

PAT 463/563 (Fall 2026)

Music & AI

Lecture 6: Spectral Analysis

Instructor: Hao-Wen (Herman) Dong

Music Representations



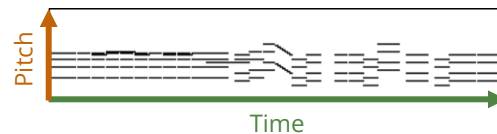
Symbolic music representations

Text-based

Image-based

```
Program_change_0,  
Note_on_60, Time_shift_2, Note_off_60,  
Note_on_60, Time_shift_2, Note_off_60,  
Note_on_76, Time_shift_2, Note_off_67,  
Note_on_67, Time_shift_2, Note_off_67,  
...
```

MIDI



Piano roll



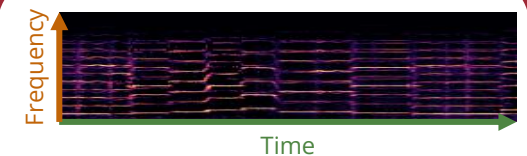
Audio-domain music representations

Time series-based

Image-based



Waveform



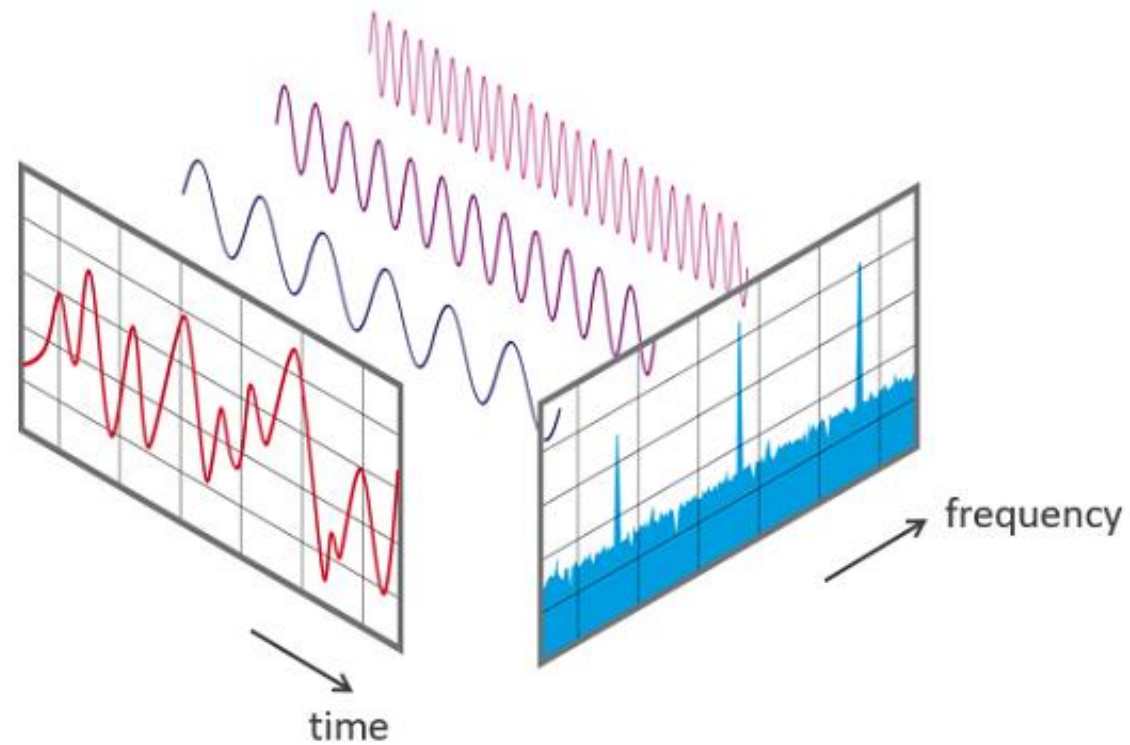
Spectrogram

Today's topic!

Spectral Analysis

| Spectral Analysis

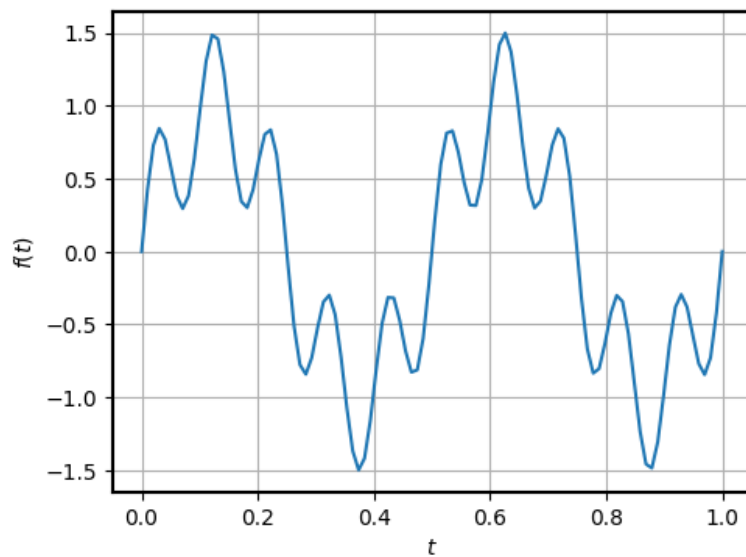
- **Goal:** Analyze the **frequency components** of a signal



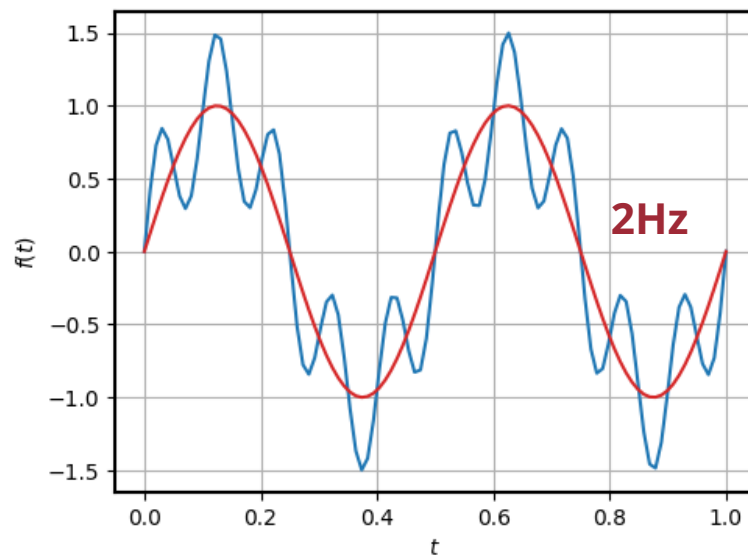
(Source: Phonocal)

Spectral Analysis

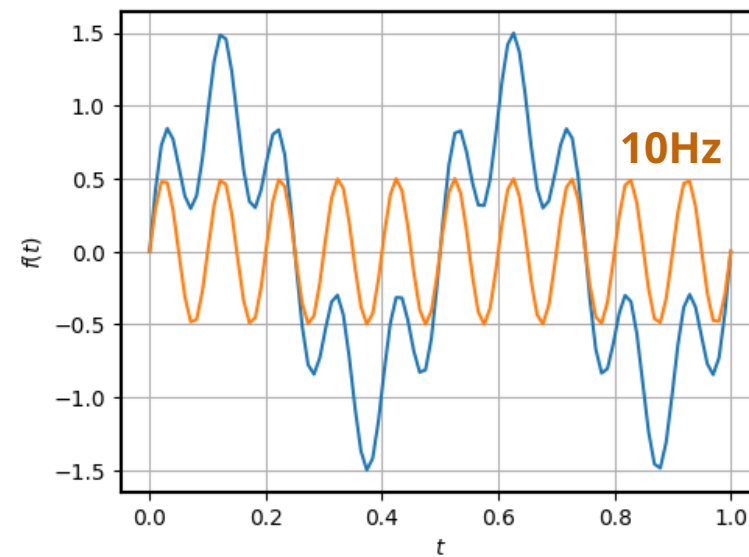
- **Goal:** Analyze the **frequency components** of a signal



$$\sin(2 \cdot 2\pi t) + \frac{1}{2} \sin(10 \cdot 2\pi t)$$

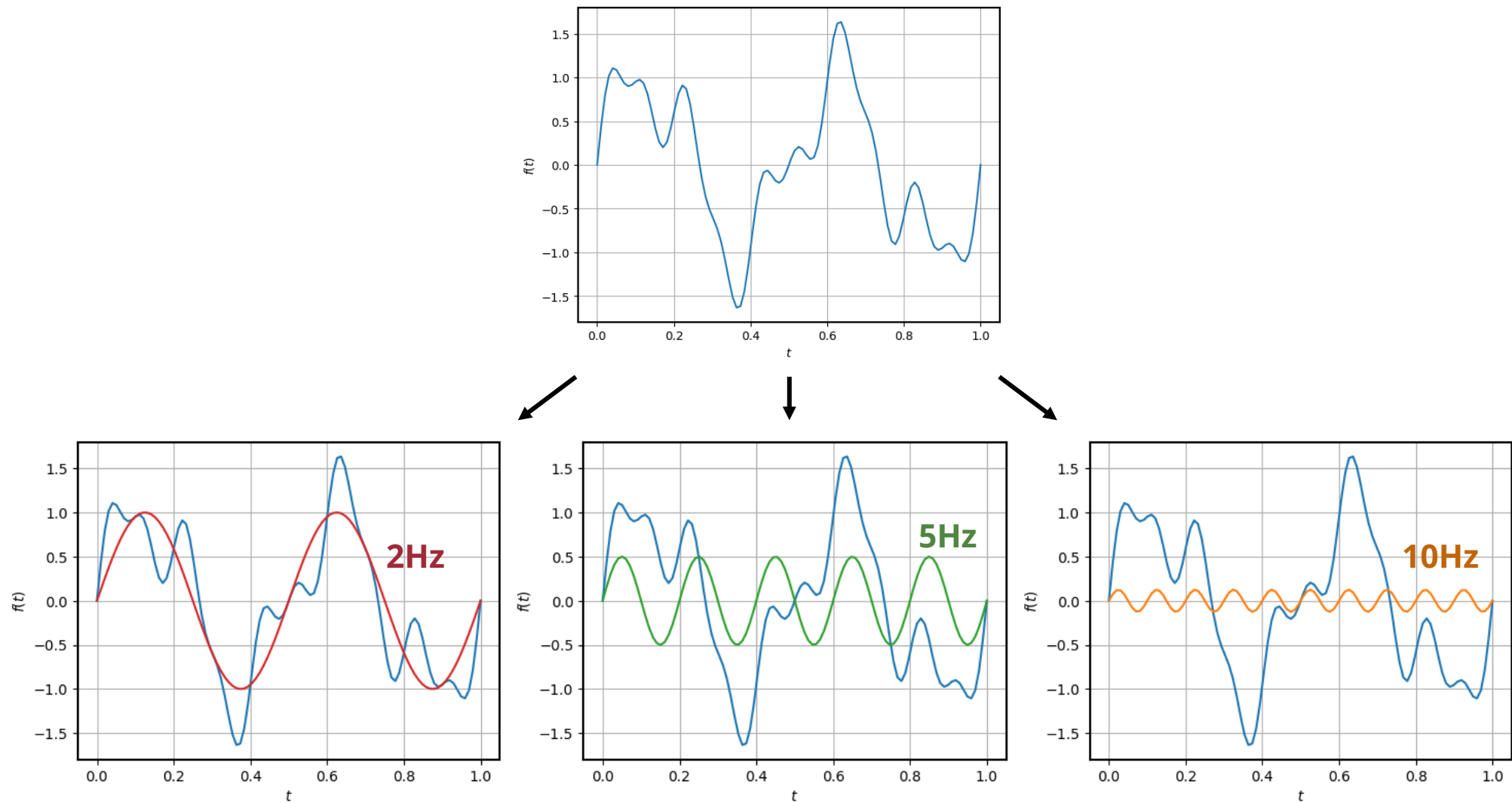


$$\sin(2 \cdot 2\pi t)$$



$$\frac{1}{2} \sin(10 \cdot 2\pi t)$$

Spectral Analysis



Fourier Transform

- **Intuition:** Decompose time-domain signals into **frequency components**
- Math formulation:

The diagram illustrates the Fourier Transform equation with various annotations:

- Output spectrum:** A green arrow points to the $F(\omega)$ term, which is enclosed in a green box.
- Frequency:** An orange arrow points to the ω symbol inside the $F(\omega)$ box.
- Input signal:** A blue arrow points to the $f(t)$ term, which is enclosed in a blue box.
- Sum over all t :** A purple arrow points to the integral symbol $\int_{-\infty}^{\infty}$, which is enclosed in a purple box.
- Complex exponential:** The term $e^{-j\omega t}$ is enclosed in a red box, with a thinking face emoji above it.
- Differential element:** The term dt is enclosed in a purple box, with a purple arrow pointing to it.

