SPRING 2023

CS395T: SPOKEN LANGUAGE TECHNOLOGIES

Fourier Analysis

DAVID HARWATH
Assistant Professor, UTCS

The University of Texas at Austin
Department of Computer Science
College of Natural Sciences

Welcome!

- Problem set 1 out on Canvas
  - Feel free to take a look, but you probably won't be able to complete any parts of it until next Wednesday (Jan 18th)

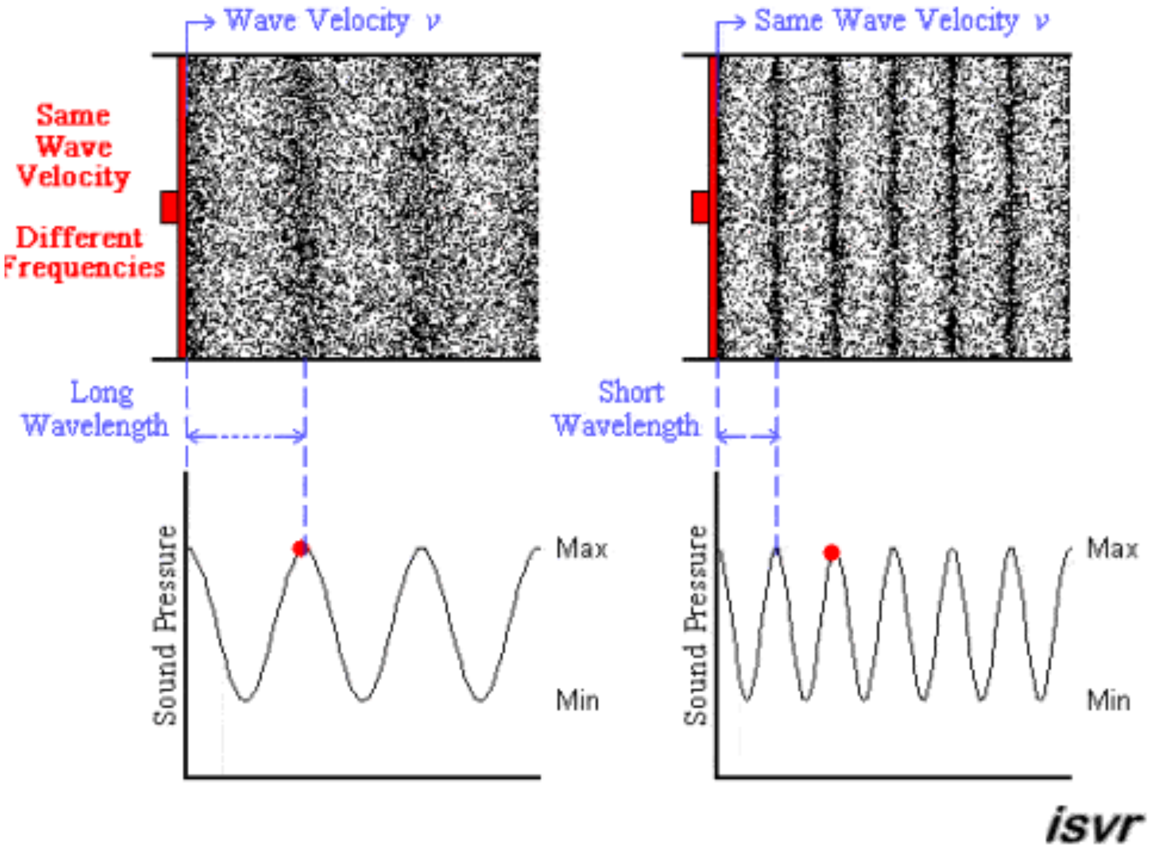
- This will be a dense lecture, with a lot of material you may not have seen before

- Try to understand the **concepts** at an **intuitive level**

Today's agenda

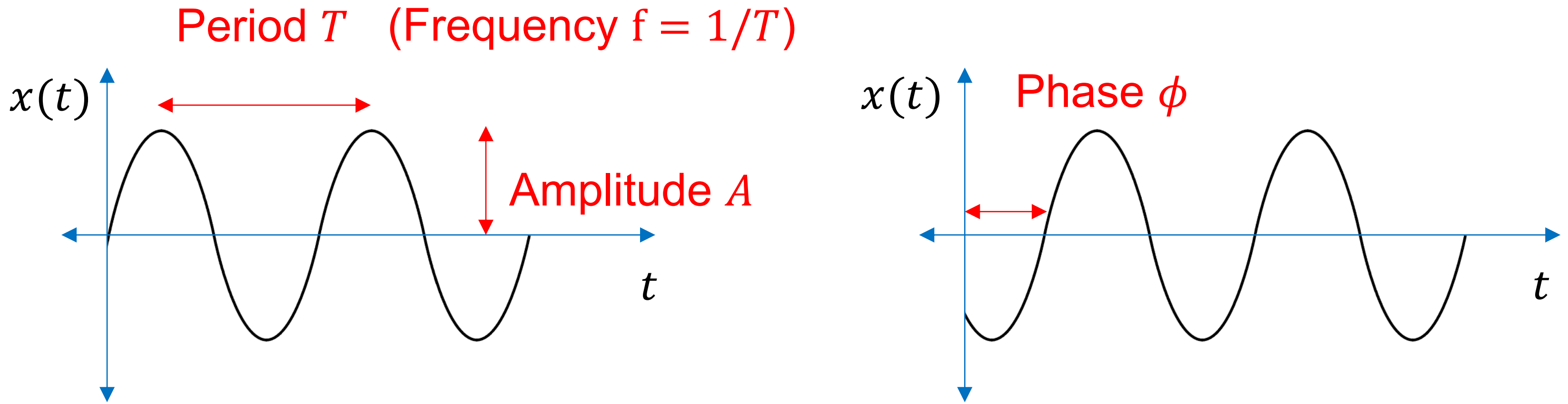
- Fourier Analysis
- Overview of human speech production
- Acoustic tubes
- Modeling the human vocal tract with concatenated acoustic tubes

Sound waves



Wave Velocity v
Same Wave Velocity v
Same
Wave
Velocity
Different
Frequencies
Long
Wavelength
Short
Wavelength
Sound Pressure
Max
Min
Sound Pressure
Max
Min
isvr

Recall: Sinusoids



$$x(t) = A\sin(2\pi f t - \phi)$$

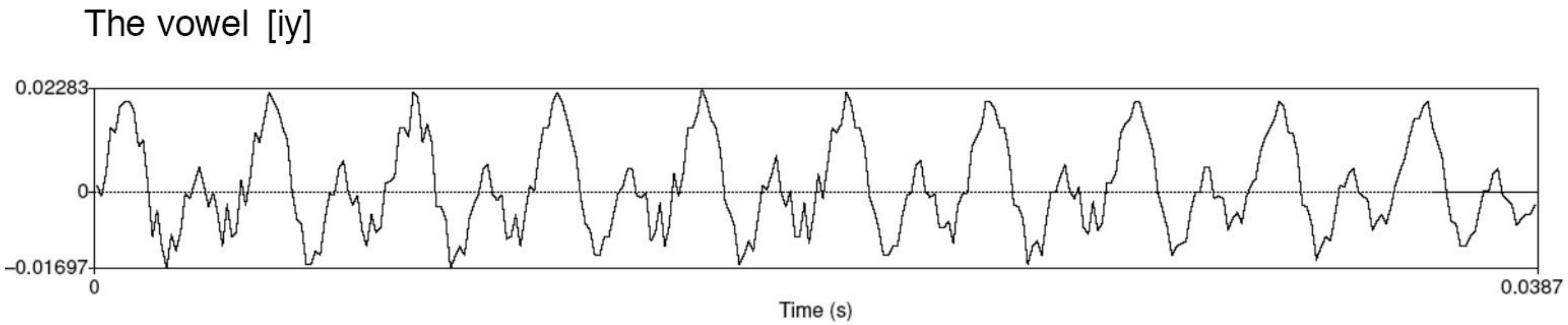
Where frequency f is measured in Hertz (cycles/second)

