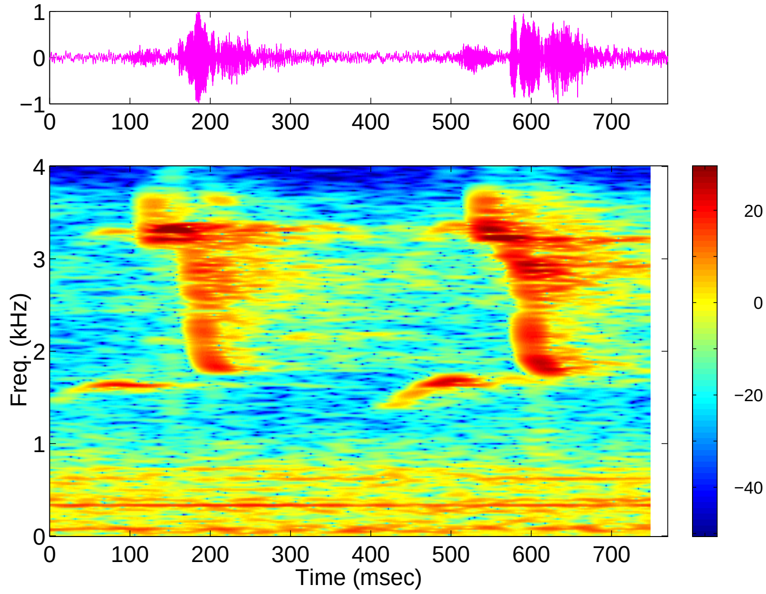

Figure 7.5: A bird call waveform (top) and its spectrogram (bottom). The window size is 512 samples and the overlap is 511 samples (increment is 1). Color indicates magnitude, with color key on the right of the figure.

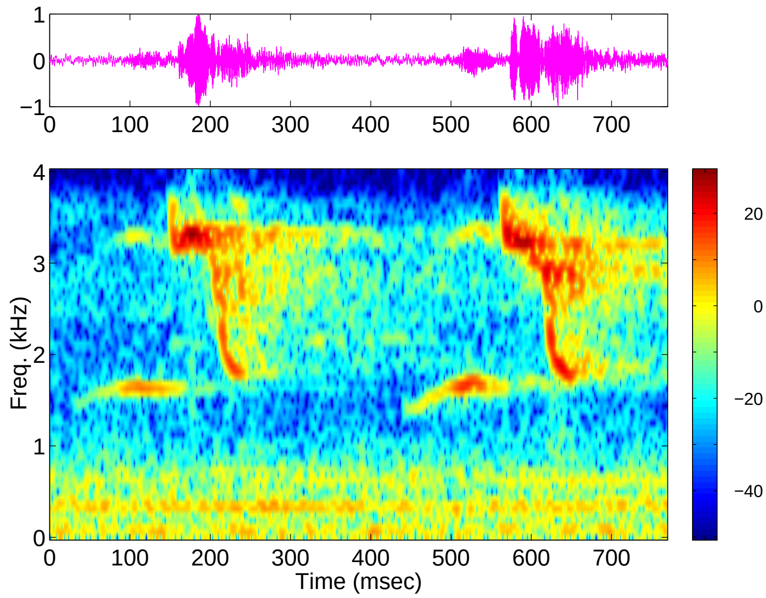

Figure 7.6: The same data as figure 7.5, but the window size is 128 samples and overlap is 127 (increment is 1).

low, as shown color bar on the right.

In figure 7.5, the window length is 512 samples with an overlap of 511. Let's check the details of the spectral components with large magnitudes. Starting at the beginning, there is a component at around 1.5kHz which ends a little before 200 milliseconds. A second component is at around 3.3kHz starting at approximately 50ms continuing until 300ms, but becoming broader at around 100ms. Just before this component's end, it has a broad band frequency range of 1.7kHz to 3.5kHz (time ranging from 150 to 250ms). A similar pattern can be seen in the span of 400–700ms.