The equation is presented as: $F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt$

Demystifying Fourier Transform

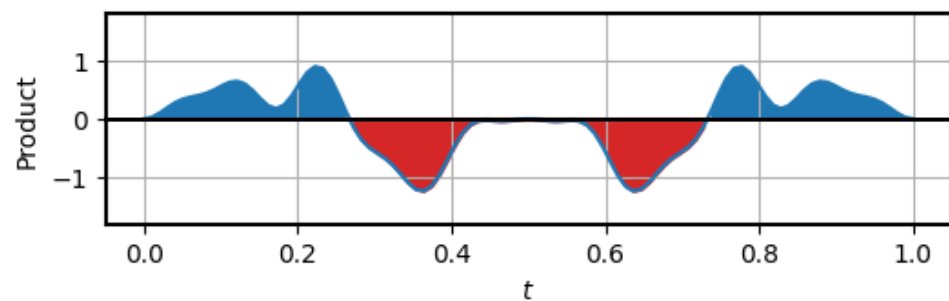
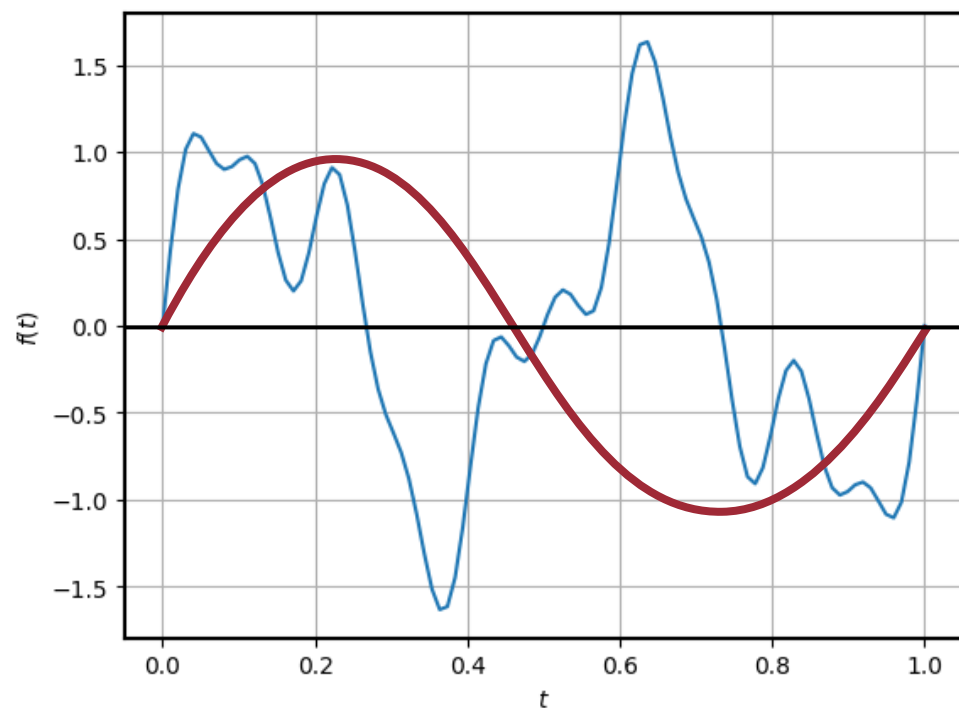
$$F(\omega) = \int_{-\infty}^{\infty} f(t) \boxed{e^{-j\omega t}} dt$$


Euler's formula

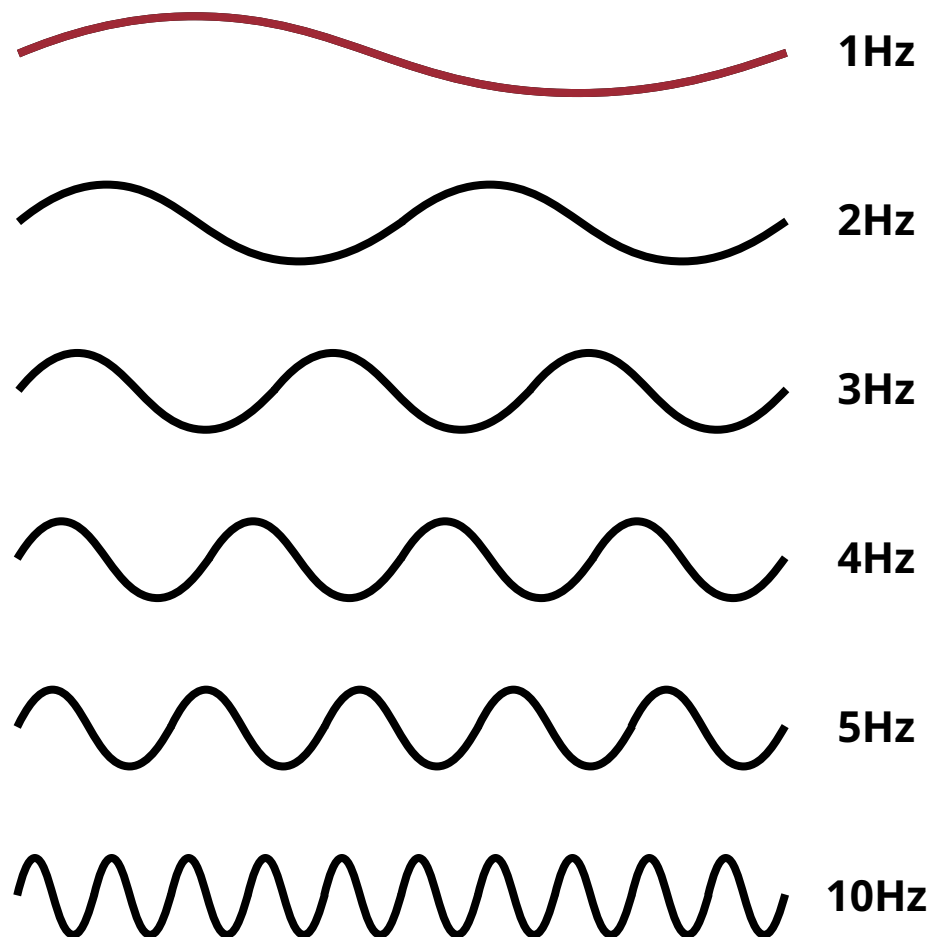
$$e^{-j\theta} = \cos \theta + j \sin \theta$$

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \boxed{\cos(-\omega t)} + j f(t) \boxed{\sin(-\omega t)} dt$$

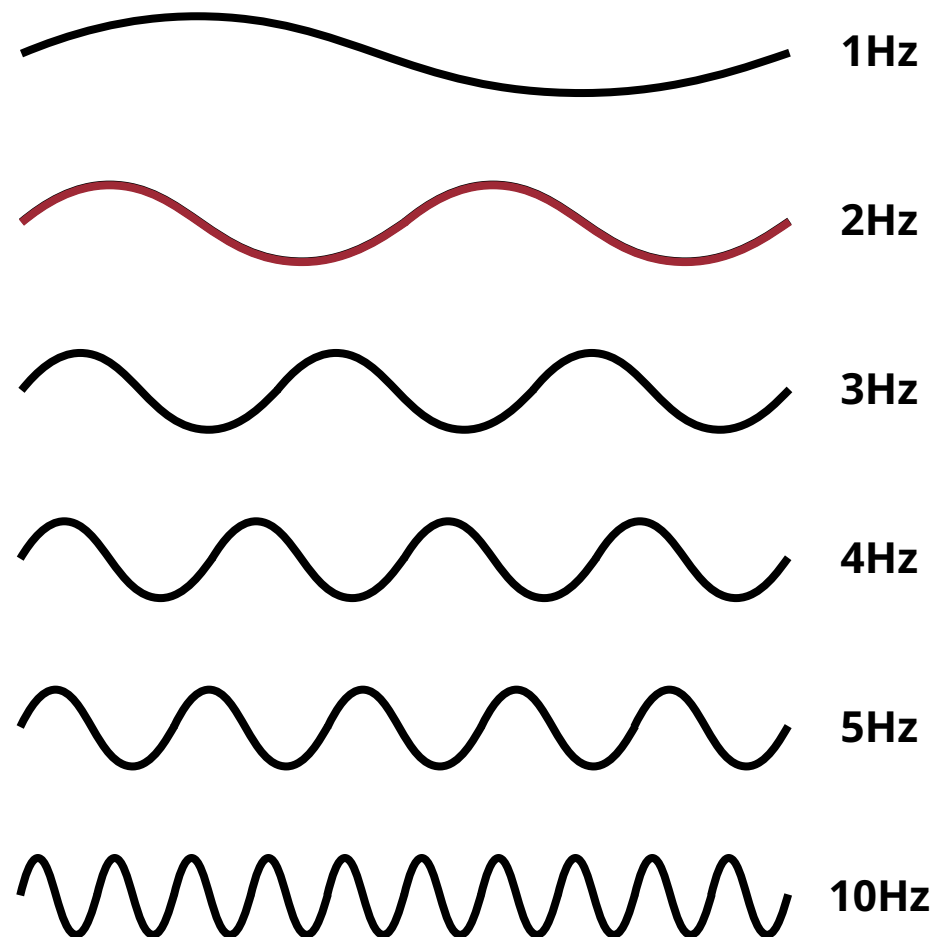
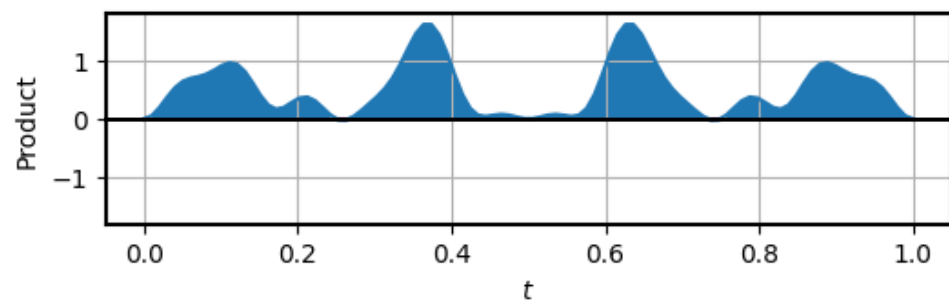
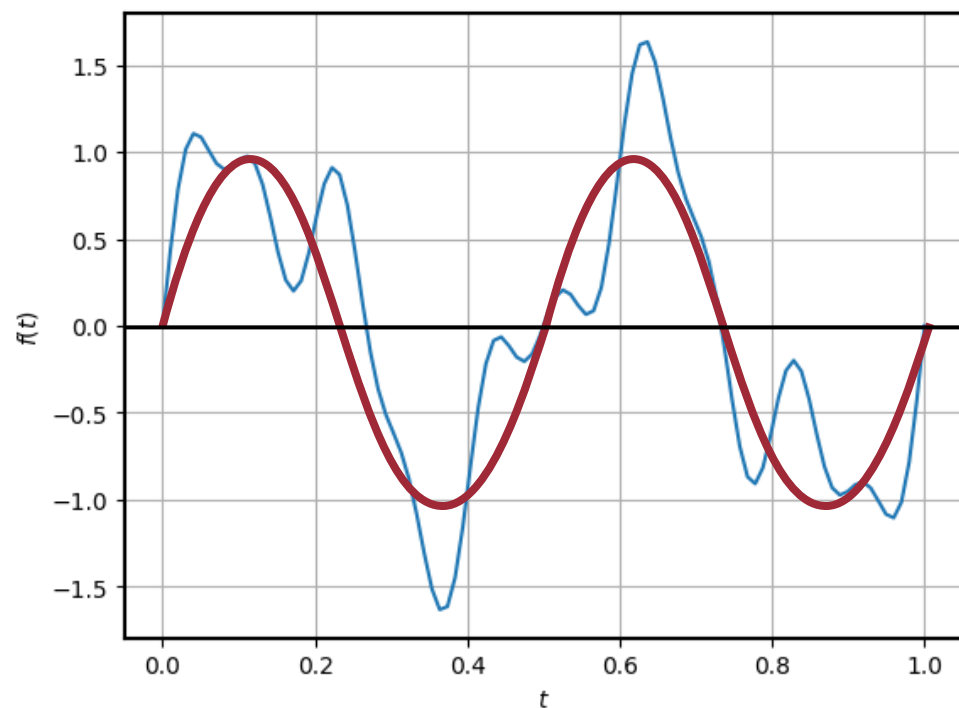
Demystifying Fourier Transform