$$x(t) = A\sin(\omega t - \phi)$$

Where angular frequency $\omega = 2\pi f$ is measured in radians/second

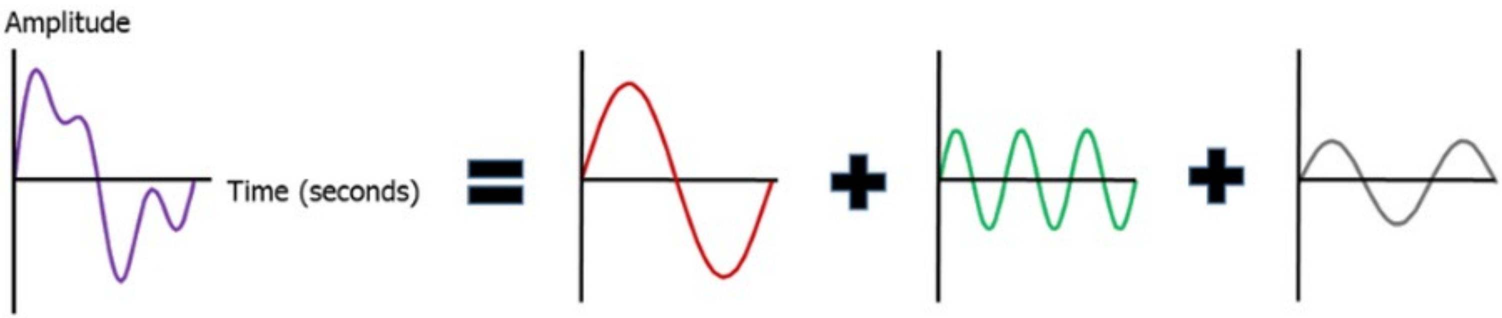
Real world sounds

- Real world sounds are not idealized sinusoids
- They have much more complex waveform shapes that arise from a *mixture of sinusoids at different frequencies.*



The Fourier Series

Basic idea (informal) of the Fourier Series: you can represent any *periodic* function with a weighted discrete sum of pure sinusoids



The Fourier Series

- What makes the Fourier Series so important for this topic is that we can decompose complex sound waves into a *superposition* of simple sound waves.

- This enables us to mathematically understand complex sound waves in terms of the frequencies, magnitudes, and phases of simple sinusoids.

The Fourier Transform

We can compute the Fourier Series coefficients of a function using the *Fourier Transform* (FT).

We can do this for any function (need not be periodic) by taking the limits of integration to infinity (equivalent to Fourier Series with period of infinity)

Let $x(t)$ be a *time domain* function

The FT maps $x(t)$ to a *frequency domain* function $X(\Omega)$

$$X(\Omega) = \int_{-\infty}^{\infty} x(t) e^{-j\Omega t} dt$$

Recall that $e^{j\Omega t} = \cos(\Omega \mathrm{t}) + j\sin(\Omega t)$ (Euler's Formula)

Sometimes we will write frequency as Ω (radians), sometimes we will write it as f (cycles/second). To convert: $\Omega = 2\pi f$

The Fourier Transform

For a given frequency Ω, $X(\Omega)$ tells us

- "How much" sinusoid at frequency Ω is contained within $x(t)$ (magnitude)
- It also tells us the phase shift of the sinusoid at Ω

We can recover both of these quantities from $X(\Omega)$ because $X(\Omega)$ has a real part $\Re\big(X(\Omega)\big)$ and an imaginary part $\Im\big(X(\Omega)\big)$

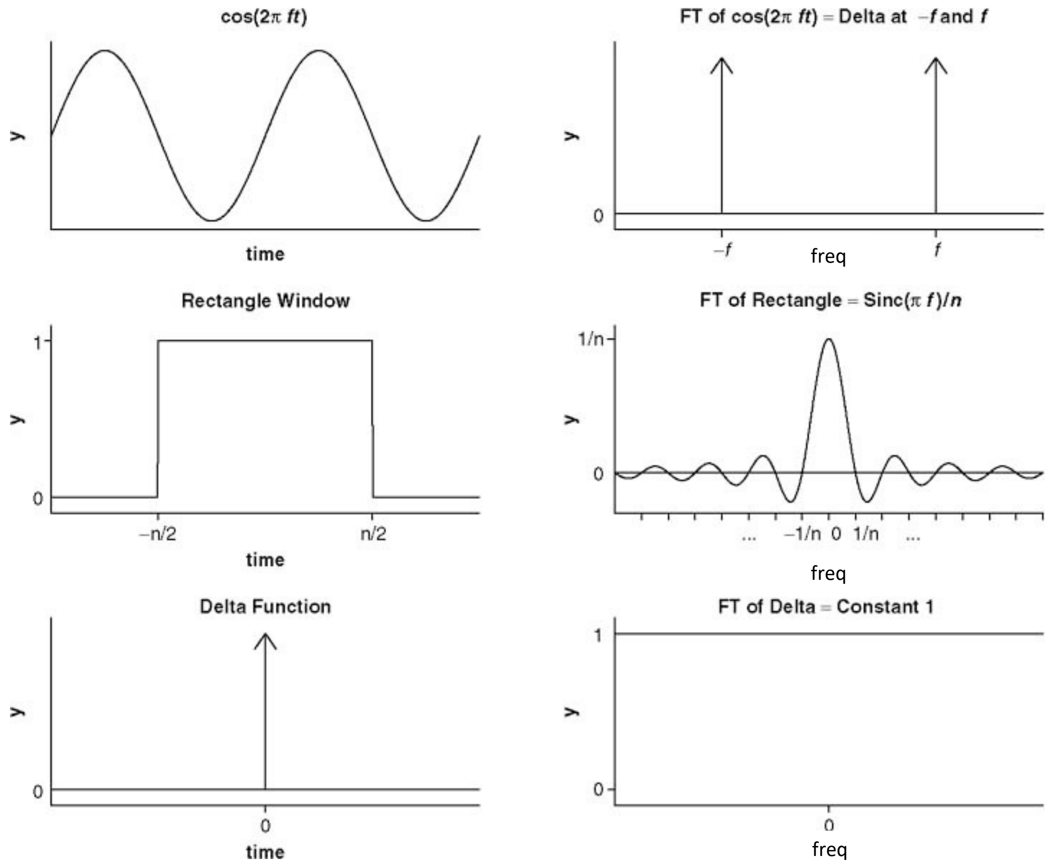
Magnitude $M(\Omega) = \sqrt{\Re(X(\Omega))^2 + \Im(X(\Omega))^2}$

Phase $\phi(\Omega) = \tan^{-1} \frac{\Im(X(\Omega))}{\Re(X(\Omega))}$

The Fourier Transform

- For our purposes, we will usually ignore phase because human hearing is generally insensitive to it
- What is important to us is the amount of *energy* at a particular frequency, which is just the squared magnitude, i.e. $\mathrm{E}(\Omega) = M(\Omega)^2$
- In this class we will always be looking at the *magnitude spectrum* $M(\Omega)$ or *energy spectrum* $\mathrm{E}(\Omega)$ of a waveform.