Compare this to figure 7.6, where the window length is 128 samples. From the point of view of time, the patterns look narrower, which means they have better temporal resolution (because of the shorter window). We can now see, for example, that there is a complex structure in the temporal evolution of the 1.7–3.5kHz frequency band. However, the patterns are broader and "blockier" along the frequency axis — the frequency resolution got worse (notice the 1.5kHz component in the area of 50–150ms) because of the short window.

Figures 7.7 and 7.8 are another pair of examples. They show similar effects from the tradeoff between time and frequency resolution.

Windowing

In the discussion of power leakage, I mentioned that turning a signal on and off suddenly, abruptly truncating it (which is what a rectangular window does) will cause power leakage from the central lobe to side lobes. To reduce leakage when we select a segment of a signal, instead of using an abrupt truncation — like a rectangular window produces — we could select a data window function w_t that has lower side lobes in the frequency domain. Examples of some popular window functions include Hamming, Hann, Bartlett, Blackman, Gaussian, etc. Each of these has its own characteristics. I will give some examples of what these windows look like and what their effects are.

Hann Window

The text book has some detail about the Hamming window; here I introduce the Hann window, which is very similar. The coefficients of an n-point Hann window are computed from the equation,

$$Hann_t = 0.5\left[1 - \cos\left(\frac{2\pi t}{n-1}\right)\right], \quad t = 0, 1, \ldots, n-1 \tag{7-14}$$

It consists of a half cycle of a cosine, droping to zero at the end-points and with a peak value of one. A plot of a Hann window is shown in figure 7.9

Bartlett Window

The n-point Bartlett window is defined as:

- For n odd

$$Bartlett_t = \begin{cases} \frac{2t}{n-1} & 0 \le t \le \frac{n-1}{2} \\ 2 - \frac{2t}{n-1} & \frac{n-1}{2} \le t \le n-1 \end{cases} \tag{7-15}$$

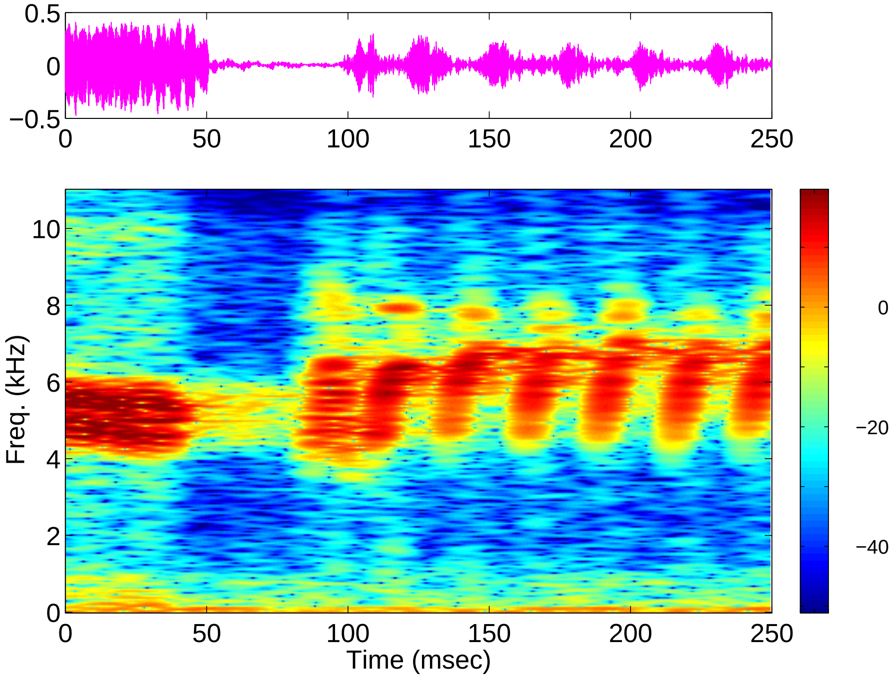

Figure 7.7: Another bird call waveform (top) and its spectrogram (bottom), with window size of 512 samples and overlap of 511 samples.