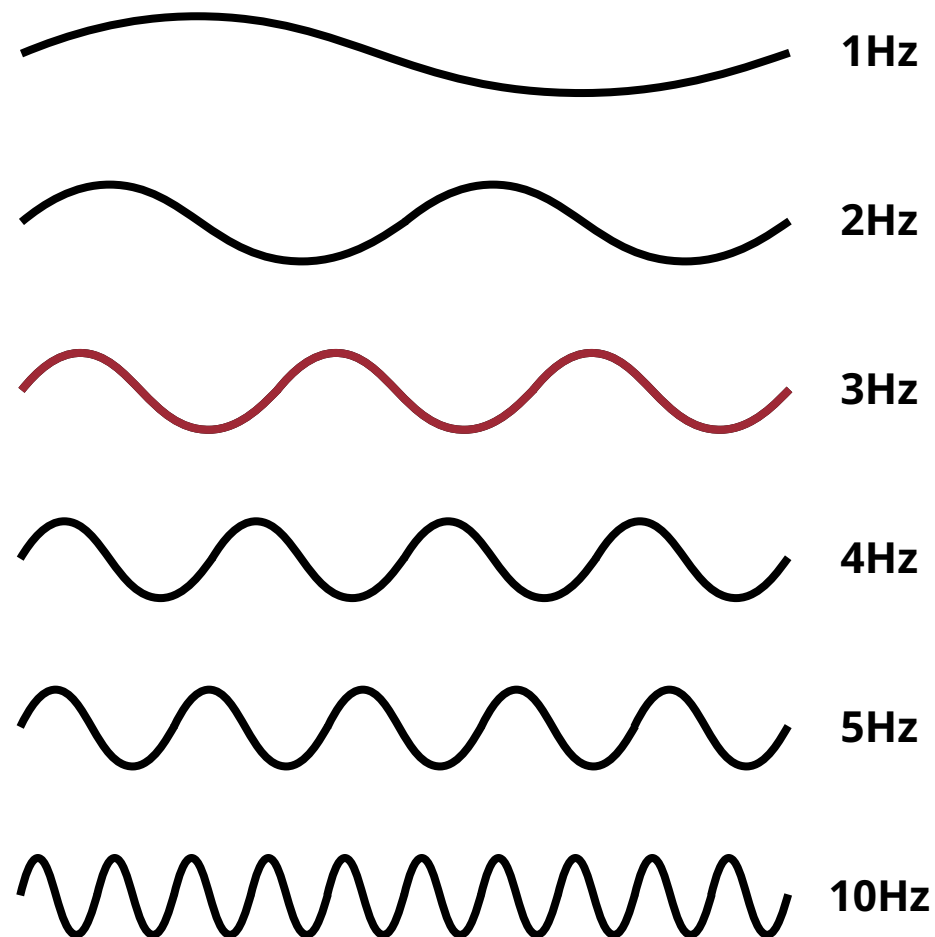
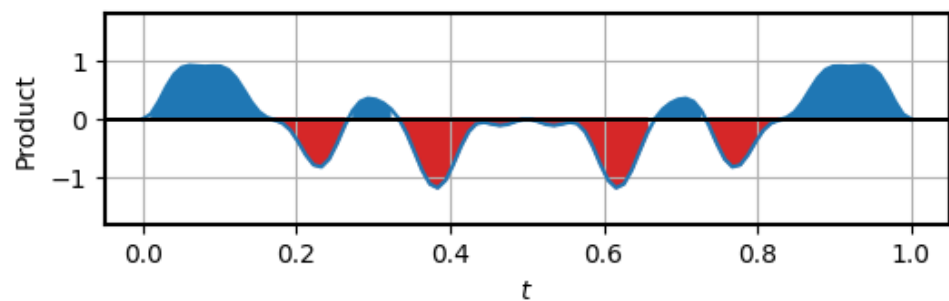
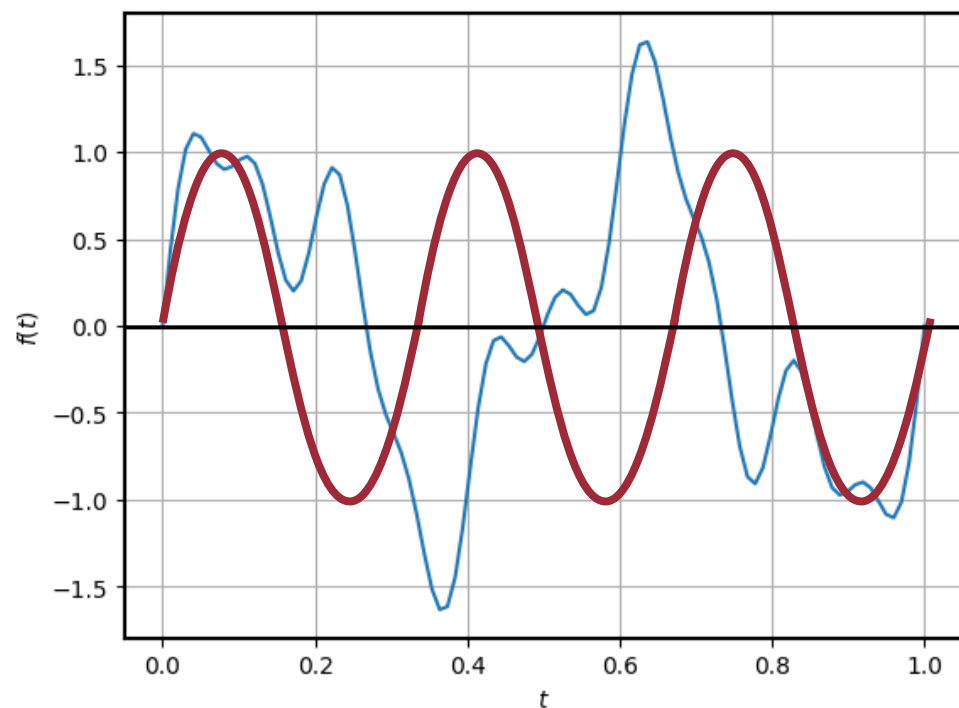
Candidate frequency components



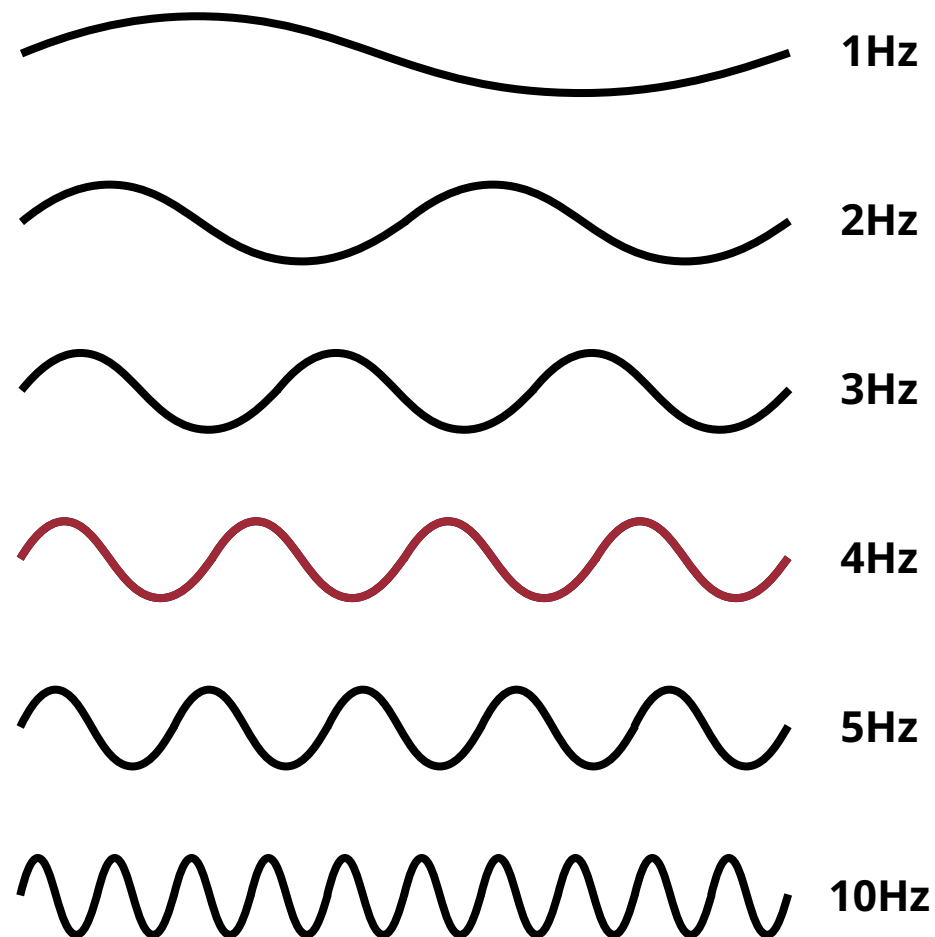
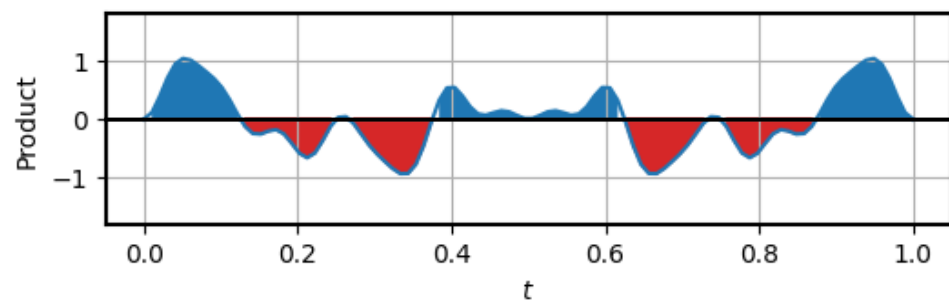
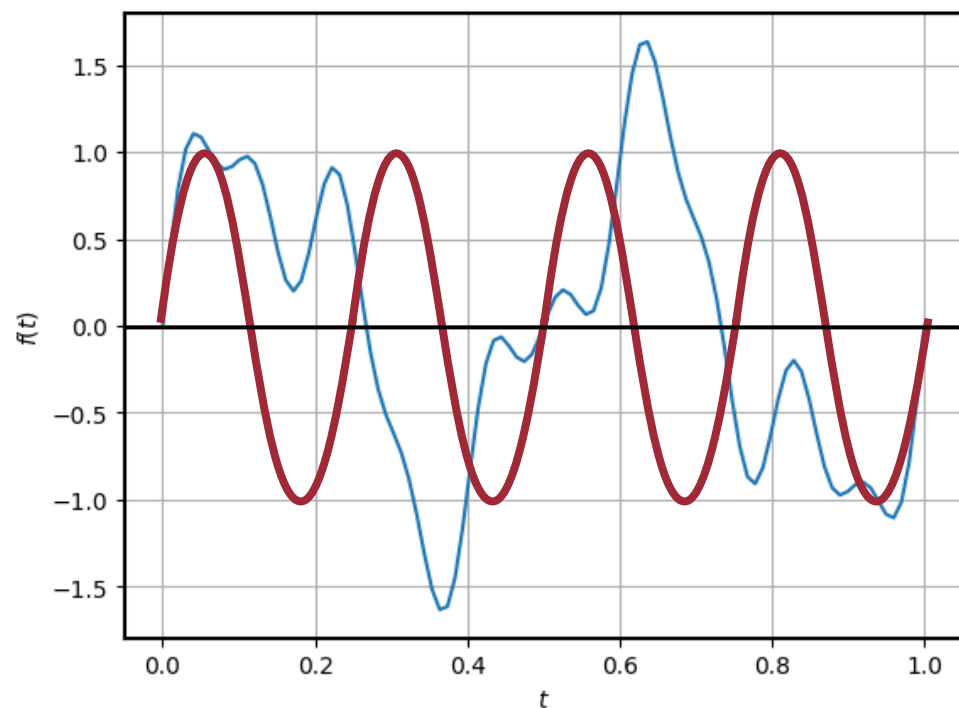
Demystifying Fourier Transform



