Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science

6.341: Discrete-Time Signal Processing

OpenCourseWare 2006

Lecture 21
Short-Time Fourier Analysis

Reading: Sections 10.1 - 10.5 in Oppenheim, Schafer & Buck (OSB).

In the discussion of spectral analysis of deterministic signals, the signal is usually processed by a system as shown in OSB Figure 10.1, where the DFT is applied, possibly with windowing to satisfy its finite-length requirement. In OSB Figure 10.2 are Fourier transforms associated with this system.

While the output of this system represents the active signal, in more complex signals such as speech or music where the signal properties change over time, it is clearly more meaningful to refer to the changing frequency content over a smaller time interval rather than over an infinite time interval. The next figure is an example of a speech waveform that possesses high spectral locality in time:
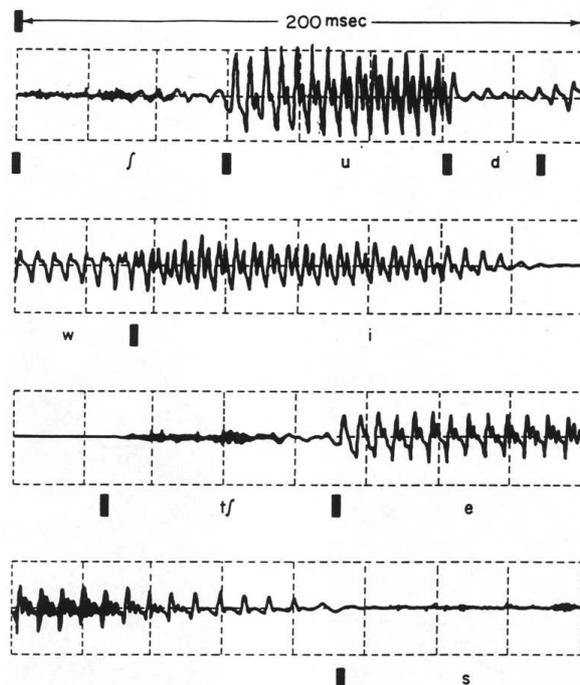
Example of a speech waveform illustrating different classes of sounds.
The utterance is "should we chase ..." (Reproduced from Chapter 3,
A.V. Oppenheim, *Applications of Digital Signal Processing.*)

1

Traditionally, such analysis is carried out by spectrum analyzers which were banks of filters. We will see later in this lecture how a filter bank carries out this task. Next time we will incorporate the FFT into the filter bank implementation.

The Time-Dependent Fourier Transform

Also known as the short-time Fourier transform (STFT), the time-dependent Fourier transform (TDFT) of a signal $x[n]$ is defined as

$$X[n, \lambda) = \sum_{m=-\infty}^{\infty} x[n+m]w[m]e^{-j\lambda m} ,$$

where $w[n]$ is a window sequence. As time progresses, the signal slides past the fixed window. Here we use λ as the frequency variable to distinguish from the conventional DTFT, and the mixed bracket-parenthesis notation $X[n, \lambda)$ to notate that n is a discrete variable while λ is a continuous variable.

As an illustration of how the window $w[n]$ is positioned relative to the shifted signal, OSB Figure 10.11 shows a linear chirp signal with the window superimposed. The signal being shifted through time is
$$x[n] = \cos(\omega_o n^2), \qquad \omega_o = 2\pi \times 7.5 \times 10^{-6} .$$

Its instantaneous frequency is $2\omega_o n$ (hence the name 'linear chirp'). $w[n]$ used is a Hamming window of length 400. Over the finite duration of the window, the frequency content of the signal changes only slightly by $400\omega_o = 2\pi(0.003)$. We can see the linearity of the signal's frequency content from OSB Figure 10.12, a STFT magnitude plot where the vertical axis represents continuous frequency λ and the horizontal axis represents discrete time n. Such a magnitude plot is referred to as a spectrogram.

The Effect of the Window

From the definition of STFT, we see that

$$X[n, \lambda) = \text{DTFT}\{x[n+m]w[m]\} = \frac{1}{2\pi}\text{DTFT}\{x[n+m]\} * \text{DTFT}\{w[m]\}$$

For a fixed n,
$$X[n, \lambda) = \frac{1}{2\pi} \int_0^{2\pi} e^{j\theta n} X(e^{j\theta}) W(e^{j(\lambda-\theta)}) d\theta .$$

This is the convolution between the window $W(e^{j\omega})$ and the shifted input sequence. The main-lobe width of $W(e^{j\omega})$ determines how well the components of the signal frequency spectrum can be resolved and the sidelobe amplitude determines the amount of leakage from these components into their neighbors. Again, trade-offs exist when selecting a proper window for short-time Fourier analysis (STFA). OSB Table 7.1 in Section 7.2 quantitatively illustrates such trade-offs.

Sampling in Frequency and Time

Since λ is continuous, explicitly computing $X[n, \lambda)$ can be done only for a finite set of sample values. Sampling $X[n, \lambda)$ at N equally spaced frequencies gives the DFT of the windowed sequence:

$$X[n, k] = X[n, \lambda)|_{\lambda = \frac{2\pi k}{N}} = N\text{-point DFT}\{x[n + m]w[m]\} \ .$$

If the window is of length L and starts at $m = 0$, i.e., $w[m] = 0$ except for $0 \leq M \leq L - 1$, by properties of the inverse DFT, the windowed sequence can be recovered as long as $N \geq L$:

$$x[n + m]w[m] = \frac{1}{N} \sum_{k=0}^{N-1} X[n, k]e^{j\frac{2\pi}{N}km} \qquad 0 \leq M \leq L - 1 \ .$$

The sequence values $x[n]$ within the interval $[n, n + L - 1]$ can be subsequently computed.