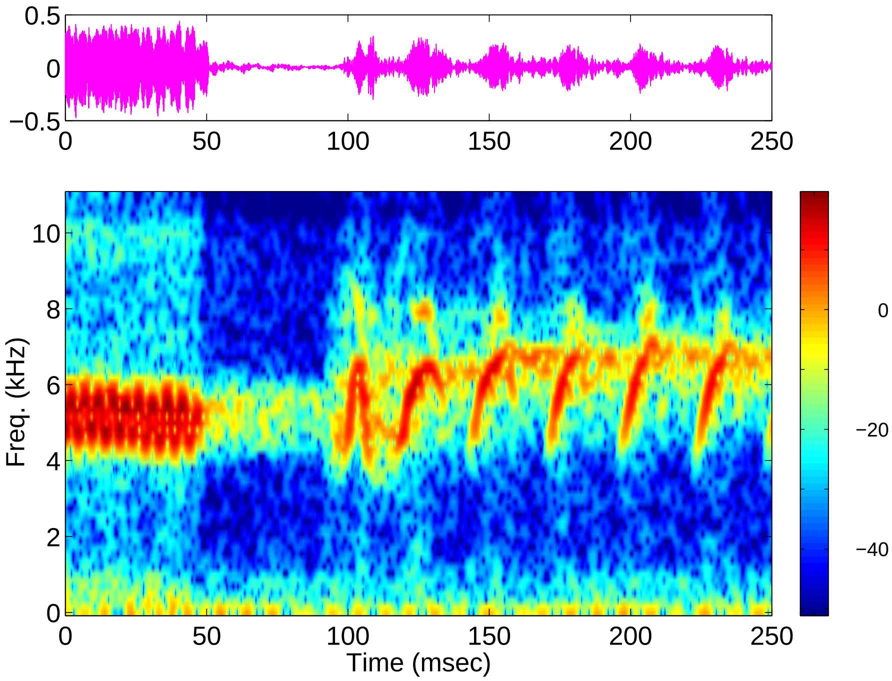

Figure 7.8: Same data as figure 7.7, but with window size of 128 samples and overlap of 127.

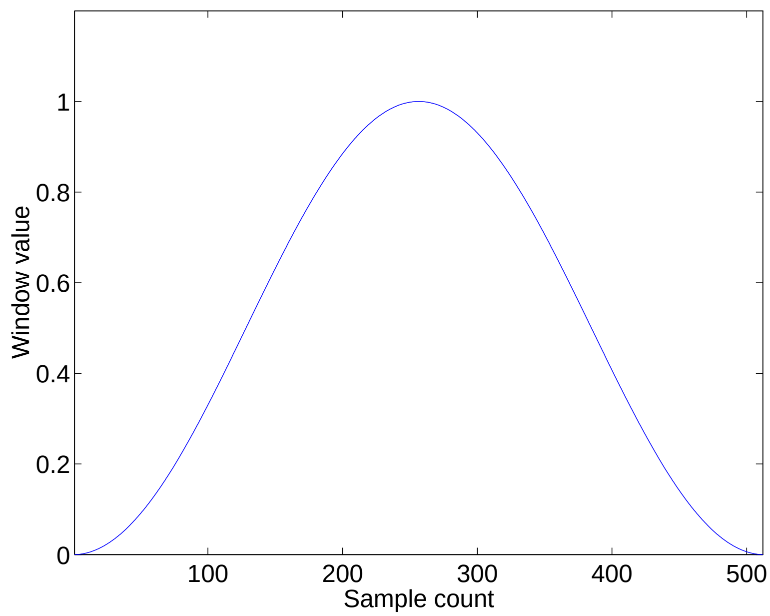

Figure 7.9: A 512-point Hann window in the time domain.

- For n even

$$Bartlett_t = \begin{cases} \frac{2t}{n-1} & 0 \le t \le \frac{n-1}{2} \\ \frac{2(n-t-1)}{n-1} & \frac{n-1}{2} \le t \le n-1 \end{cases} \tag{7-16}$$

This is a triangular window with maximum height of one and zeros at samples 0 and $n-1$. It is plotted in figure 7.10.

