Outline

Today: Chapter 7 pp. 309-320

1. Review from Last Time (Chapter 6 pp. 276-296)

2. Time-Frequency Analysis

3. Short-Time Fourier Transform
   - Fourier Transform View
   - Filtering View

4. Time-Frequency Resolution Tradeoffs
Last Time
Homomorphic Signal Processing

- Short-time Homomorphic Analysis
  - Periodic Sequences
  - Homomorphic Filtering of Speech

- Analysis / Synthesis Structures
  - General Framework
  - Effect of Phase on Synthesis

- Contrasting Linear Prediction with Homomorphic Analysis

- Administrivia
  - Homework #3 posted online due Thursday 03/12
  - Midterm #1 in class Tuesday 03/17: Chapters 1-6
  - One two-sided 8.5x11" sheet of notes allowed
Speech content is (on some time scale) slowly time-varying
- Formant locations change (changing articulatory configuration)
- Pitch period changes (changing state of vocal cords)

A Fourier representation $\hat{f}(\omega)$ of an entire utterance $f(t)$ provides a global view of spectral content, but the timings and locations of individual frequencies are hidden in the complex phase.

This motivates the idea of working with windowed segments of the speech signal, as we have been doing informally thus far.

Time-Frequency analysis is the study of representations that combine the features of $f(t)$ and $\hat{f}(\omega)$ into a single function called a time-frequency representation.
An intuitive solution to the idea of a joint time-frequency analysis is to cut our function $f(t)$ into pieces through the application of a window $w(t)$—and then take the Fourier transform of each piece.

Equivalently, we may define a transformation that correlates $f$ with all possible time-frequency shifts $(\tau, \omega)$ of the window $w(t)$:

$$S_w f(\tau, \omega) = \int_{-\infty}^{\infty} f(t) w^*(t - \tau) e^{-i\omega t} dt.$$ 

In practice, the window $w(t)$ is typically real and symmetric, so that $w^*(t - \tau) = w(t - \tau) = w(\tau - t)$. This convention is adopted implicitly in the text, and for now we will do the same.

We have already seen a related time-frequency representation: the spectrogram $S(\tau, \omega) := |S_w f(\tau, \omega)|^2$. 

Time-Frequency Analysis

Short-Time Fourier Transform
Continuous-time formulation
Different Approaches to Time-Frequency Analysis
We will skin this cat a number of ways

- In this course we will approach time-frequency analysis from two different, intimately connected, points of view

- Signal processing
  - Transforms as filter banks
  - Emphasis on algorithms
  - Based on material in Chapters 7 and 8 of Quatieri

- Harmonic Analysis
  - A rigorous mathematical framework
  - Provides insight into mechanics and limitations of existing methods
  - Based on text and supplementary material

- We will examine the short-time Fourier transform from all these angles, beginning with this lecture...
Discrete-Time Short-Time Fourier Transform

- Recall from Chapter 3 the text’s definition of the discrete-time STFT $X(n, \omega)$ of a sequence $x[n]$:

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

- If the continuous frequency variable $\omega$ is replaced by a discrete variable $k$, the text terms the result a discrete STFT (in analogy to the DFT):

$$X(n, k) = \sum_{m=-\infty}^{\infty} x[m]w[n - m]e^{-j\frac{2\pi}{N} km}$$

- The window $w[n]$ is assumed to be non-zero only in an interval of length $N_w$ and is referred to as the analysis window.

- The sequence $f_n[m] = x[m]w[n - m]$ is called a short-time section of $x[m]$ at time $n$ (note time-reversal).
Computing the Discrete STFT

a. Flip window $w[m] \rightarrow w[-m]$ (note time-reversal)
b. Slide the window sample-by-sample over the sequence
c. Multiply the sequence by the displaced window
d. Take the Fourier transform of the windowed segment
Subsampling in Time and Frequency
Can we subsample and still invert?

- Consider sliding the analysis window along by $L$ samples, rather than one sample, at a time
  - (a) Window positions
  - (b) Time-Frequency subsampling

- What conditions are necessary for reconstruction of the original sequence?
- We will examine this question in depth when we study frames
If we fix the value of $\omega$ at $\omega_0$, and consider the STFT as a function of $n$, then the STFT equation may be written as:

$$X(n, \omega_0) = \sum_{m=-\infty}^{\infty} \left( x[m] e^{-j\omega_0 m} \right) w[n - m]$$

This is a convolution of the sequence $x[n] e^{-j\omega_0 n}$ with $w[n]$.

In this view, the signal $x[n]$ is modulated by $e^{-j\omega_0 n}$ and passed through a filter whose impulse response is the window $w[n]$.

We can view this as (de)modulating of a band of frequencies centered around $\omega_0$ down to baseband, and then filtering by $w[n]$. 
Filtering View of STFT Analysis at Frequency $\omega_0$

(a) Block diagram of complex exponential demodulation followed by a lowpass filter

(b) Equivalent formulation in the frequency domain
A Slightly Modified Filtering View
Bandpass filtering followed by demodulation

Equivalently, we may think of a bandpass filtering operation followed by demodulation:

(a) Block diagram of bandpass filtering followed by a complex exponential demodulation

(b) Equivalent formulation in the frequency domain

This point of view lends itself well to reconstruction...
Analysis/Synthesis with Discrete STFT

(a) The discrete STFT (analysis) as the output of a bank of bandpass filters

(b) Filter bank summation procedure for synthesis from the discrete STFT
Consider a Gaussian window of the form \( w[n] = e^{-\frac{1}{2}(n-n_0)^2} \)

As we have seen, the discrete STFT having DFT length \( N \) can be written as a bank of filters with impulse responses given by

\[
h_k[n] = e^{-\frac{1}{2}(n-n_0)^2} e^{j \frac{2\pi}{N} kn}
\]

Each impulse response has a modulated Gaussian envelope

If we were to assume a sampling rate of 10000 samples/s and selected \( N = 50 \), then the bandpass filters are spaced by 200Hz as follows:
STFT Example 2/2

- Gaussian window $w[n]$ (a) (also output of $k$th bandpass filter with $k = 0$)
- Output of $k$th bandpass filter with $k = 5$ (b), 10 (c) and 20 (d)
- Notice the frequency of the output increases with increasing $k$, while the Gaussian envelope remains intact
The Uncertainty Principle
What is so special about Gaussian windows?

- Recall from Chapter 2 the **uncertainty principle** for Fourier transforms stating that the product of signal duration and bandwidth is bounded from below by a fixed limit.
- Moreover, if a signal has **compact support** in time, then it cannot be compactly supported in frequency (and vice-versa).
- This implies that we cannot simultaneously achieve arbitrarily good resolution in both time and frequency.
- A **Gaussian window** can be shown to achieve the lower bound on joint time-frequency concentration set by the uncertainty principle.
- We will come to study this in more depth in the sequel...
Effect of the Length of Analysis Window

Short-time analysis of a linear chirp

- Frequency-modulated sinusoid (25ms) with linearly decreasing frequency from 1250Hz to 625Hz
- Transforms for different window lengths: 5ms (b, solid), 10ms (c, dashed) and 20ms (d, dotted)
- Short window gives good temporal but poor frequency resolution
- Long window gives poor temporal but good frequency resolution
Window Length in Analysis of Harmonic Spectra

- Schematic of the convolutional view of time-frequency resolution tradeoff with long and short analysis windows for harmonic spectra.
- Long window (a) better represents the spectral harmonicity.
- Short window (b) blurs the spectrum, but better captures changes over time in harmonicity and spectral envelope.