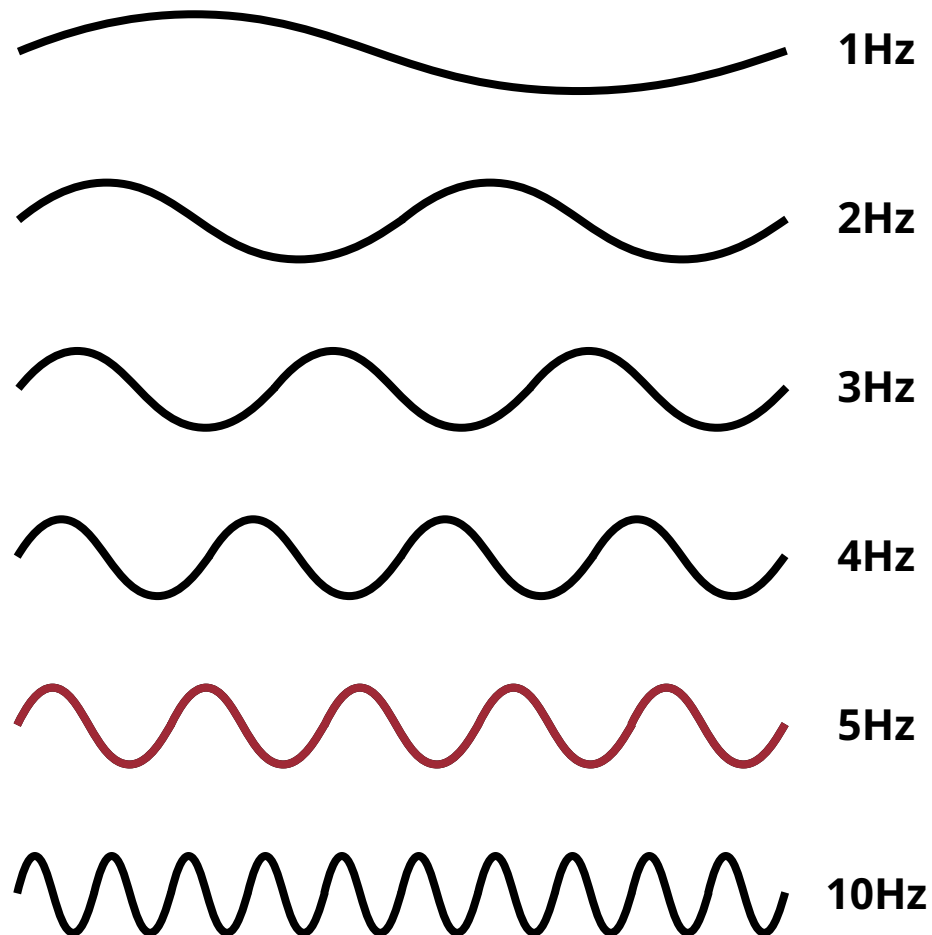
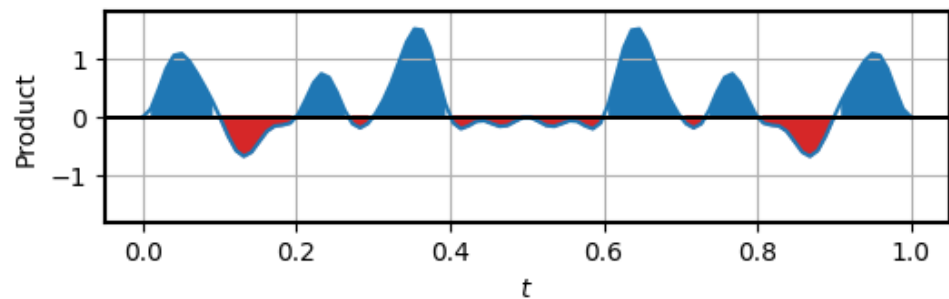
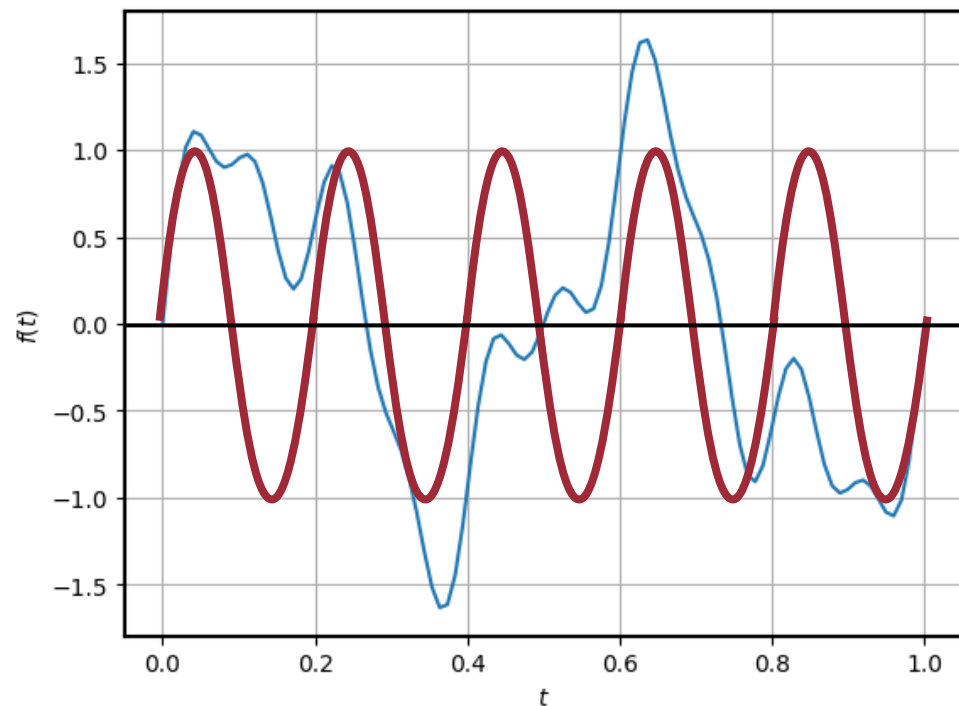
Demystifying Fourier Transform



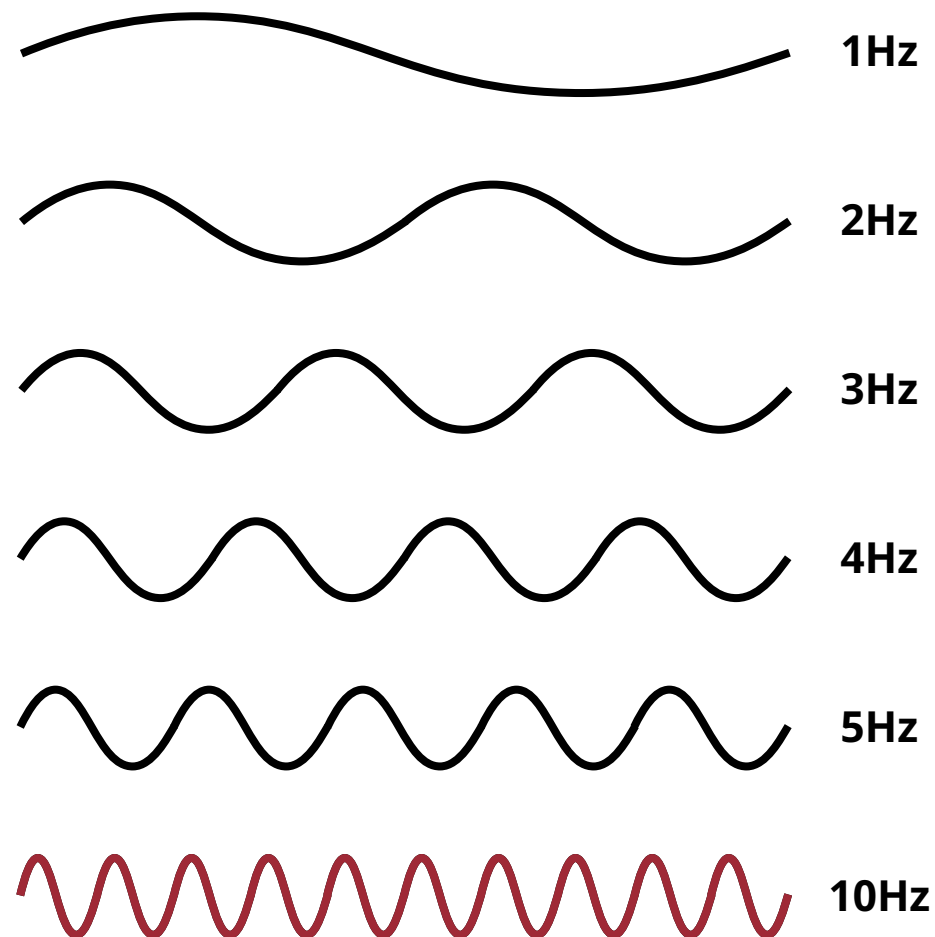
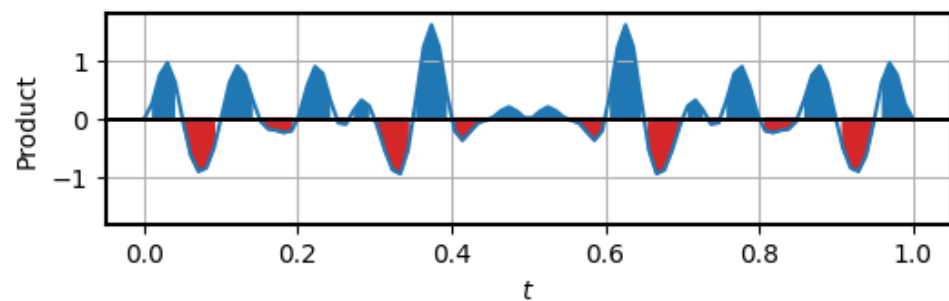
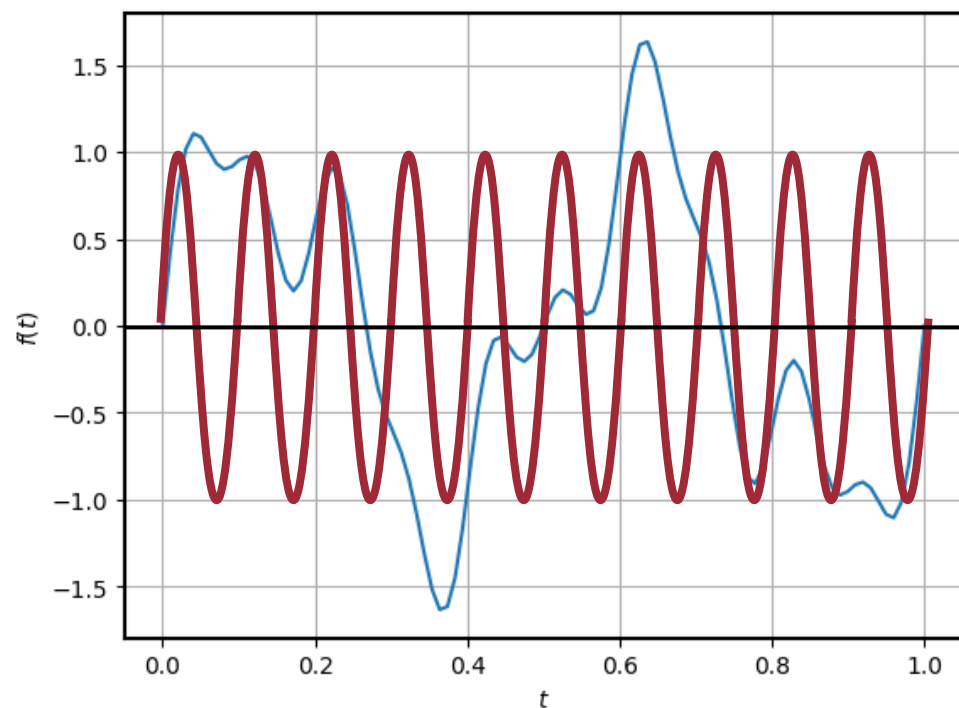
Demystifying Fourier Transform



Demystifying Fourier Transform



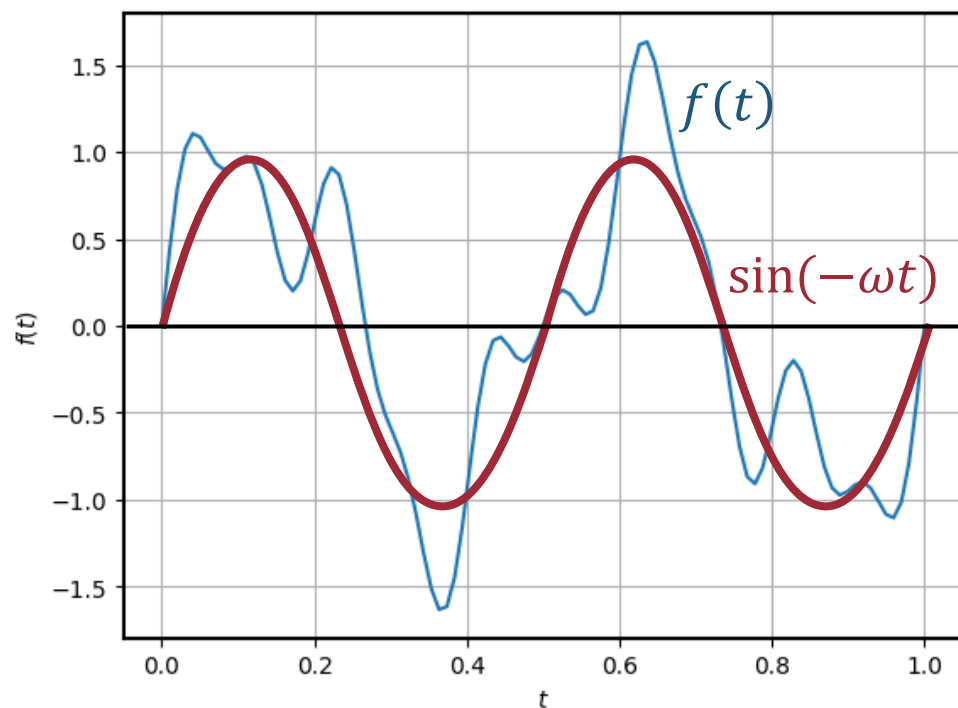
Demystifying Fourier Transform



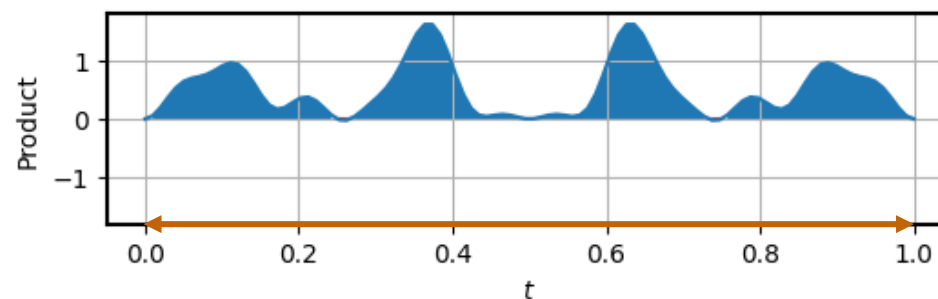
Demystifying Fourier Transform

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \cos(-\omega t) + j f(t) \sin(-\omega t) dt$$