Using Window Functions

How do we use window functions? For a general signal sequence x_t, the time domain relationship between the windowed sequence y_t and original sequence is

$$y_t = x_t w_t \tag{7-17}$$

where w_t is a window function in time domain. Previously, I discussed the special case of a rectangular window; now I will consider w_t as a general function. When X, Y,and W are the frequency contents of x_t, y_t, and w_t, the relationship (7-17) in the frequency domain is given by

$$\begin{aligned} Y(\omega) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} X(\theta)W(\omega-\theta)d\theta \\ &= X(\omega) * W(\omega) \end{aligned} \tag{7-18}$$

Equation (7-18) is a convolution between $X(\omega)$ and $W(\omega)$. Recall the definition of convolution we leaned in the lesson, “The Z Transform and Convolution”; instead of an infinite, discrete summation, here it is a finite, continuous integral. It is important to notice that the independent variable is frequency here: this is a frequency domain convolution. The convolution property of the transform means that the time domain relationship is the point to point multiplication in (7-17).

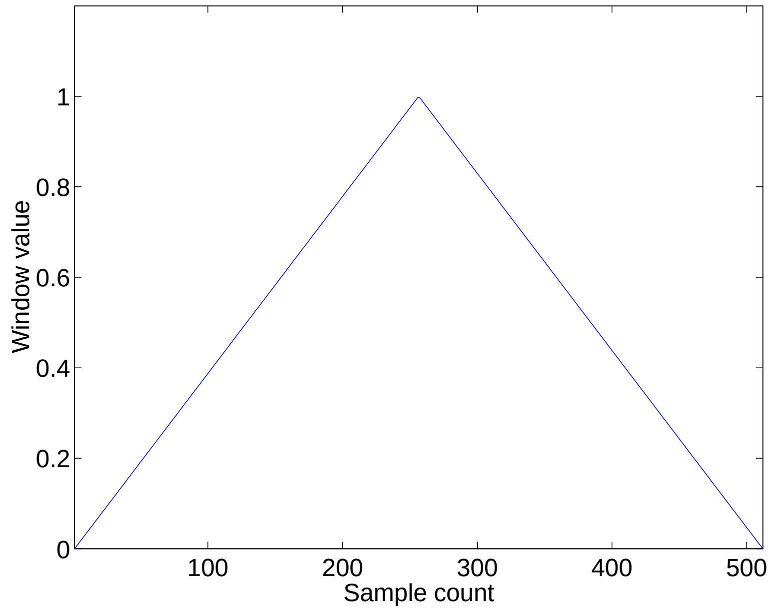

Figure 7.10: The 512-point Bartlett window in the time domain.

Example: Windowed Sinusoid

Consider again a sinusoid signal x_t,

$$x_t = \cos(\omega_0 t); \tag{7-19}$$

Instead of using a rectangular window, let's first use a Hann window and see the results. The windowed signal y_t is

$$y_t = x_t Hann_t = \cos(\omega_0 t) Hann_t; \tag{7-20}$$

where the Hann window is as described in (7-14). y_t is shown in figure 7.11 (top, red curve); its spectrum is the red curve in the bottom plot. Since x_t only has one component $\omega_0 = 0.2\pi$, ideally there should be only one δ function peak at frequency 0.2π. However, as you now know, that is not likely to be the case because of power leakage. The magnitude value is gradually damped down away from the center lobe around ω_0. The leakage is smaller with the Hann window than a rectangular one (yellow curves). On the other hand, the rectangular windowed signal has a narrower peak than the Hann windowed one. So, we can see that the Hann window decreases power leakage by sacrificing peak resolution.

Let's see the result of using a Bartlett window. For the same cosine waveform, the results are shown in figure 7.12. We can reach similar conclusions about the Bartlett window. But, there is a difference between the two windows, as you can see in figure 7.13, which is a comparison between the Hann and Bartlett windows for the cosine wave. Though it appears that the Hann window is superior, there are a number of issues (beyond the scope of this course) that I haven't discussed; the choice of window function involves a number of tradeoffs and depends on the signals being processed and the goal of the processing.