Fourier Transform Pairs

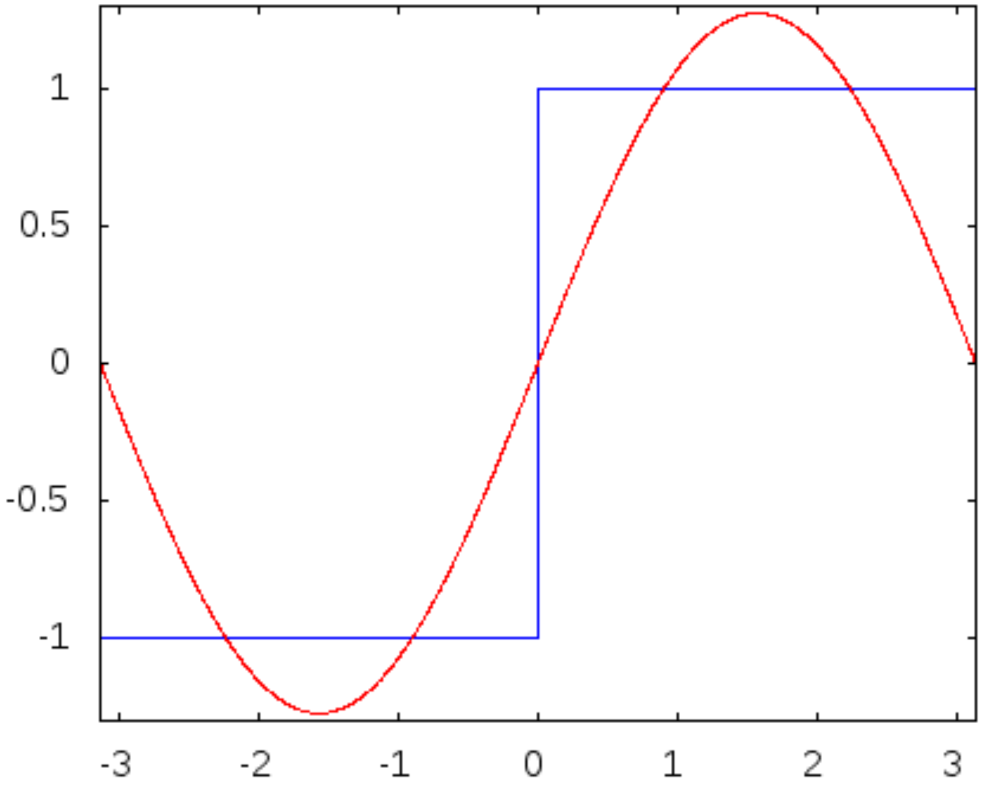


Notice that the FT is defined over *positive* as well as *negative* frequencies

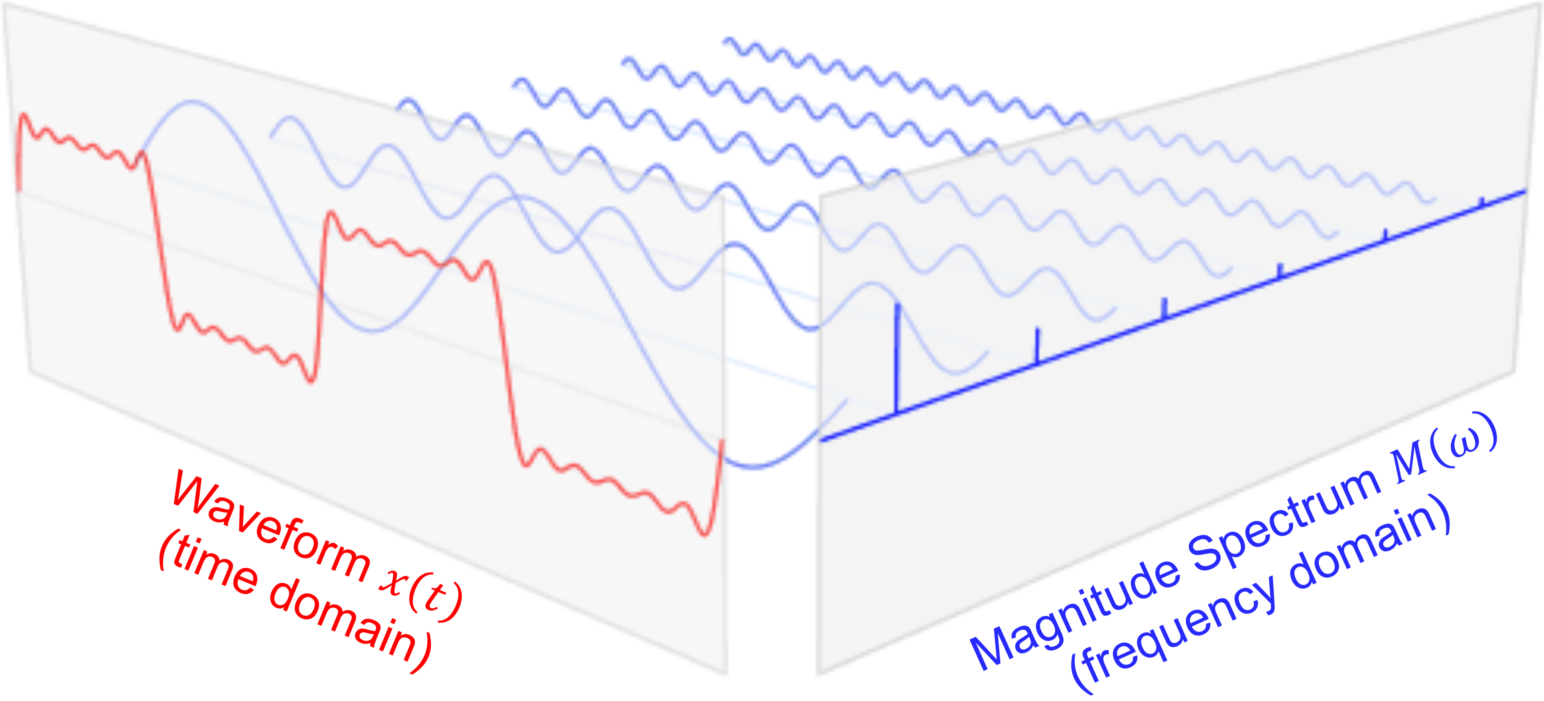
However, real-valued time domain signals (such as sound waveforms) have a conjugate symmetric FT (but symmetric magnitude spectrum)

In this course, when we draw FTs we will generally only draw the positive frequencies.