Sum over all t



$f(t) \sin(-\omega t)$

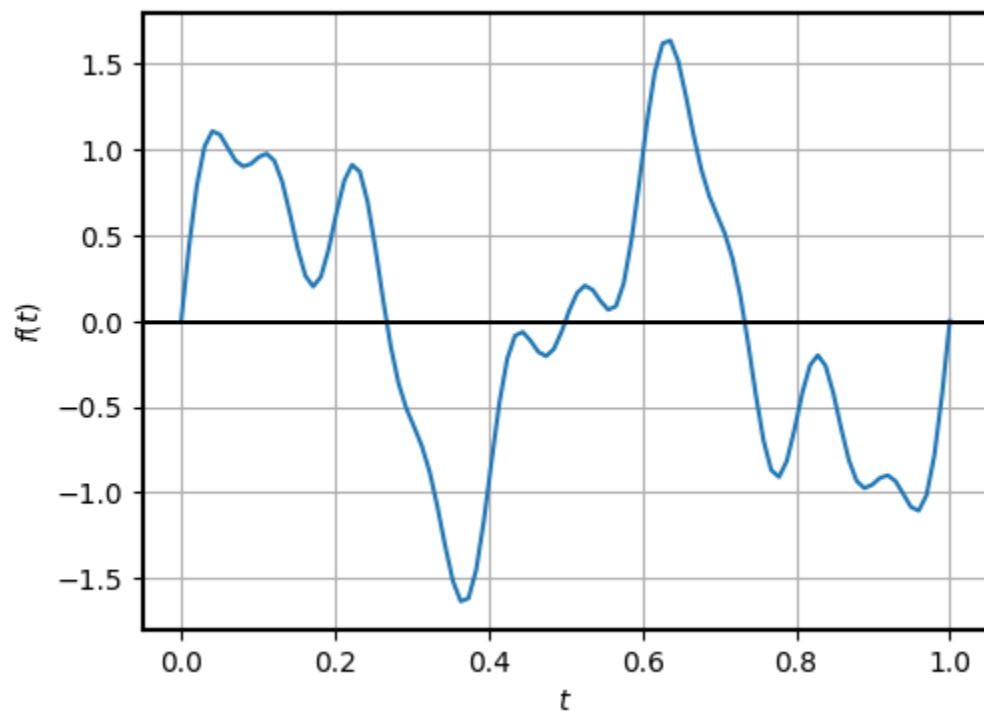


Sum = 0.495

Demystifying Fourier Transform

Signal

(time-domain)

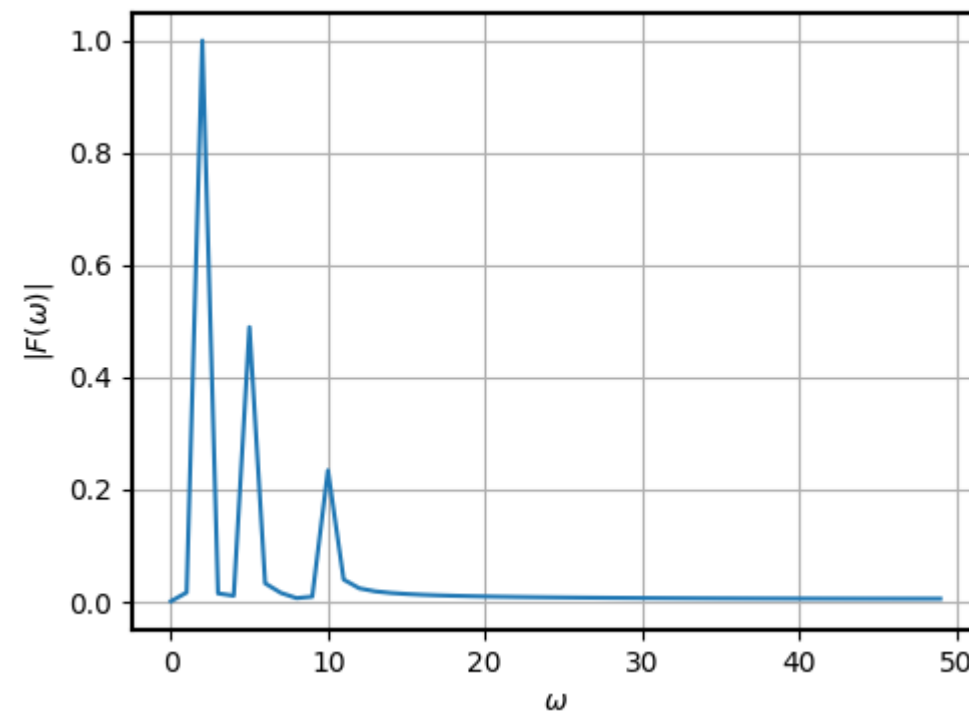


**Fourier
Transform**



Spectrum

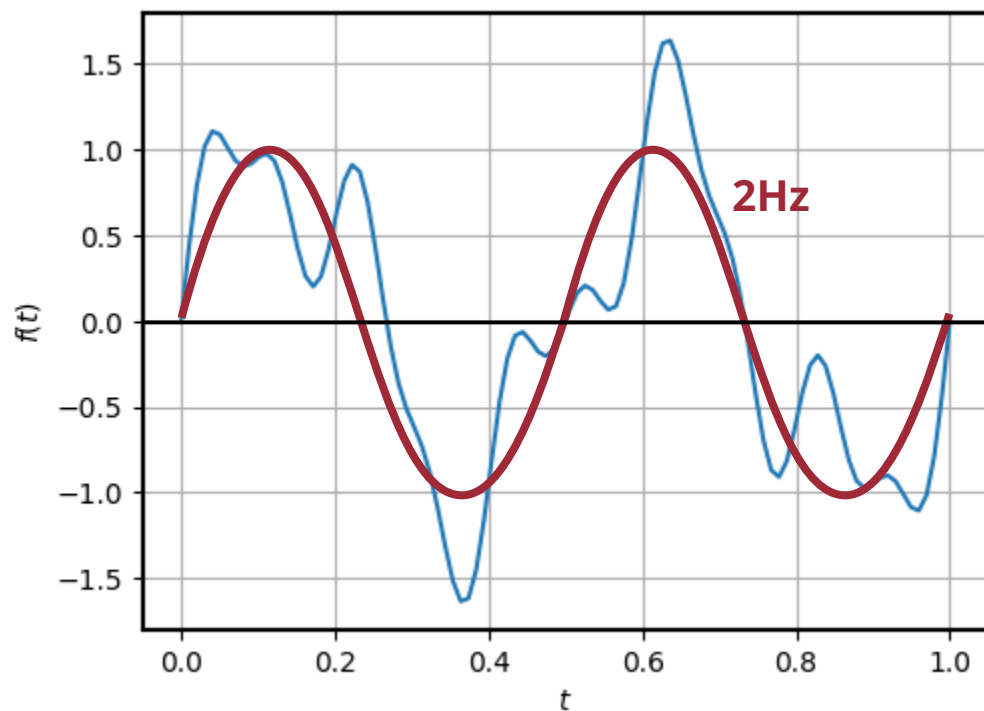
(frequency-domain)



Demystifying Fourier Transform

Signal

(time-domain)

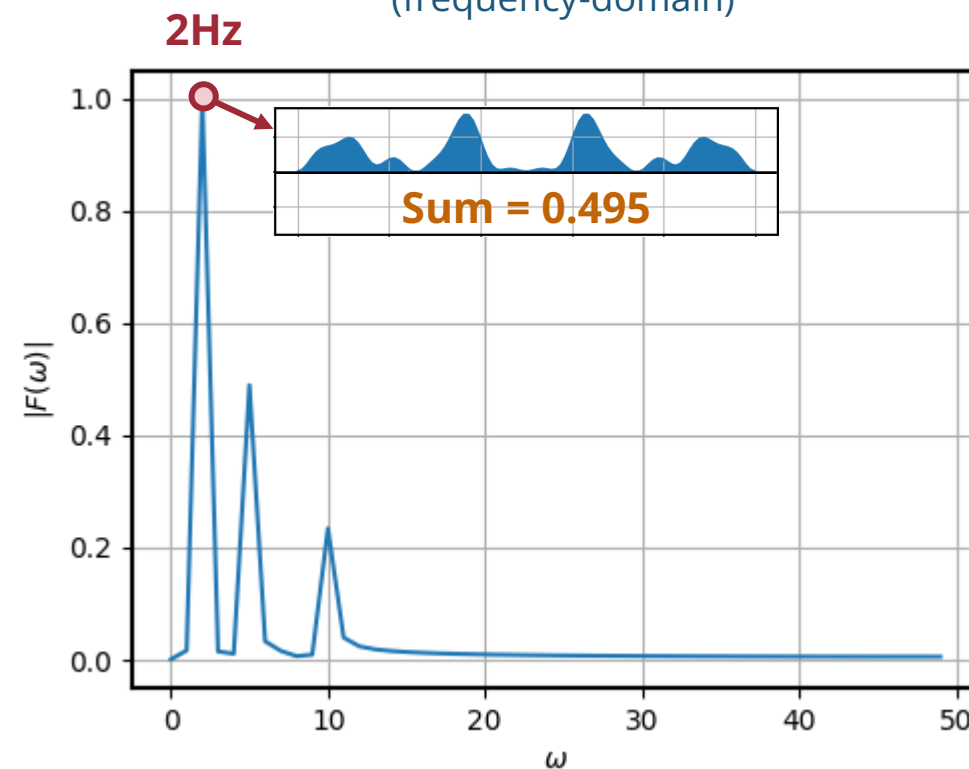


Fourier Transform



Spectrum

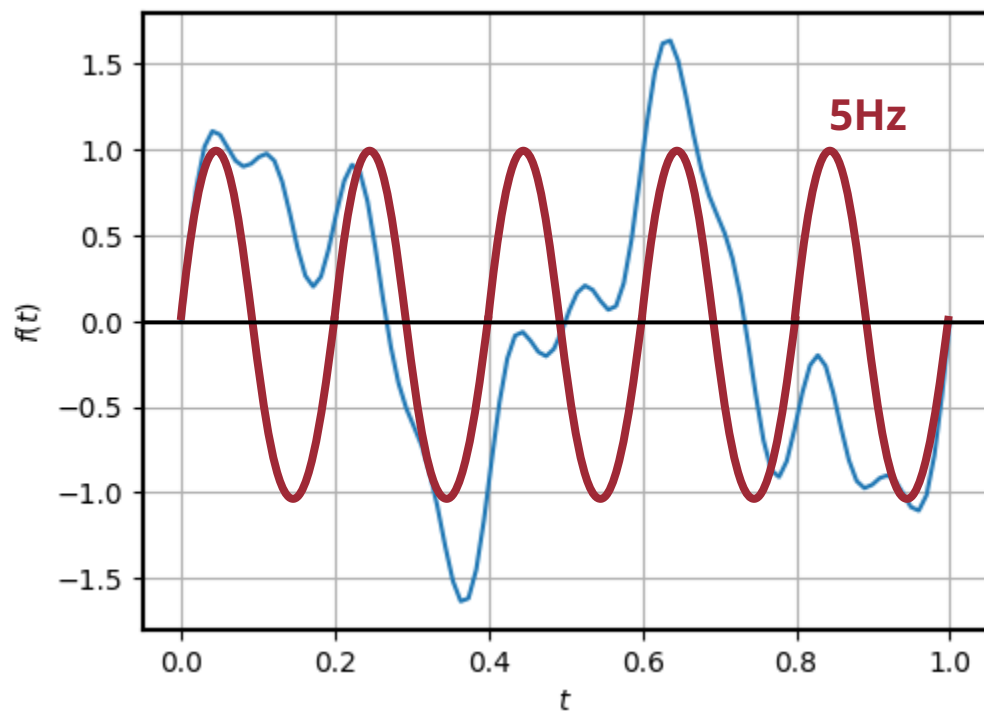
(frequency-domain)