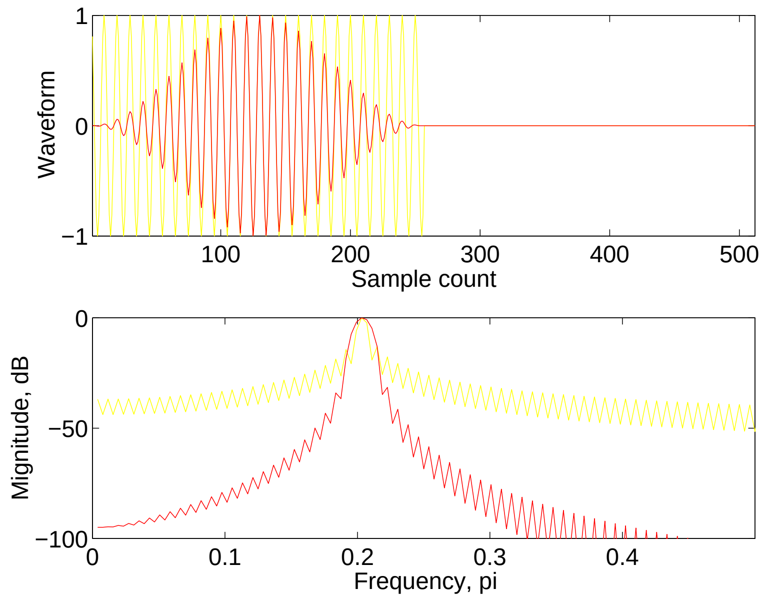

Figure 7.11: Hann windowed cosine waveform ($\omega_0 = 0.2\pi$) and its spectrum, compared to rectangular window. The top is the cosine waveform windowed by rectangular (yellow) and Hann (red) windows, the bottom is their corresponding spectra.

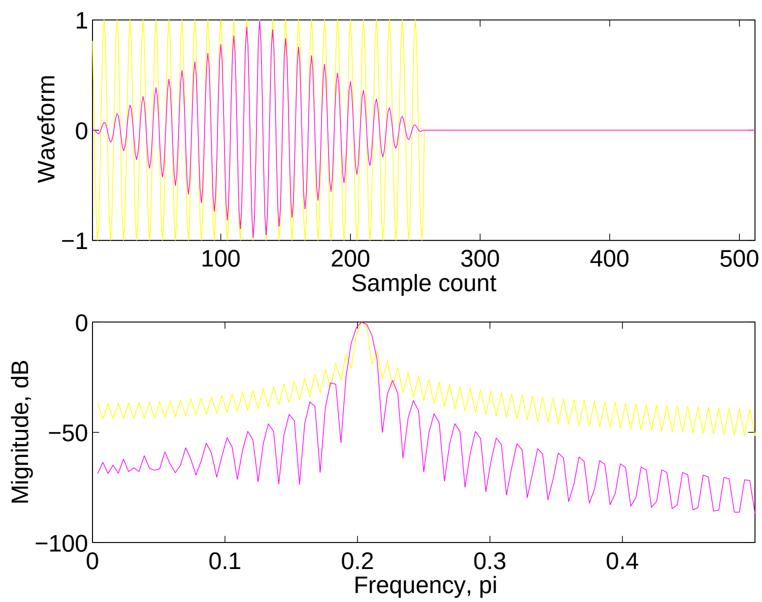

Figure 7.12: Bartlett windowed cosine waveform and its spectrum, compared to rectangular window. The top is the cosine waveform windowed by rectangular (yellow) and Bartlett (magenta), the bottom is their corresponding spectra.