Fourier Series of a Square Wave



Visualizing Fourier Transforms

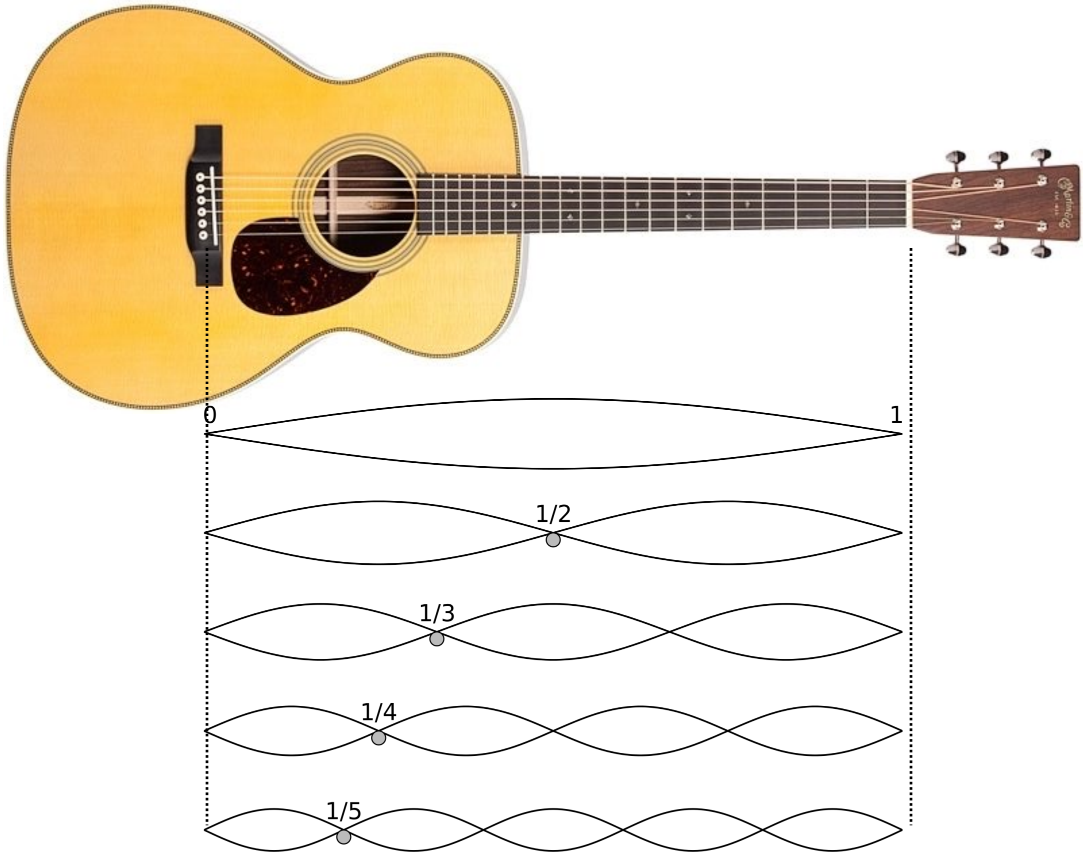


Pure Tones, Harmonics, Timbre

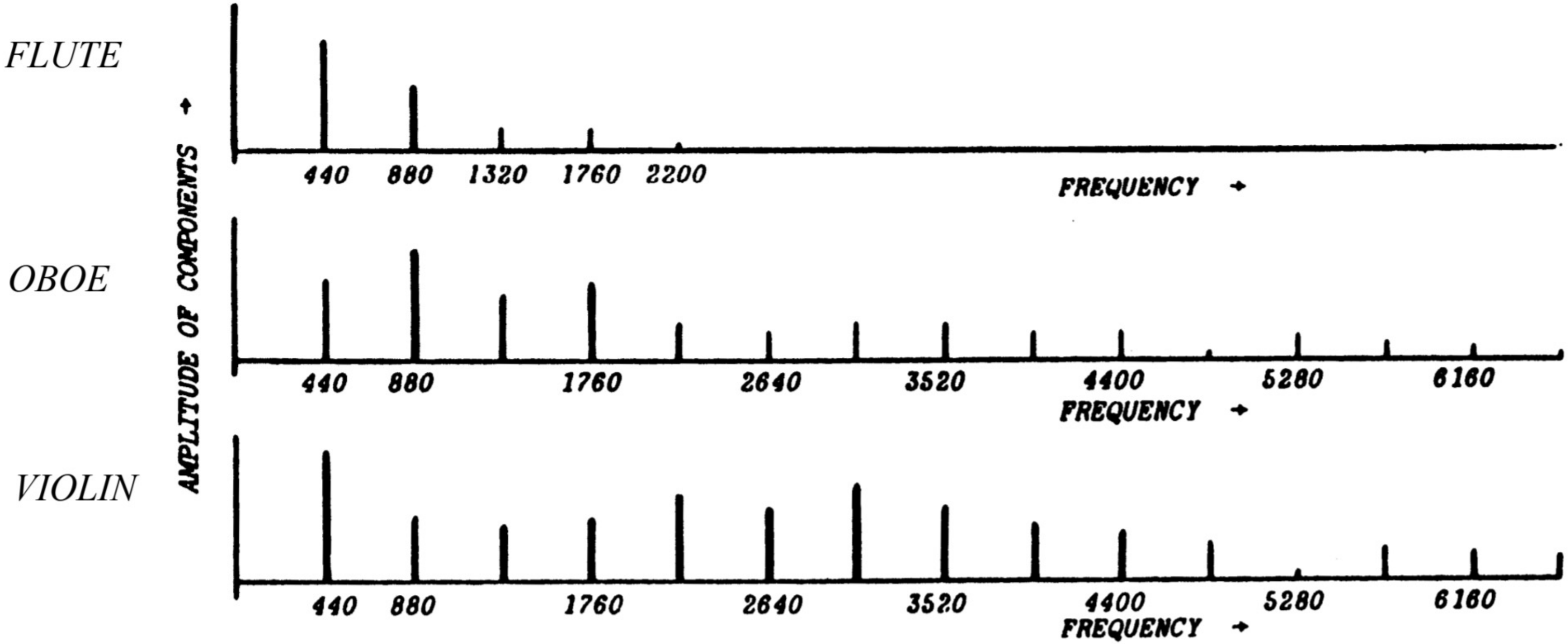
- Sounds in the real world are rarely made up of a single frequency
  - A tuning fork is a good example of a sound that is almost a pure tone at one specific frequency

- Most sounds that we perceive to have a particular pitch are actually made up of a superposition of a fundamental frequency and harmonics

Pure Tones, Harmonics, Timbre



Pure Tones, Harmonics, Timbre



Questions?

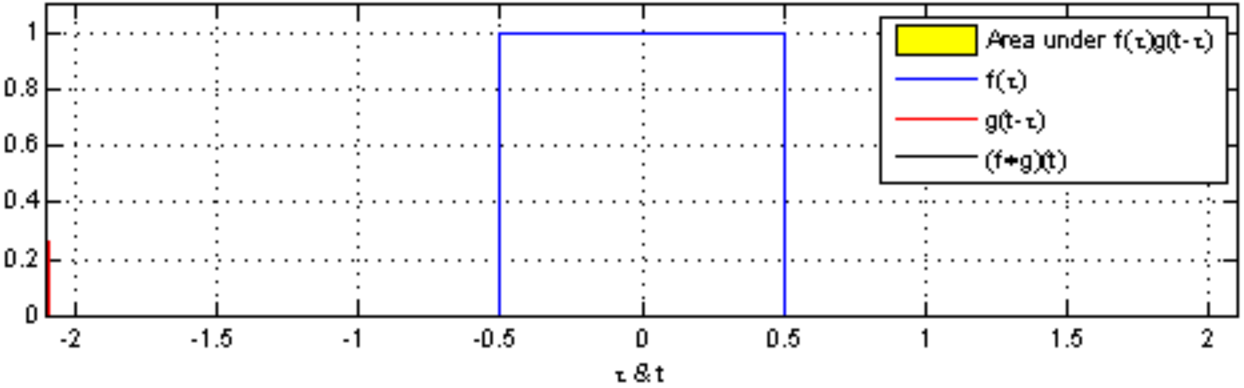
The Inverse Fourier Transform

- The Fourier transform is invertible – we can go from the frequency domain back to the time domain

$$x(\mathrm{t}) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\Omega)e^{j\Omega t}d\Omega$$

1-D Convolution

$$f(t) * g(t) = \int_{-\infty}^{\infty} f(\tau)g(t-\tau)d\tau$$



Source: Wikimedia commons

The Convolution Theorem

- Let $x_1(t)$ and $x_2(t)$ be time-domain signals
- Let $\mathcal{F}(\cdot)$ be the Fourier Transform operator, and $\mathcal{F}^{-1}(\cdot)$ be the Inverse Fourier Transform operator
- The Convolution Theorem states that:

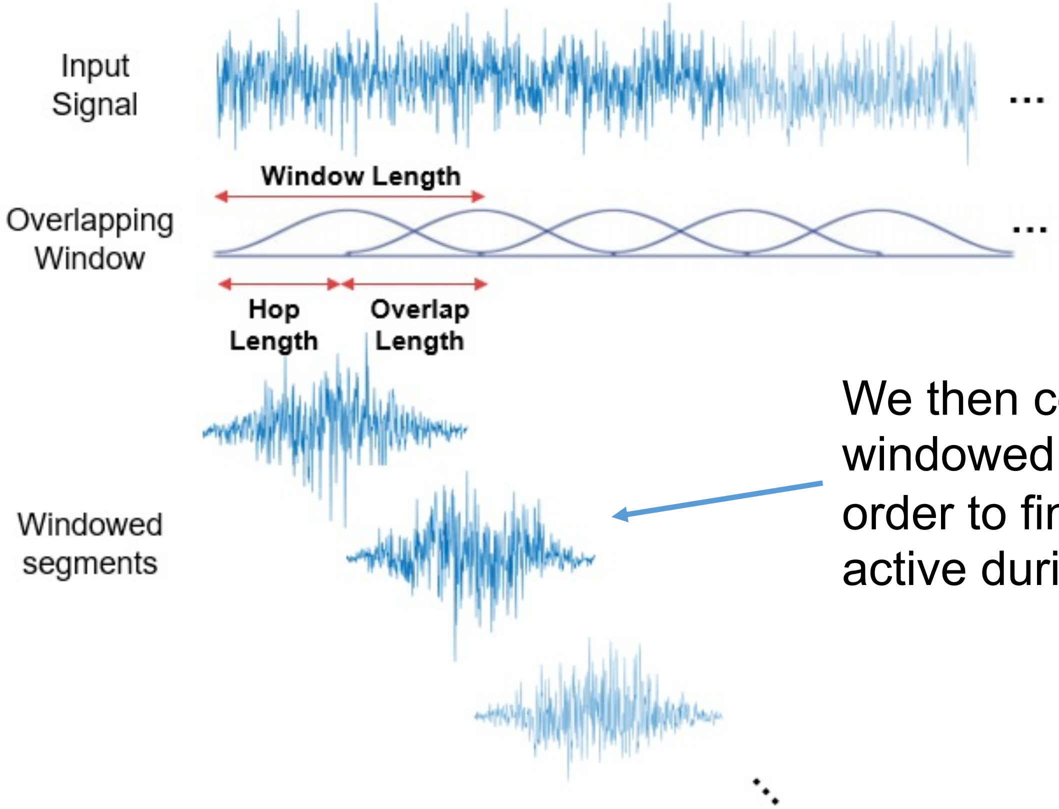
$$x_1(t) * x_2(t) = \mathcal{F}^{-1}(\mathcal{F}(x_1(t))\mathcal{F}(x_2(t)))$$
$$x_1(t)x_2(t) = \mathcal{F}^{-1}\left(\mathcal{F}\left(x_1(t)\right) * \mathcal{F}\left(x_2(t)\right)\right)$$

- In plain English: Convolution in the time domain is equivalent to multiplication in the frequency domain (and vice versa)

The Short-Time Fourier Transform

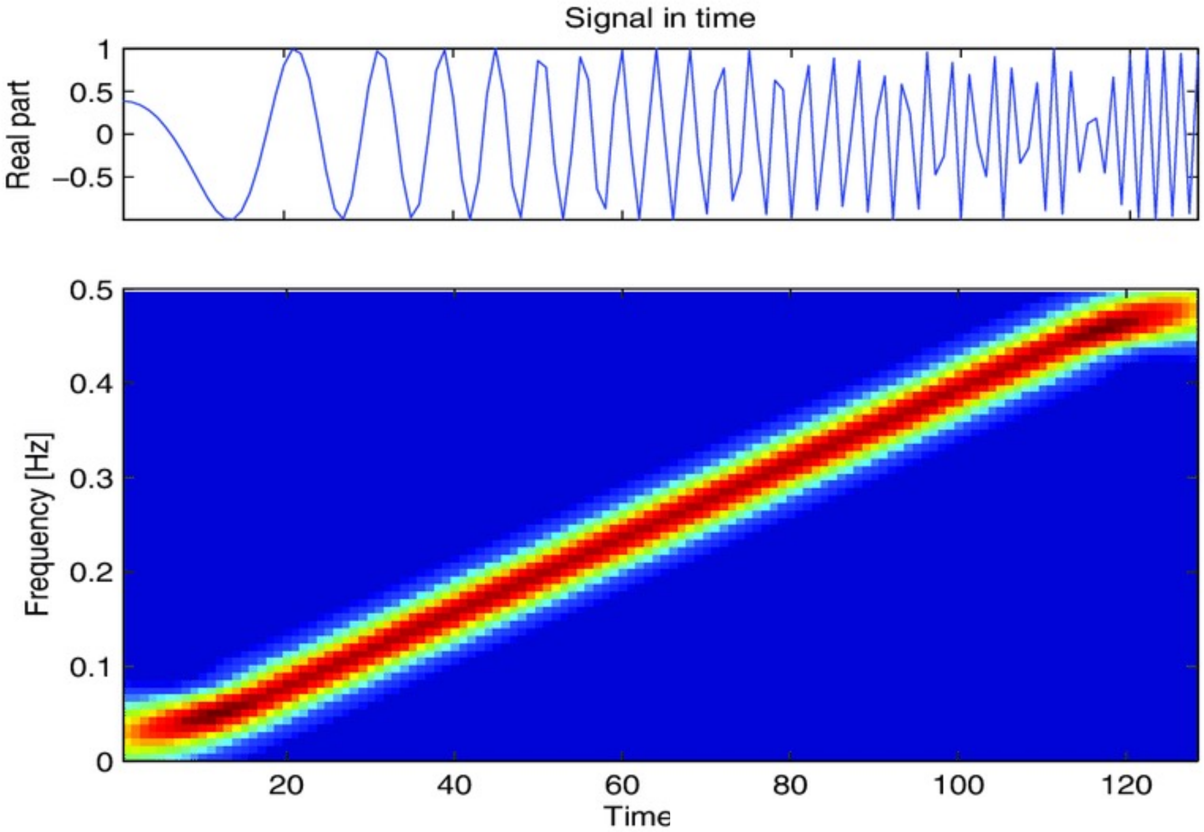
- Problem: Some waveforms (such as speech) are locally periodic, but change over longer timespans. In the case of speech, we care about how the spectrum changes over time because these dynamics are what indicate the sequence of phones (and words) that were spoken
- Solution: Apply a sliding temporal window over the waveform, and compute a Fourier transform for each windowed piece of the waveform.
- This is called the Short-Time Fourier Transform (STFT)