Demystifying Fourier Transform

Signal

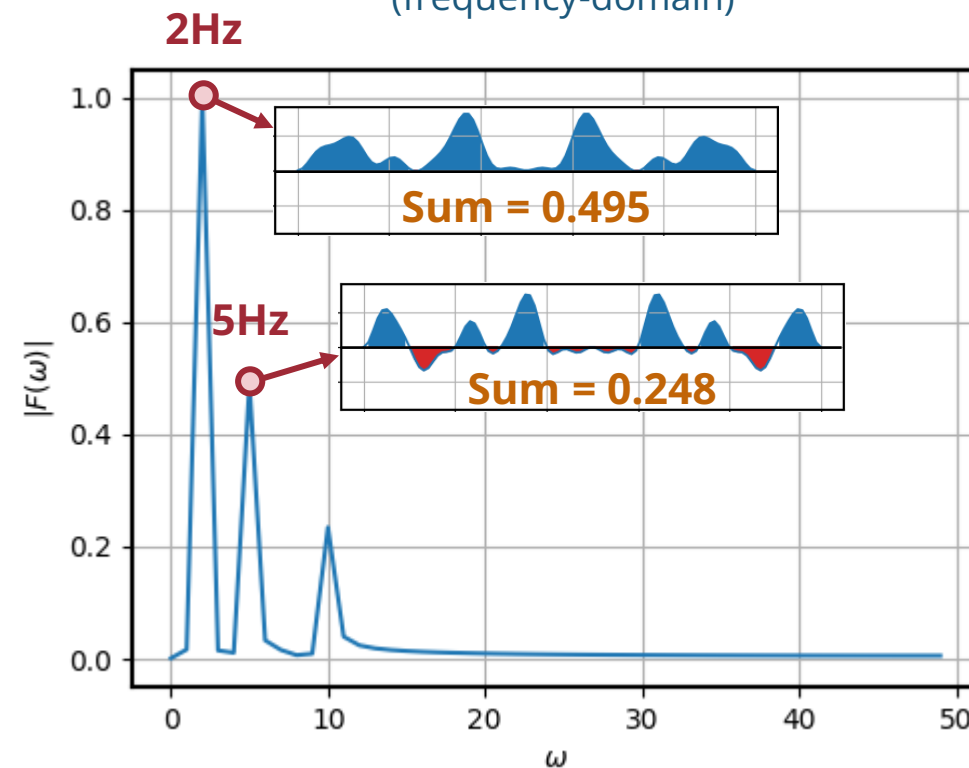
(time-domain)



Fourier Transform
→

Spectrum

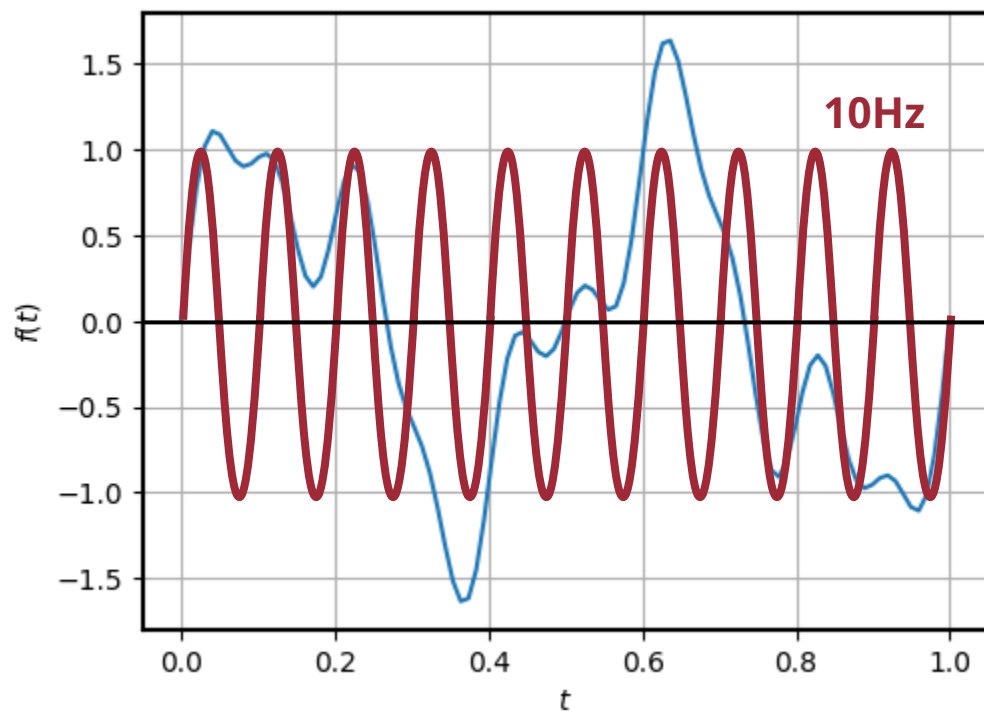
(frequency-domain)



Demystifying Fourier Transform

Signal

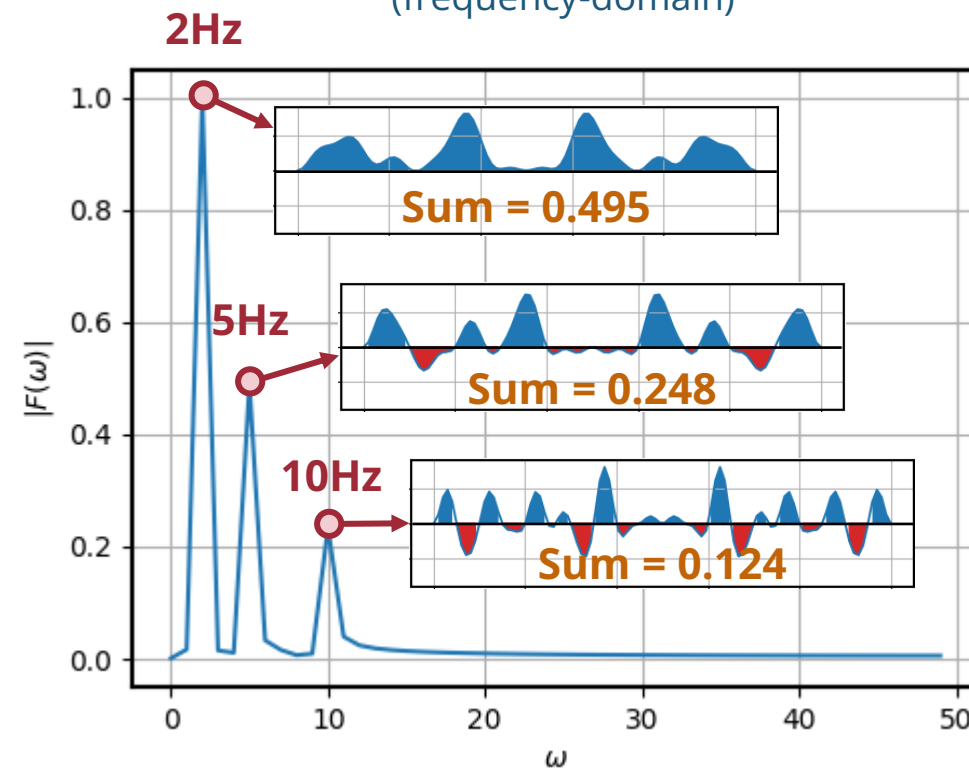
(time-domain)



Fourier Transform
→

Spectrum

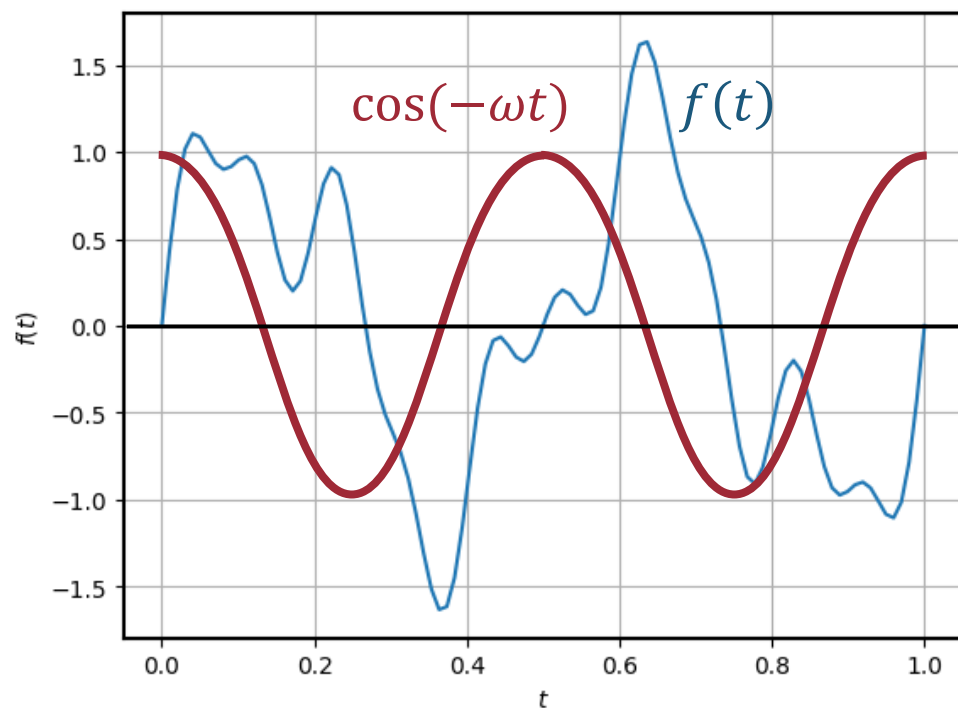
(frequency-domain)



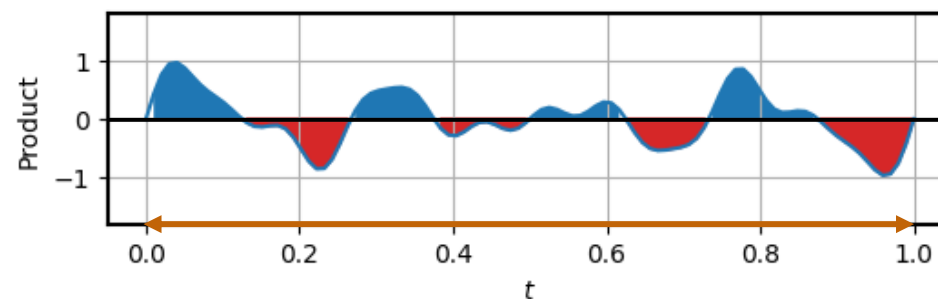
Demystifying Fourier Transform

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \cos(-\omega t) + j f(t) \sin(-\omega t) dt$$