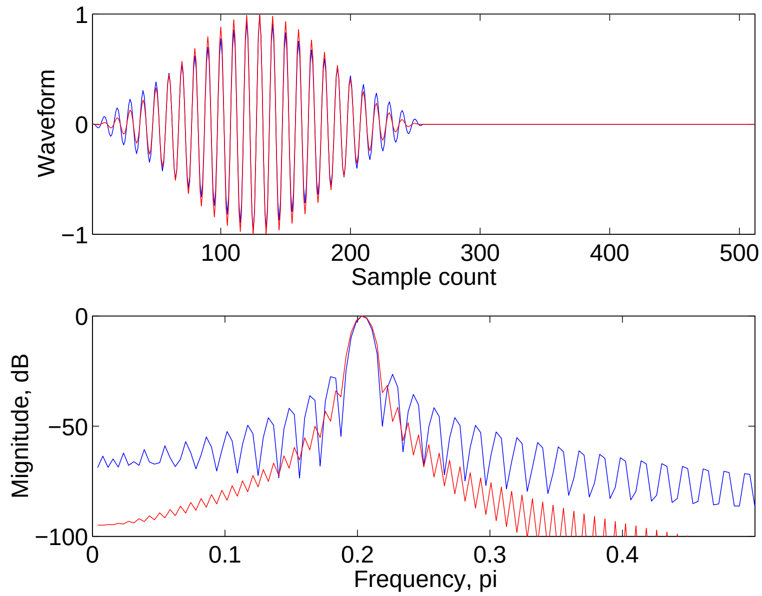

Figure 7.13: Comparison between Hann (red) and Bartlett (blue) windows for a cosine wave.

Example: Windowed Bird Call

Here, I will once again examine the bird call spectrogram I introduced earlier. When I computed that spectrogram, I cheated: I used a Hann window, instead of a rectangular one. Now, instead of using a Hann window in figure 7.6, if I use rectangular one, I get figure 7.14. The increased power leakage into the side lobes of the two main frequency components is readily apparent. This leakage messes up the spectral appearance and affects how accurately we can estimate the shape of the frequency components. That illustrates how important it is to choose the right window.

Figure 7.15 presents a similar result for the bird call originally Hann windowed in figure 7.8.

Self-Test Exercise

1. Plot the Hann and Hamming windows in the time domain and compare their shapes. (*Popup Answer: Use the MATLAB built-in commands* `hamming()` *and* `hanning()`).

Assignment 7

1. Text book chapter 10 problem 10.1.
2. Text book chapter 10 problem 10.5.
3. Text book chapter 10 problem 10.6.
4. Given the bird call data (press button to get the data file amoriole2-1.txt (4000 samples)), and its sampling frequency $F_s = 8kHz$, use MATLAB to compute its spectrogram, plot