The Short-Time Fourier Transform

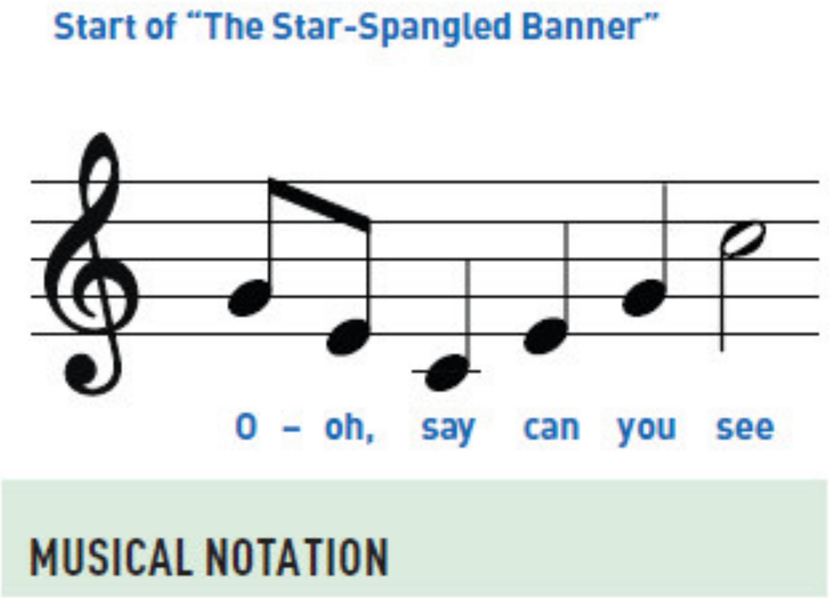


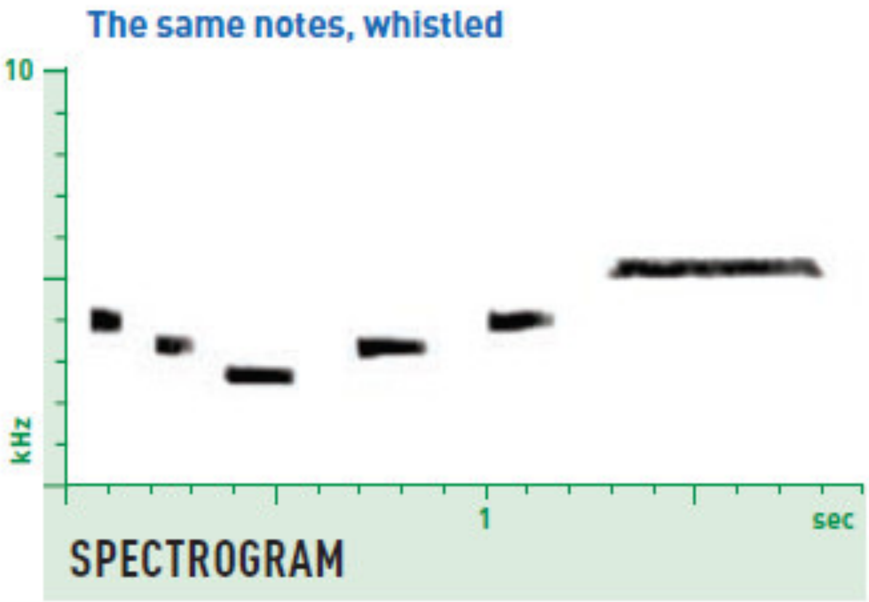
We then compute the FT of each windowed segment independently in order to find out what frequencies are active during a particular window of time

Spectrograms



Analogy to musical notation



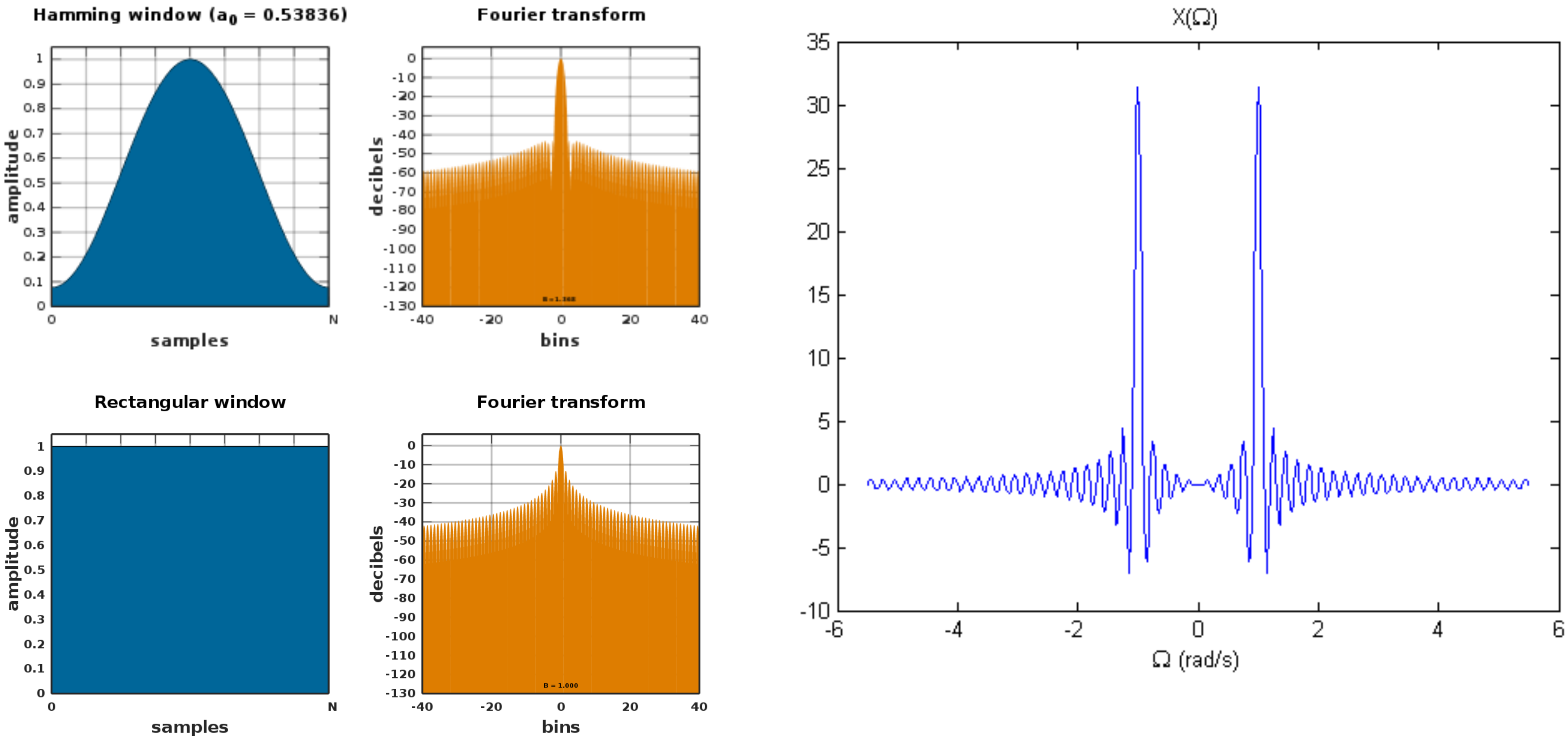


The Short-Time Fourier Transform

- Caveat: The application of the sliding window $w(t-\tau)$ to the signal $x(t)$ *distorts* the spectrum
- This is because we aren't computing the spectrum of $x(t)$, we're computing the spectrum of $x(t)w(t-\tau)$
- Recall the Convolution Theorem: Multiplication in the time domain is equivalent to convolution in the frequency domain, and vice versa