Given the STFT of a sequence at time n_o and exact form of the window, we are able to synthesize $x[n]$ for $n_o \leq n \leq n_o + L - 1$. This suggests that the frequency sampled STFT is redundant, thus can be further downsampled in time. Such a sampled STFT is defined as

$$X[rR, k] = X[rR, \frac{2\pi k}{N}] = \sum_{m=0}^{L-1} x[rR + m]w[m]e^{-j\frac{2\pi}{N}km} \ ,$$

where $-\infty < r < \infty$, $0 \leq k \leq N - 1$, and R is the sampling interval in the time domain. Exact reconstruction of the signal is possible if $R \leq L \leq N$. An alternative notation for $X[rR, k]$ is $X_r[k]$. OSB Figure 10.15 shows the region of support for $X[n, \lambda)$ in the $[n, \lambda)$-plane as lines and for $X_r[k] = X[rR, k]$ as discrete grid points. The amount R by which the window slides from block to block depends on how fast the frequency of the signal changes, and the length L of the window should be chosen according to the desired frequency resolution (long window), or time resolution (short window).

Modulated Filter Banks for Short-Time Fourier Analysis

A common method for short-time Fourier analysis before the development of the FFT is to use a filter bank in which each filter has the same frequency response shape, but different center frequencies. An example is shown below where the prototype lowpass filter is shifted $N - 1$ times to obtain a bank of N filters with complex impulse responses.

Modulated filter banks for short-time Fourier analysis
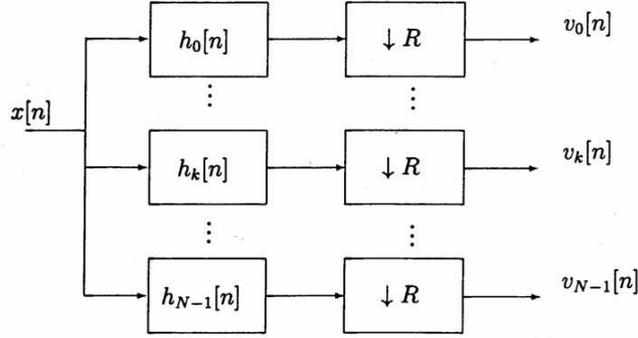
Figure 1

$$h_k[n] = e^{j\omega_k n}h_0[n] \qquad \omega_k = \frac{2\pi k}{N} \text{ for } N \text{ equally spaced filters}$$

$$h_0[n] = \text{lowpass prototype filter}$$
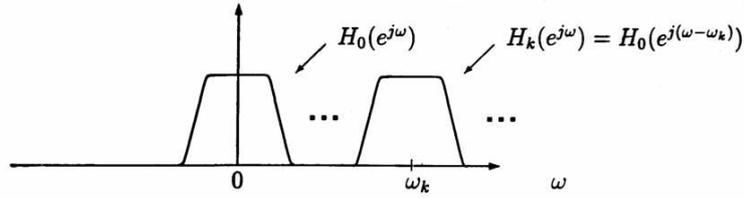
Figure 2

$$\text{For } \omega_k = \frac{2\pi k}{N} \qquad H_k(z) = H_0(e^{-j\frac{2\pi k}{N}}z)$$

In this system, the output $y_k[n]$ from filter $h_k[n]$ to the downsampler satisfies the following equation:

$$y_k[n] = \sum_{r=-\infty}^{\infty} e^{j\omega_k r}h_0[r]x[n-r] = \sum_{r=-\infty}^{\infty} e^{-j\omega_k r}h_0[-r]x[n+r]$$

$$= \text{DTFT}\{x[n+r]h_0[-r]\}|_{\omega=\omega_k} .$$

Let the window sequence be $w[m] = h_0[-m]$, then

$$y_k[n] = X[n, \lambda)|_{\omega=\omega_k} .$$

This is the STFT of $x[n]$ sampled at $\omega = \omega_k$. If $h_0[-r]$ is N-point long, and $\omega_k = \frac{2\pi k}{N}$, then

$$y_k[n] = X[n, k] = N\text{-point DFT}\{x[n+r]h_0[-r]\} .$$

The downsampler with rate R removes time redundancies in $y_k[n]$ to obtain $v_k[n] = X[nR, k]$. Because of the existence of the downsampler, a polyphase implementation for this filter bank is possible, we will explore in detail how this is carried out in the next lecture.

In typical spectral analysis of filter banks, the filter frequency spacing ω_k is chosen to be approximately equal to the filter bandwidth. Also, we typically look at the magnitude of the outputs only, although the filter impulse responses are complex. OSB Figure 10.16 shows the frequency response for sets of bandpass filters obtained from rectangular and Kaiser windows. The Kaiser window has narrower sidelobes but a wider mainlobe. In either case, the filter responses overlap significantly and are of infinitely length. Nonetheless, it can be shown that time aliasing of the reconstructed signal does not occur when inverting the time and frequency sampled TDFT, hence exact reconstruction is possible. Problem 10.40 in OSB illustrates this point.

In addition to using filter banks, the short-time Fourier analysis can also be performed with only one filter $h_0[n]$ by sliding the signal $x[n]$ in frequency, i.e., modulate $x[n]$ with $e^{j\omega_k n}$ for different ω_k values and filter the modulated signal with $h_0[n]$ to obtain $y_k[n]$.

The short-time Fourier transform plays a significant role in discrete-time processing of speech. For a much more detailed discussion of this topic, see OSB Section 10.5 and Chapter 3 Digital processing of Speech, A. V. Oppenheim, *Applications of Digital Signal Processing*, downloadable through the MIT RLE Digital Signal Processing Group web site at

http://www.rle.mit.edu/dspg/documents/DigitalSpeech.pdf