Sum over all t

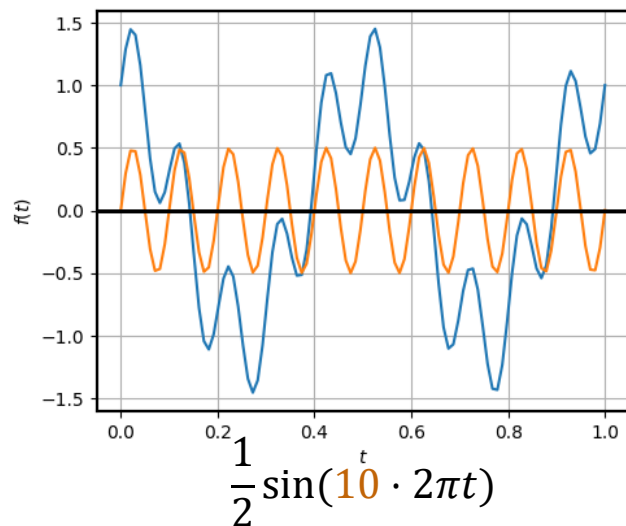
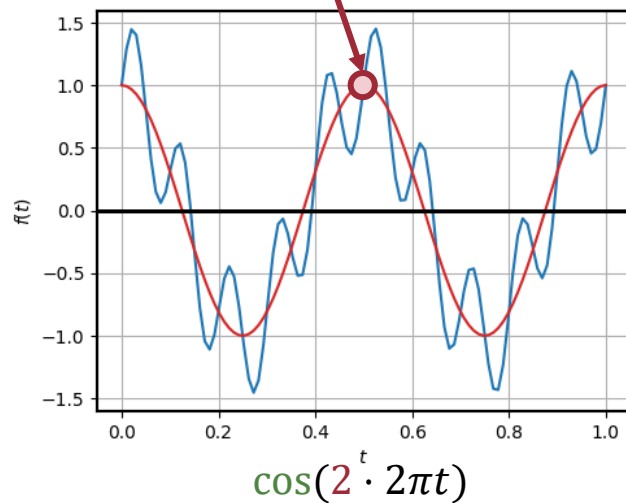
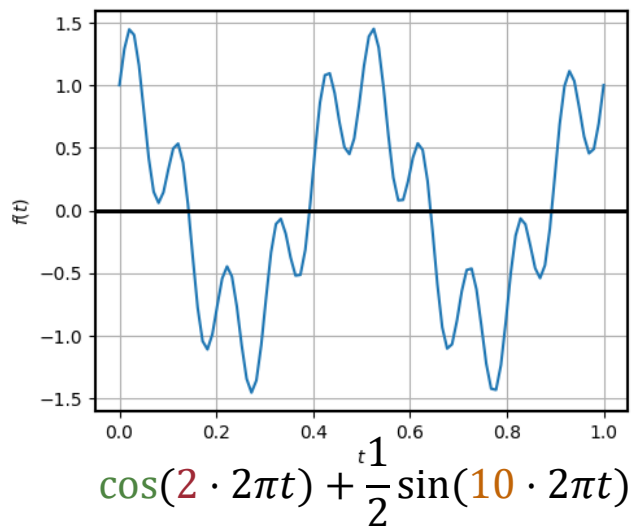
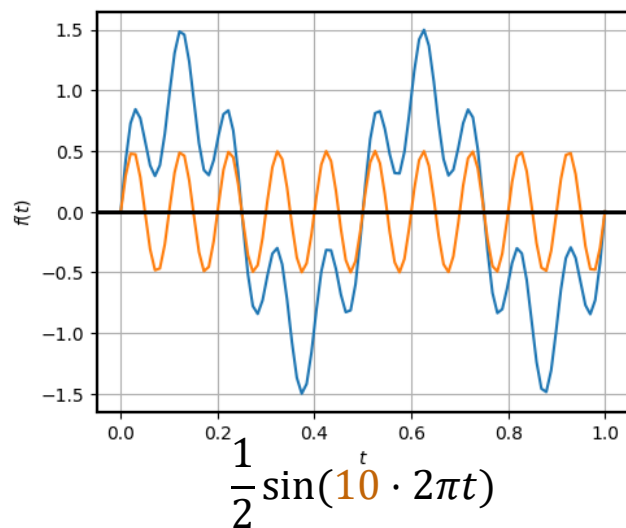
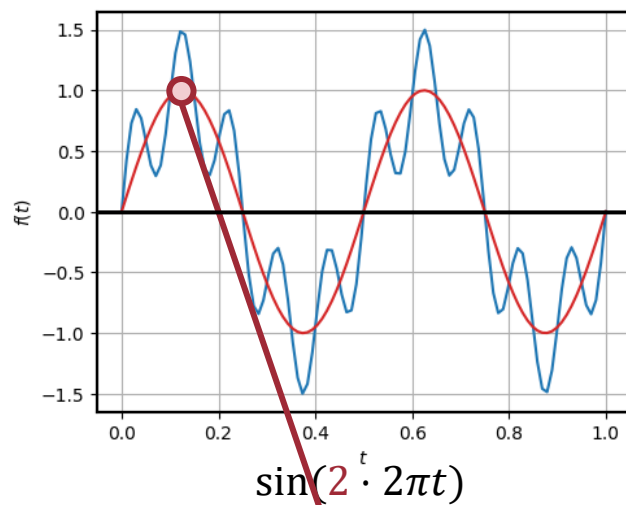
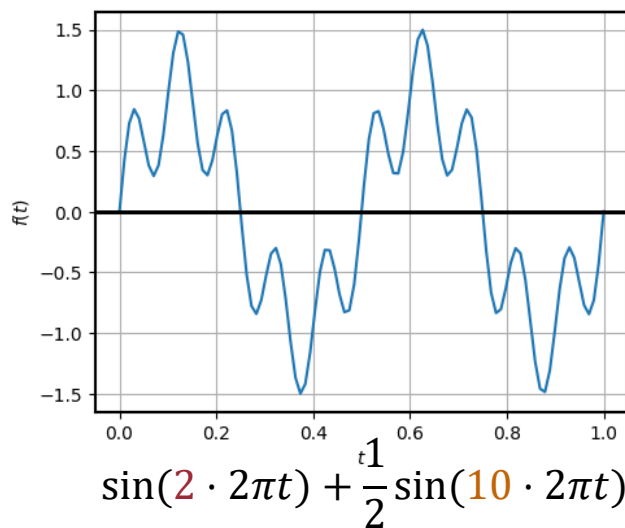


$f(t) \cos(-\omega t)$



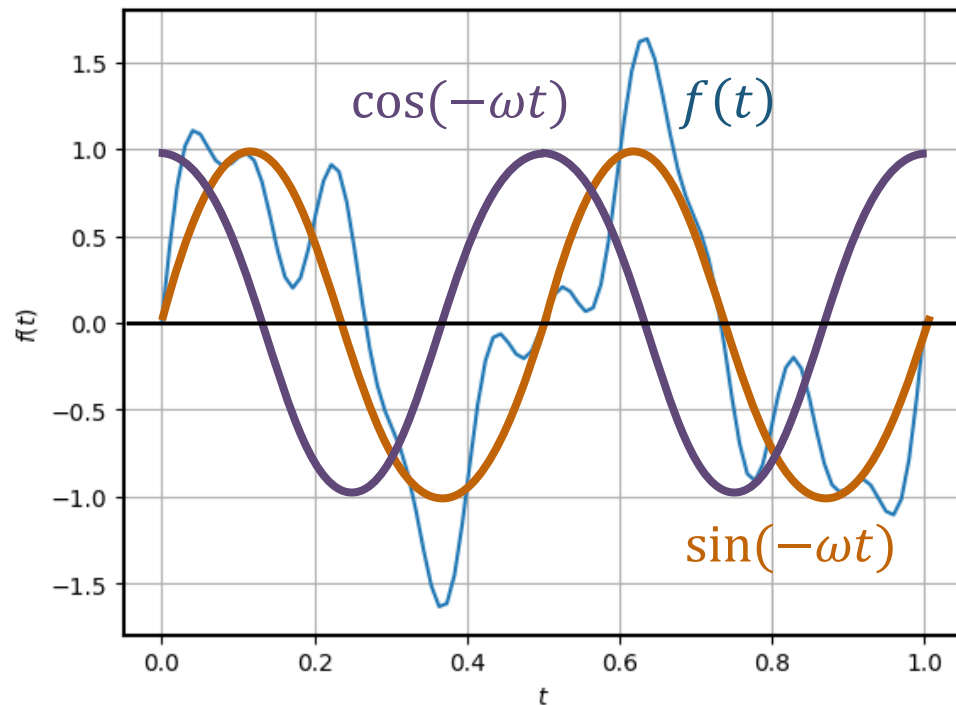
Sum = 0

Phase of a Signal



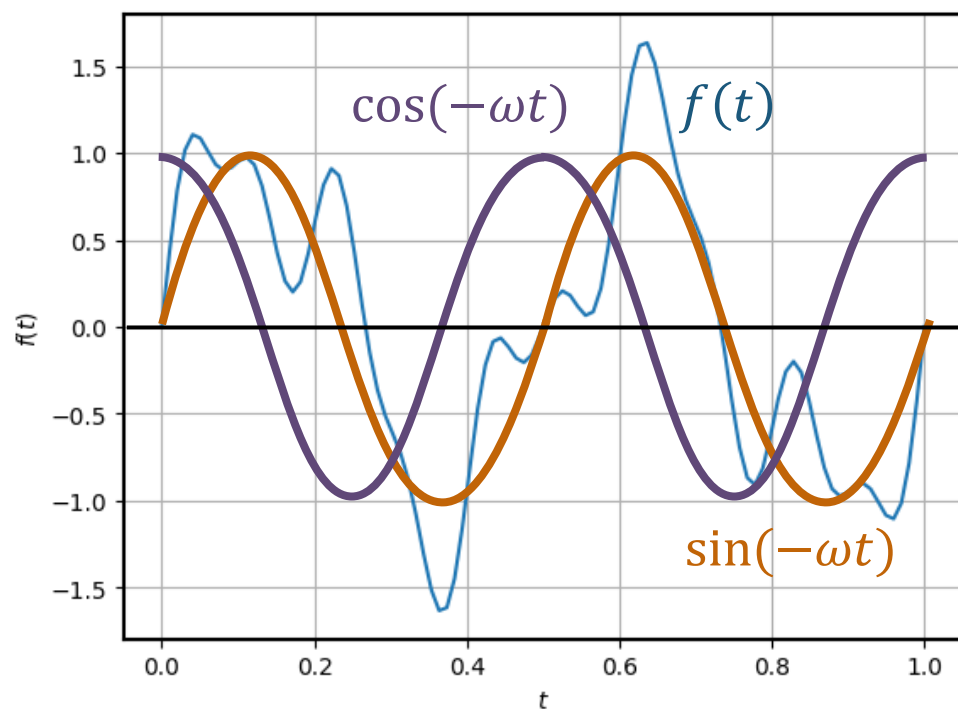
Demystifying Fourier Transform

$$F(\omega) = \int_{-\infty}^{\infty} \underbrace{f(t) \cos(-\omega t)}_{\text{Real part}} + \underbrace{j f(t) \sin(-\omega t)}_{\text{Imaginary part}} dt$$



Demystifying Fourier Transform

$$F(\omega) = \underbrace{\int_{-\infty}^{\infty} f(t) \cos(-\omega t) dt}_{\text{Real part}} + j \underbrace{\int_{-\infty}^{\infty} f(t) \sin(-\omega t) dt}_{\text{Imaginary part}}$$

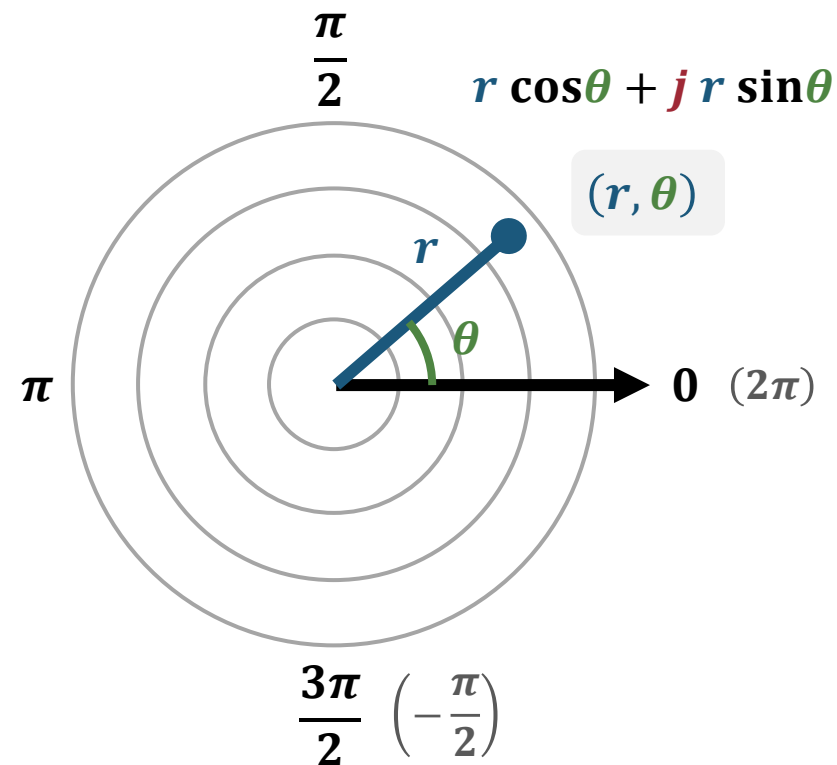
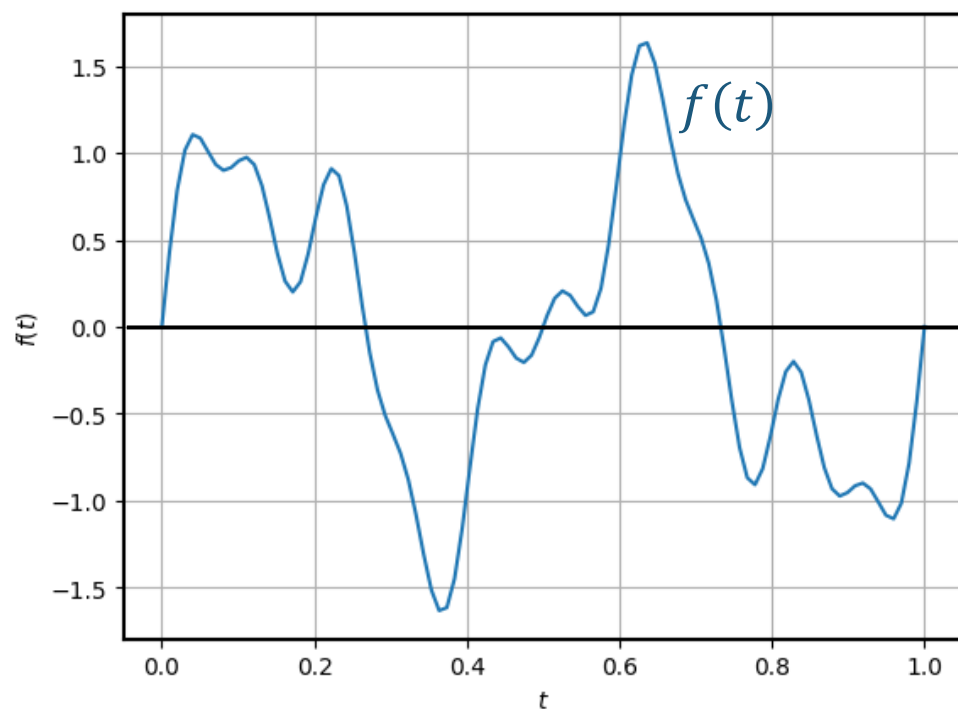


