**X. Spectral Analysis/Processing and the FFT**   
read: **D/J**, chapter 7, p.244-258; Roads, chapter 13, pp. 533-577.  
suppl. reading: Strawn pp. 35-67  
  
**harmonic**a frequency that is a whole number (i.e. integer) multiple of another. The smaller number is called the fundamental, or the first harmonic: the second harmonic is the result of the fundamental times two, the third harmonic is the fundamental times three, etc.  
  
**Fourier analysis**  
any periodic waveform (time domain) can be expressed as the sum of one or more harmonically related sine waves called its spectrum (frequency domain), each with a particular frequency, amplitude, and phase. The mathematical procedure for converting a waveform into its spectral components is called the Fourier Transform. The Inverse Fourier Transform converts a spectrum back to a waveform without any loss of information. The harmonic frequencies are called "bins". The amplitudes (or magnitudes) of each bin are known as the "real" part of the spectral signal; the phases are called the "imaginary" part, because imaginary numbers (that use the square root of -1) are required to compute phase.  
  
**Spectra of simple waveforms**  
A sawtooth wave contains all harmonics, each with 1/n times the amplitude of the fundamental (rolloff 6dB/octave.) A square wave consists of only odd harmonics, each with 1/n times the amplitude of the fundamental (rolloff = 6dB/octave). A triangle wave consists of odd harmonics only, each with 1/n2 times the amplitude of the fundamental (rolloff = 12 dB/octave.)   
  
**Pulse wave**the narrower the pulse-width (smaller duty cycle), the richer the spectrum (the greater the amplitude of each harmonic), to the limit that the narrowest possible pulse (called an "impulse" in a digital system) has equal amplitudes of all representable harmonic frequencies.  
  
**Timbre**  
timbre is the name for the perceptual mechanisms by which we classify sound into families. The "Classical Theory of Timbre" says that the spectral composition of a waveform primarily determines the sensation of timbre or tone-color. Additive synthesis methods derived from this theory create timbres by adding together spectral components of varying strengths.  
  
**Inharmonicity**  
truly harmonic spectra are theoretical ideals, and real physical sounds exhibit a range of spectral characteristics ranging from relatively harmonic (like strings and vibrating air-columns) to inharmonic (like percussion instruments and bells) to random-like distributions of frequencies (noise).  
  
**Spectral Evolution**  
spectral evolution describes the amplitude envelope of each spectral component plotted regularly over time, graphically represented in three dimensions (freq, amplitude, time). The spectral evolution of most instruments is different throughout the pitch range and also varies with loudness. The most radical changes in spectral progression occur during the attack portion of a sound, when frequencies change quickly and are usually not harmonic. Analysis and additive resynthesis of acoustic instrument tones using spectral evolutions approximated by 8 line segments per component envelope can be accurate enough to recreate recognizable instrument timbres.  
  
**Formants**  
formants are spectral peaks in absolute frequency regions associated with the resonant modes of vibrating bodies. Timbral identity may often depend more on the recognition of formant similarity than on the similarity of spectral evolution envelopes. Subtractive synthesis methods derived from this type of analysis begin by modeling an excitation source (like vocal cords or reeds) filtered by a resonating body (like the vocal cavity or a horn.)  
  
**Fast Fourier Transform (FFT)**  
the Discrete Fourier Transform (DFT) is an application of the Fourier Transform to digital signals, where values are discrete (finite) and not continuous. The Fast Fourier Transform (FFT) is a restricted version of the DFT that computes faster, and is the classic computational tool used in spectral analysis/resynthesis. A DFT analysis takes place on a specific size of "N" samples at a time, which is called the "analysis window". The FFT requires N to be a multiple of 2, and it divides the signal into N/2 frequency bins. For example, if N is 512 samples, then the number of frequencies that can be analyzed is limited to 256. Assuming a sampling rate of 44.1KHz, we obtain 256 bins equally spaced over a bandwidth from 0 Hz to 22.5 KHz, so the bins are separated from each other by 22,500/256 cycles, or 87.89 Hz. If we want finer frequency resolution, we need to increase size of the analysis window "N", so for example if N = 2048, then our resolution is about 22 Hz. We now have a more accurate description of frequency, but it is averaged over a longer period of time: this is an inevitable time/frequency tradeoff.   
  
**Window functions**  
The FFT performs its analysis on successive groups of N samples, called "time windows" or "frames", in an audio signal or soundfile. Since Fourier analysis depends on the idea that the signal in each frame is a made up of sine waves, or repeating waveforms, the fact that it has a discrete beginning and end where samples are unlikely to match creates errors. Window functions are used to minimize the effect of these errors, effectively linking up the beginning and end of the analyzed segment of the waveform, helping them to behave like single periods of a periodic function even if they are not. A window function is used by multiplying its spectrum with that of the input - the "Hamming" Window is one commonly used function with a curve that never goes to zero, thus allowing inverse transformation to time domain without loss of data

**Overlap-add**  
We compute average spectrum content of the first window Nw samples, then advance the window to the next block of Nw , then the next, etc. If Nw is small, short-term variations in spectrum will be discerned, but since Nw is small, its frequency resolution is also limited. To increase resolution, windows can be overlapped, so that each new set of samples begins at a skip-factor S expressed as a fraction of Nw samples from where the last set began. But there is a trade-off between the size of S and the size of Nw: if S = Nw there is minimum spectrum data and the worst resolution; if S=1/4Nw there is 4 times the spectrum data and 4 times the resolution, but features are 1/4 as prominent as before.   
**Time/Frequency Uncertainty:** we can tell that an event occurred at a precise time but we cannot say exactly what frequencies it contained; conversely, we can pinpoint frequency content only over a longer time interval.  
  
**Phase Vocoder (VOice CODER)**is an analysis/resynthesis process. The analysis section was originally a set of contiguous band-pass filters, whose amplitude envelopes approximated the spectrum of the input at a given point in time. Modern phase vocoders use the FFT in the analysis stage to produce a pair of numbers, one real and one imaginary, for each of Nw /2 "channels" or "bins. The synthesis section implements additive synthesis: each harmonic frequency is computed as a sine wave with particular amplitude and phase, summed for each analysis window to recreate the time domain signal. This stage is called the IFFT, or Inverse Fast Fourier Transform. Between the analysis and resynthesis the frequency, amplitude, and spectrum can be independently manipulated: this allows time-stretching or time-compression without change of frequency, transposition of frequency without stretching or compressing the time (as in tape vari-speed or conventional sampling) and without spectrum compression (as in linear frequency modulation), and various other frequency or amplitude-based spectrum modifications.

**Other Spectral Analysis Techniques**  
Pitch-synchronous analysis first determines the fundamental frequency (pitch-tracking) of the sound and adjusts the analysis window size to match. The amplitude and phase pairs thus represent harmonics of the actual fundamental of the sound, instead of an arbitrary analysis frequency. Tracking Phase Vocoder s (TPV) find amplitude peaks in the spectrum and construct "guides" to smoothly track the prominent frequencies in the spectrum. Another frequency sensitive analysis method uses multiple analysis windows, called wavelets, each with its own start time, duration and frequency, to analyze the same sound.