$$x(t)w(t) \leftrightarrow X(\Omega) * W(\Omega)$$

Spectral leakage from windowing



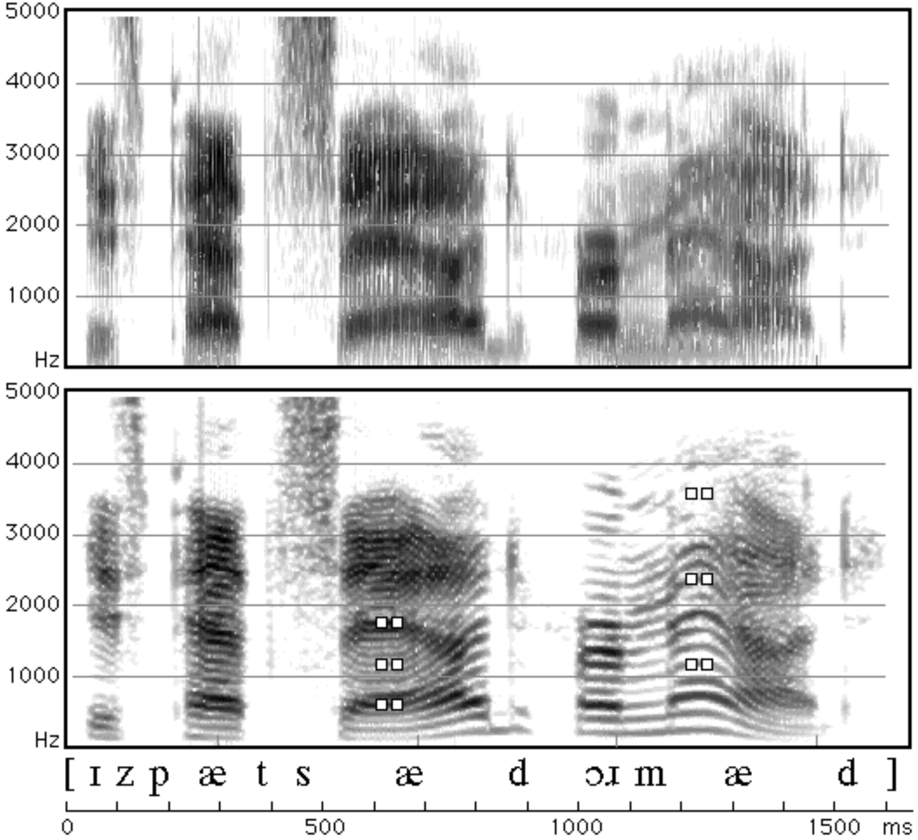
STFT Window Size

- The size of the sliding window controls the tradeoff between frequency resolution and time resolution
- Longer window: more oscillations captured means better frequency resolution, but poorer time resolution
- Shorter window: poorer frequency resolution, but better time resolution
- This is often called the time-frequency “uncertainty principle” in signal processing

Narrowband vs. Wideband Spectrograms

Wideband spectrogram: Short STFT window blurs together harmonics, but gives sharper time detail

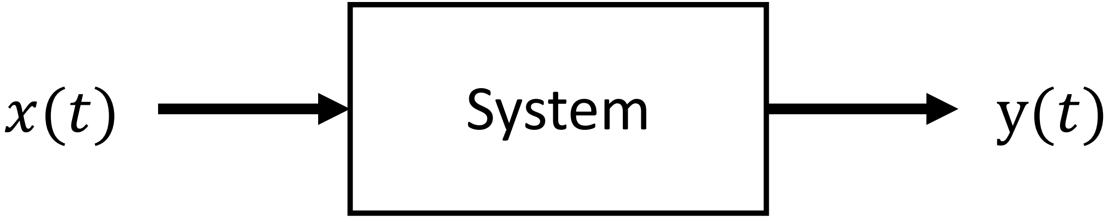
Narrowband spectrogram: Long STFT window reveals voicing harmonics, but with worse time detail (e.g. for stop consonants)



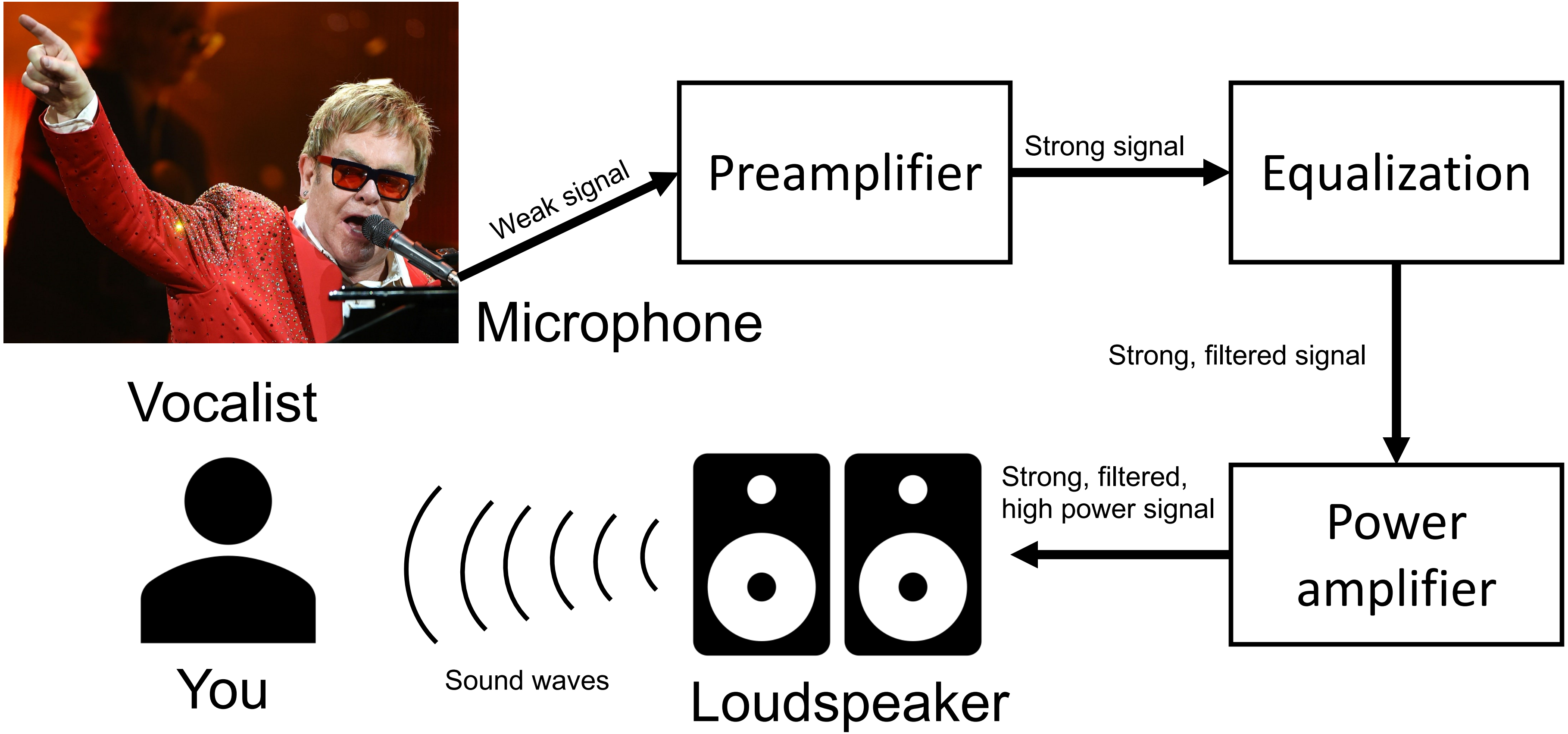
Questions?

Signal processing systems

- One of the most useful and common applications of Fourier analysis is modeling systems
- In signal processing terminology, a *system* is a black box that has an input signal and and output signal



Everyday examples of systems



Linear, time-invariant systems

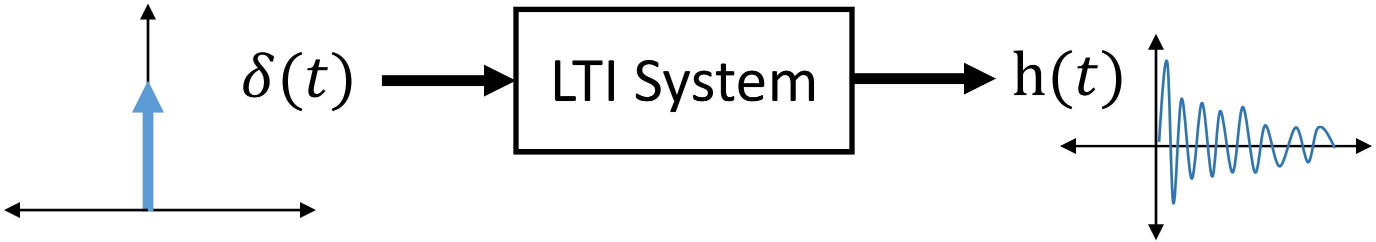
- A special type of system is the *linear, time-invariant (LTI)* system
- It is special because it is mathematically easy to analyze and commonly encountered in the world

$$x(t) \longrightarrow \boxed{\text{LTI System}} \longrightarrow \mathrm{y}(t)$$

- Linear: $Ax_1(t) + Bx_2(t) \rightarrow Ay_1(t) + By_2(t)$
- Time-Invariant: $x(t+\tau) \rightarrow y(t+\tau)$

