[11:00am-11:40am] Time-frequency analysis and the magnitude spectrogram

Fourier analysis allows us to see what frequencies occur on average, but does not allow us to localize frequencies in time. A joint time-frequency analysis allows localization in both dimensions.

A magnitude spectrogram is a common visualization that shows the amount of signal power as a function of both time and frequency.

When constructing a magnitude spectrogram, the symmetry of the Fourier series coefficients makes the negative frequencies visually redundant, and it's common to leave them out of the visualization.

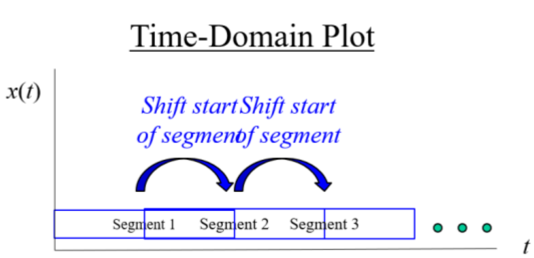
When a signal is sampled at a rate f_s, we can only represent frequencies up to $f_s/2$. The discrete-time Fourier series only has a finite number of harmonic frequencies (up to $\frac{1}{2}f_s$ and down to $-\frac{1}{2}f_s$) and hence we have a finite number of Fourier series coefficients.

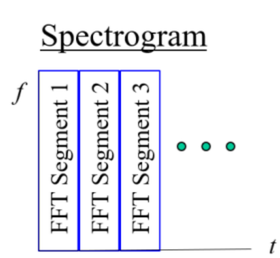
The short-time Fourier transform (STFT) is an intermediate step in constructing the magnitude spectrogram, but includes the magnitude as well as the phase. The STFT includes a window function $w[n-m]$ that isolates the part of the signal centered around m

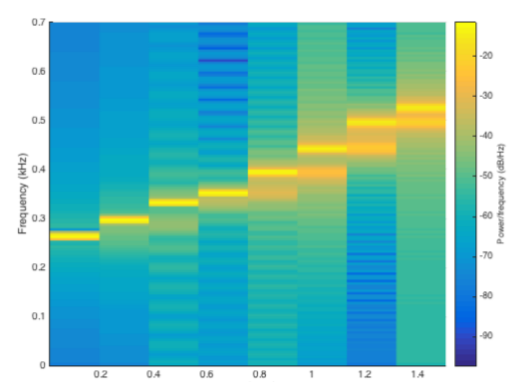
$$\textbf{STFT}\{x[n]\}[m,k] = \sum_{n=0}^{N-1} x[n]w[n-m]e^{-j2\pi\frac{k}{N}n}$$

There is a fundamental trade-off between the time and frequency resolution, since a signal cannot be finite in both time and frequency. Larger values of the window length N will increase the frequency resolution at the cost of lower time resolution.

We can only compute the Fourier series for periodic signals. If we perform Fourier series analysis of a finite length signal, the result is equivalent to periodically extending it. The periodic extension can result in high frequency artifacts whenever the first sample differs from the last sample. Using a tapered window forces the first and last samples to be equal to zero, reducing the discontinuity in the periodic extension.







[11:40-11:50] Chirp signals

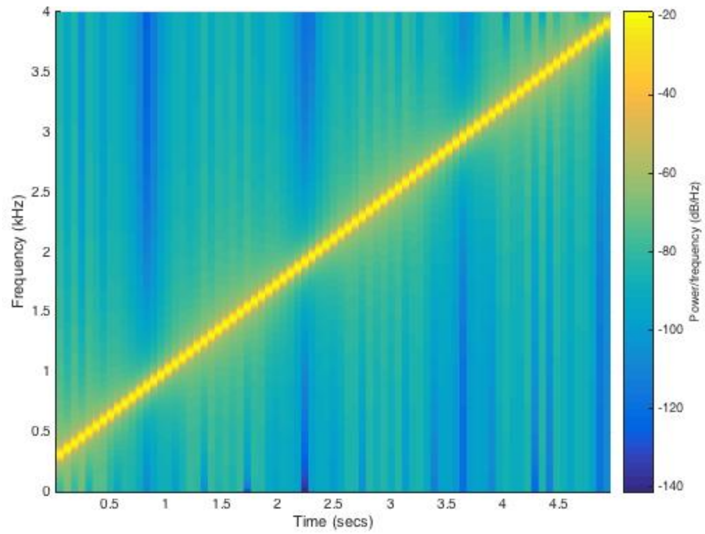
A chirp signal is a sinusoid with increasing or decreasing frequency

$$x(t) = A\cos(\psi(t)), \qquad \psi(t) = 2\pi\mu t^2 + 2\pi f_0 t$$

The instantaneous frequency $f_i(t)$ in Hertz is the scaled slope of the angle $\psi(t)$