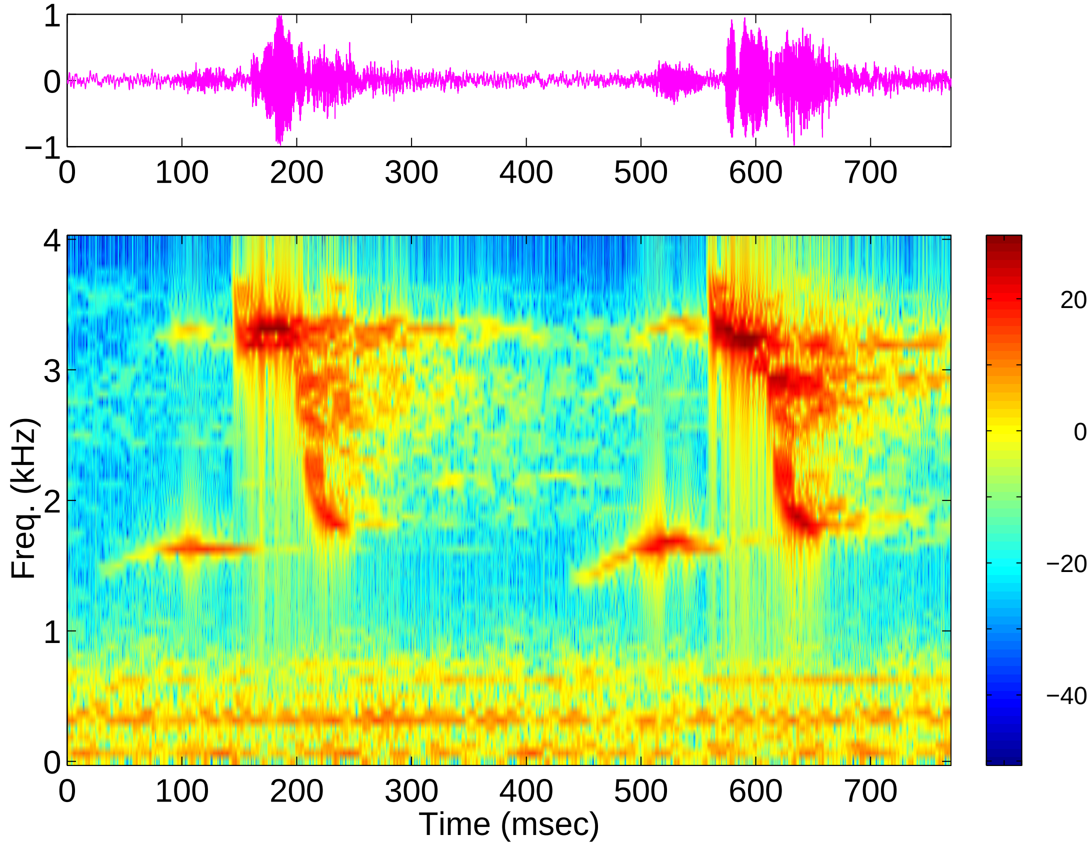

Figure 7.14: Similar spectrogram as figure 7.6, but here a rectangular window is used.

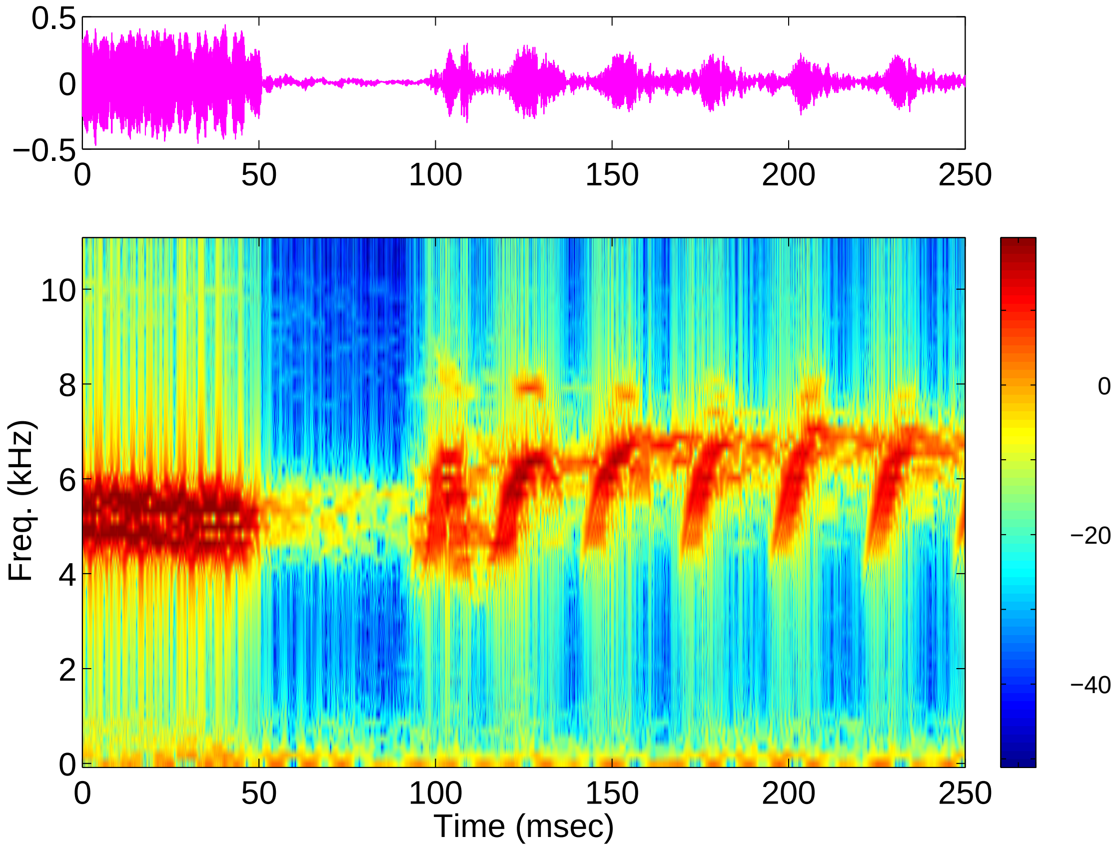

Figure 7.15: Similar spectrogram as figure 7.8, but here a rectangular window is used.

the waveform and its spectrogram (use window sizes of 128 and 512 with Hann (hanning) windows and an increment of 1). Submit the resulting figures and code.