Impulse response of LTI systems

- A special property of LTI systems is that their behavior is completely characterized by their *impulse response*
- Impulse response: the output of a system when a delta function is fed as input
  - Or: “smack the system with a hammer and let it ring”



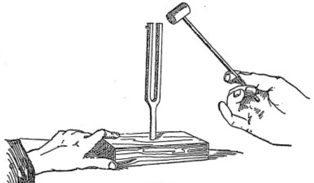


FIG. 7

Impulse response of LTI systems

If a system is LTI, then its input-output relationship will be the convolution of the input $x(t)$ with the system's impulse response $h(t)$

$$x(t) \longrightarrow \boxed{\text{LTI System}} \longrightarrow y(t) = x(t) * h(t)$$

Frequency response of LTI systems

If we analyze an LTI system in the frequency domain instead of the time domain, the input-output relationship is determined by multiplication of $X(\Omega)$ with the system's *frequency response* $H(\Omega)$

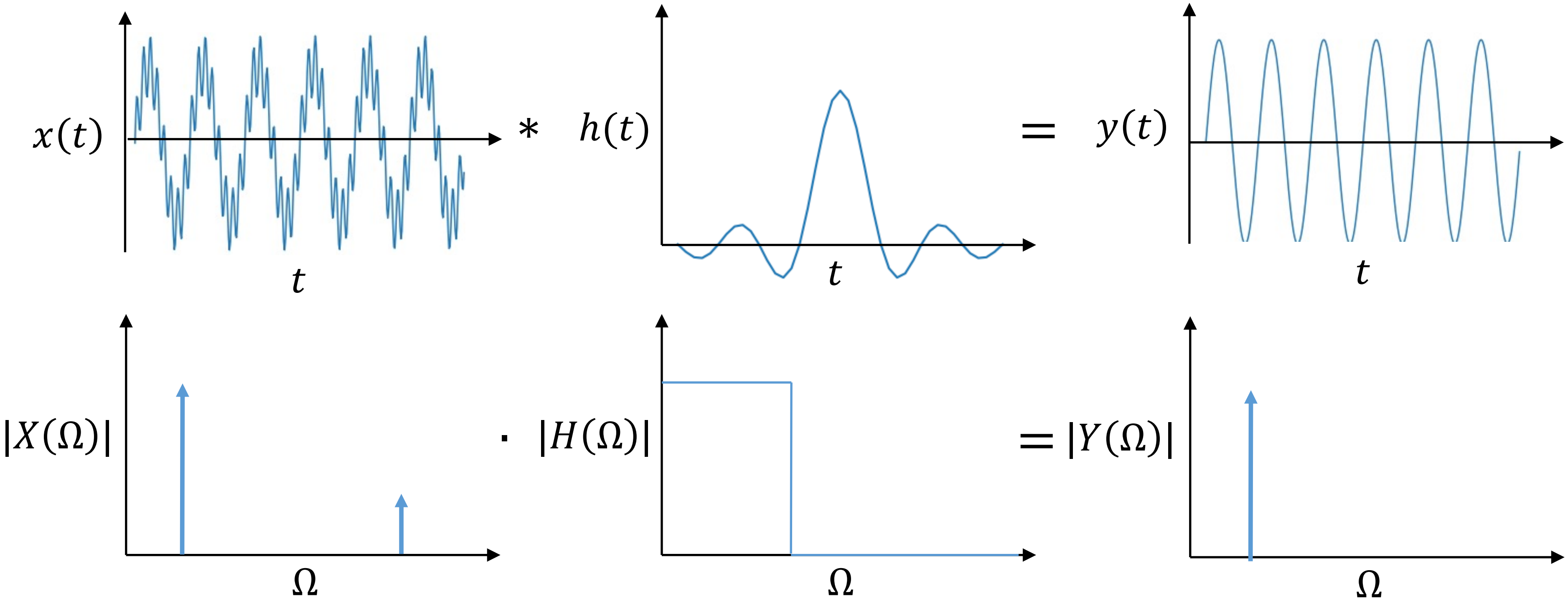
Time domain $\quad x(t) \longrightarrow$ LTI System $\longrightarrow y(t) = x(t) * h(t)$

Frequency Domain $\quad X(\Omega) \longrightarrow$ LTI System $\longrightarrow Y(\Omega) = X(\Omega)H(\Omega)$

Filters

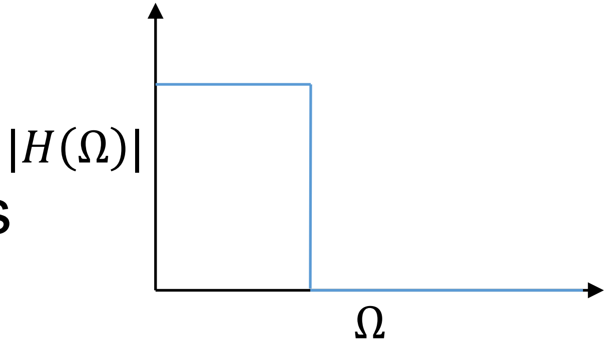
- One of the most useful types of LTI systems is the *filter*
- A filter is used to selectively adjust the amplitude of different frequencies in an input signal
- Example: the equalizer on a home audio system

Filtering example

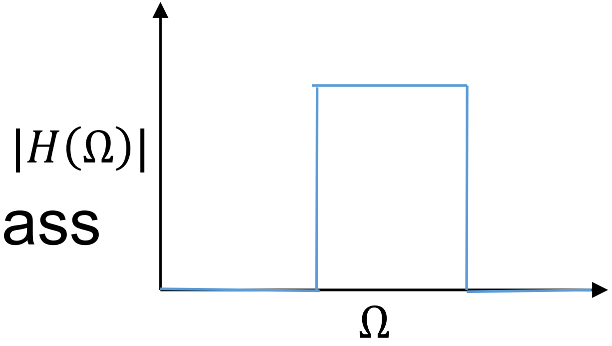


Basic filter types (idealized)

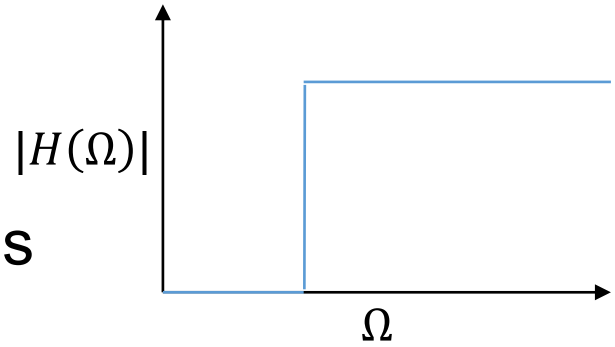
Low pass



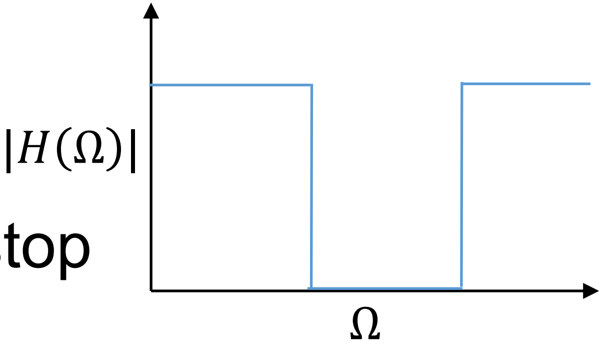
Band pass



High pass



Band stop



Poles and Zeros

- The resonance frequencies of a system are called *poles*
- The anti-resonance frequencies are called *zeros*

We can make complex filter shapes by manipulating the parameters of an LTI system that control the locations of the poles and zeros