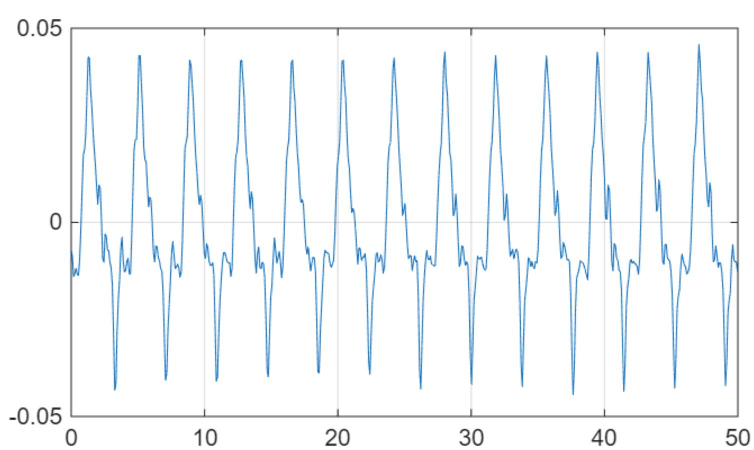
$$f_i(t) = \frac{1}{2\pi}\frac{d}{dt}\psi(t) = 2\mu t + f_0$$

Example: linear chirp from 261Hz to 3951 Hz



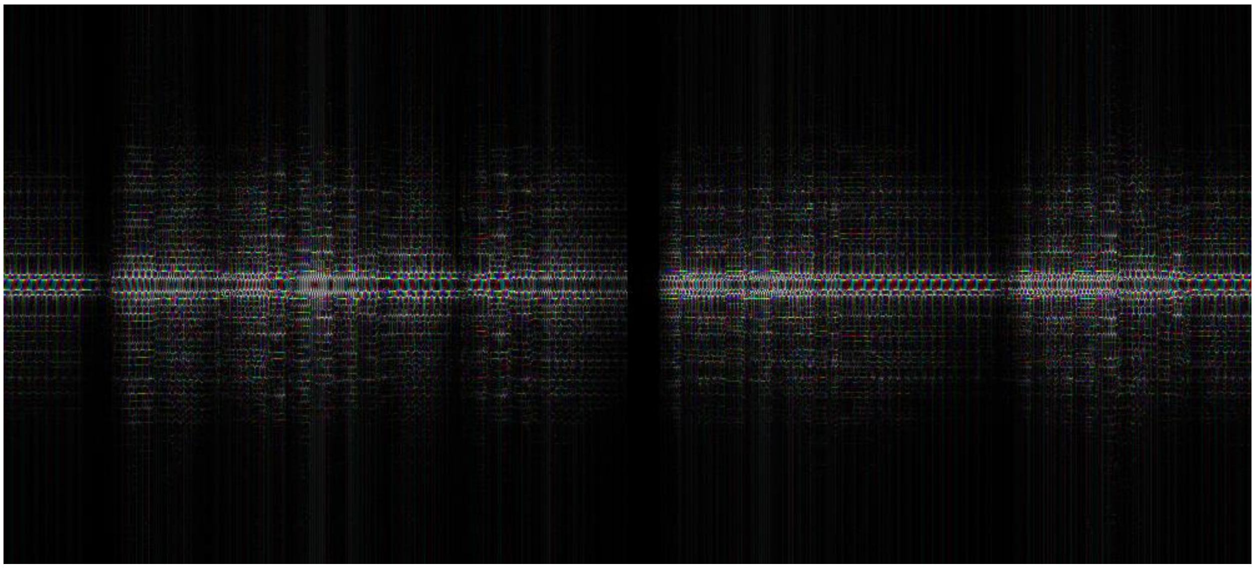
[12:00-12:30] Tune-up and mini project

The average value of an audio signal is the DC frequency component (0 Hertz). Since we cannot hear below 20 Hertz, the average value can be removed. One way to estimate the principle frequency of an audio signal is by eyeballing the plot. Example: 13 periods over 50ms = 260 Hz.



For the mini project, you should map the complex values of the STFT to a color.

Example 1: Magnitude→Value and phase→color using HSV color representation. See the hints for more information on how to convert between RGB and HSV colorspace in MATLAB. On the cello clip, the PSNR of the recovered audio is 34 decibels using this representation.



Example 2: Real part→top half of image, Imaginary part→bottom half of image. Since the Fourier series coefficients of a real-valued signal are conjugate symmetric, we can use half of the image to represent the real component and half of the image to represent the imaginary component. In this example, large magnitude positive values are mapped to red, large magnitude negative values are mapped to blue, and values near zero are mapped to white. On the cello clip, this gives 30 dB PSNR for the reconstruction.