Demystifying Fourier Transform

Real part

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \cos(-\omega t) dt + j \int_{-\infty}^{\infty} f(t) \sin(-\omega t) dt$$

Imaginary part

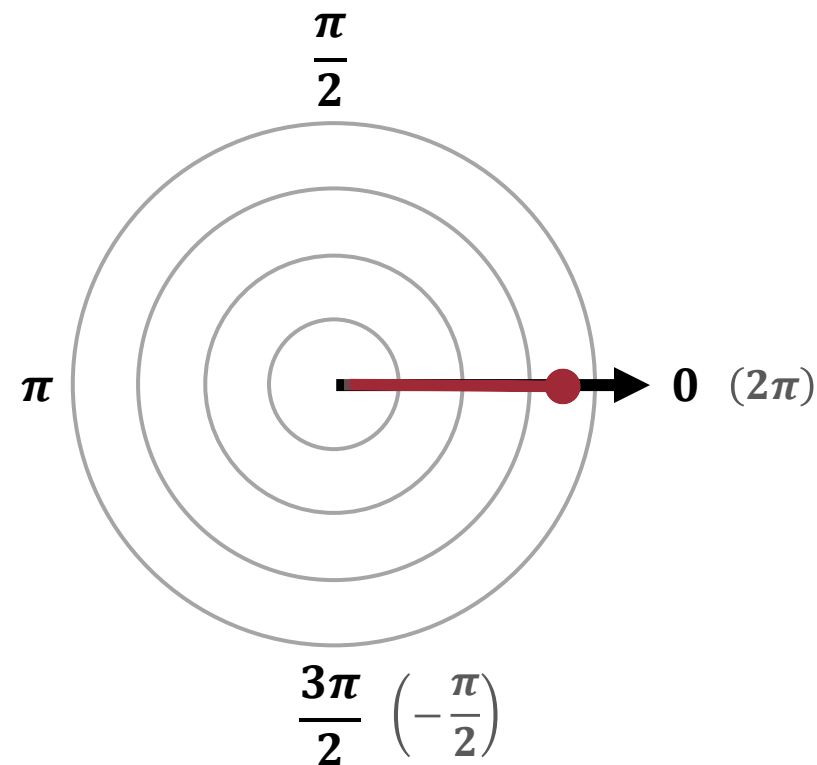
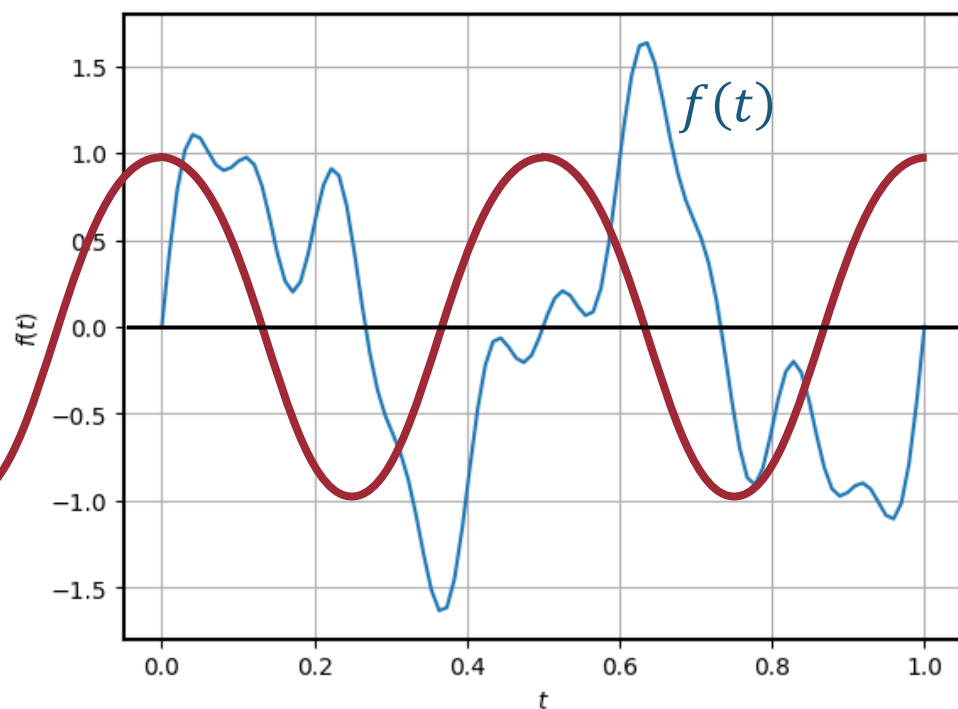


Demystifying Fourier Transform

Real part

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \cos(-\omega t) dt + j \int_{-\infty}^{\infty} f(t) \sin(-\omega t) dt$$

Imaginary part

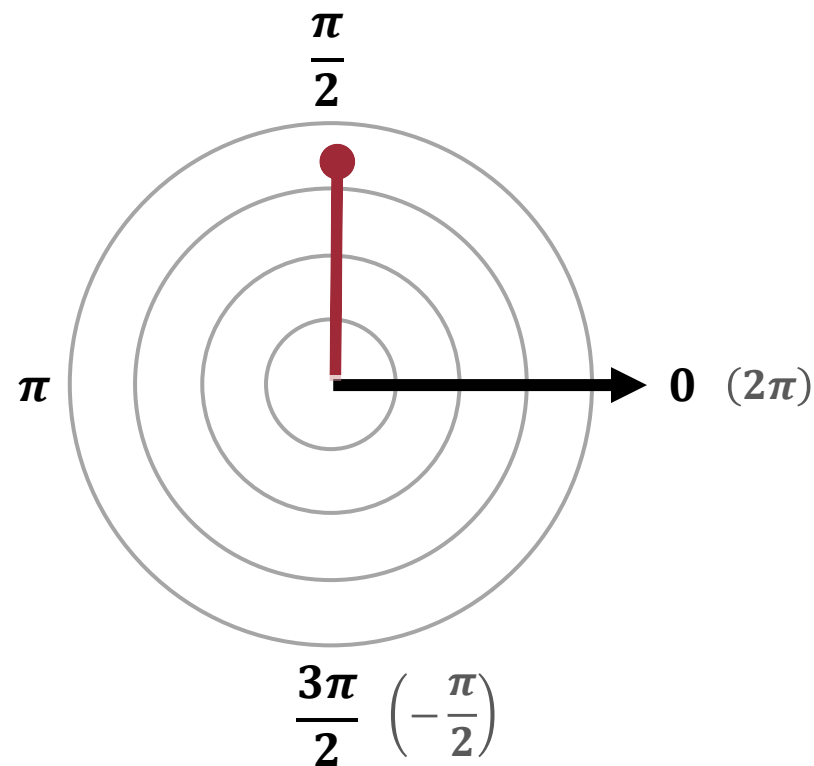
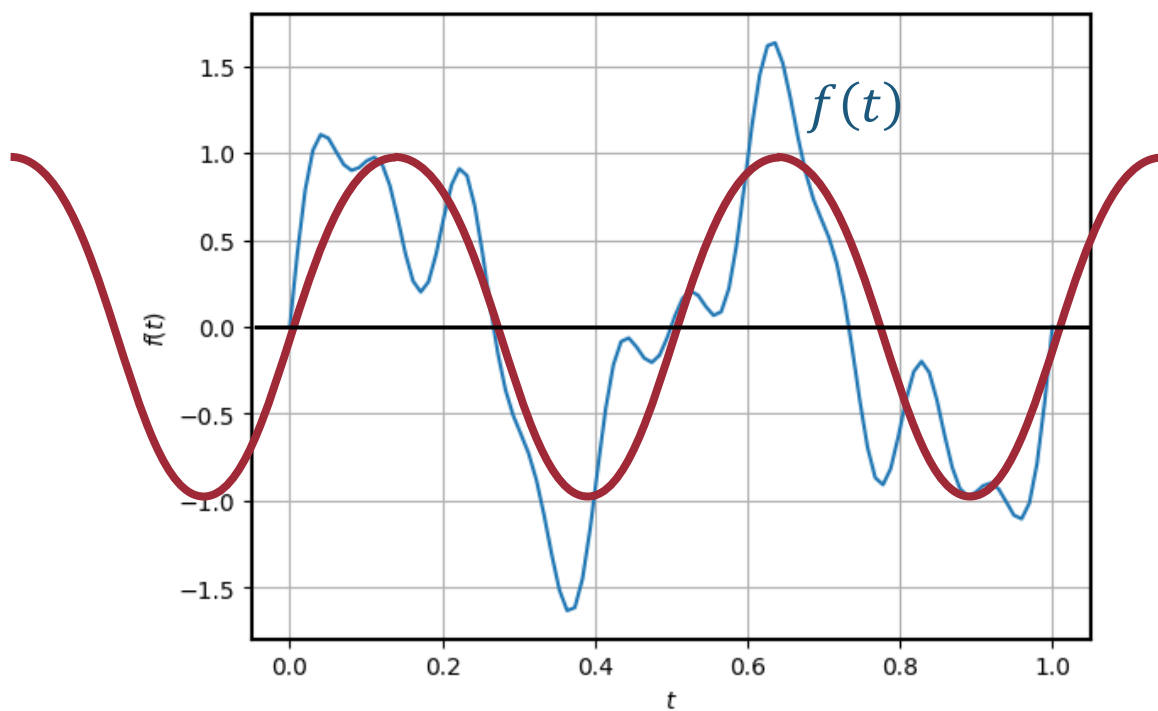


Demystifying Fourier Transform

Real part

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \cos(-\omega t) dt + j \int_{-\infty}^{\infty} f(t) \sin(-\omega t) dt$$

Imaginary part

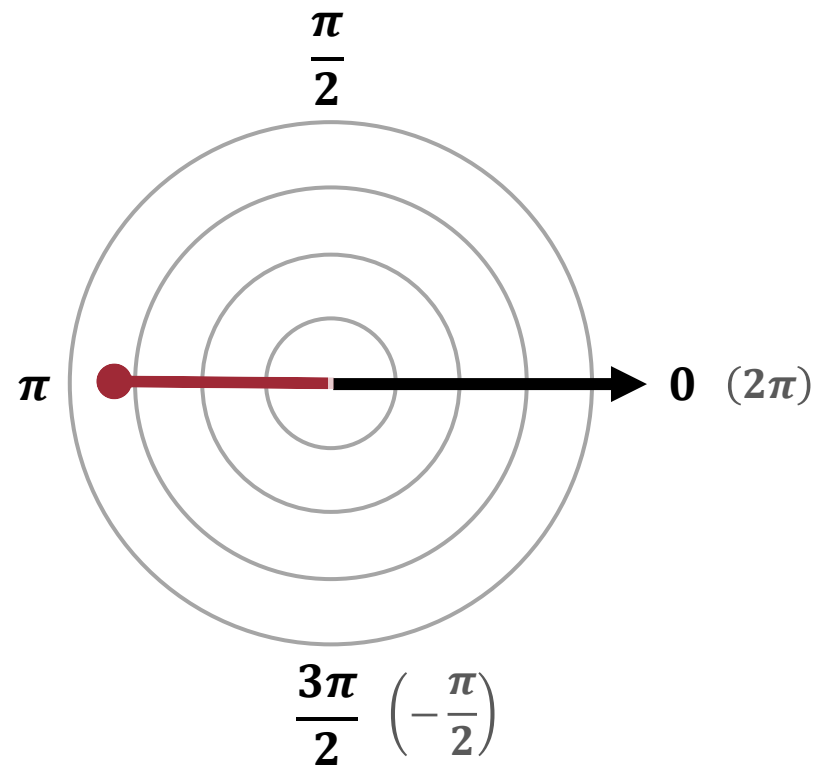
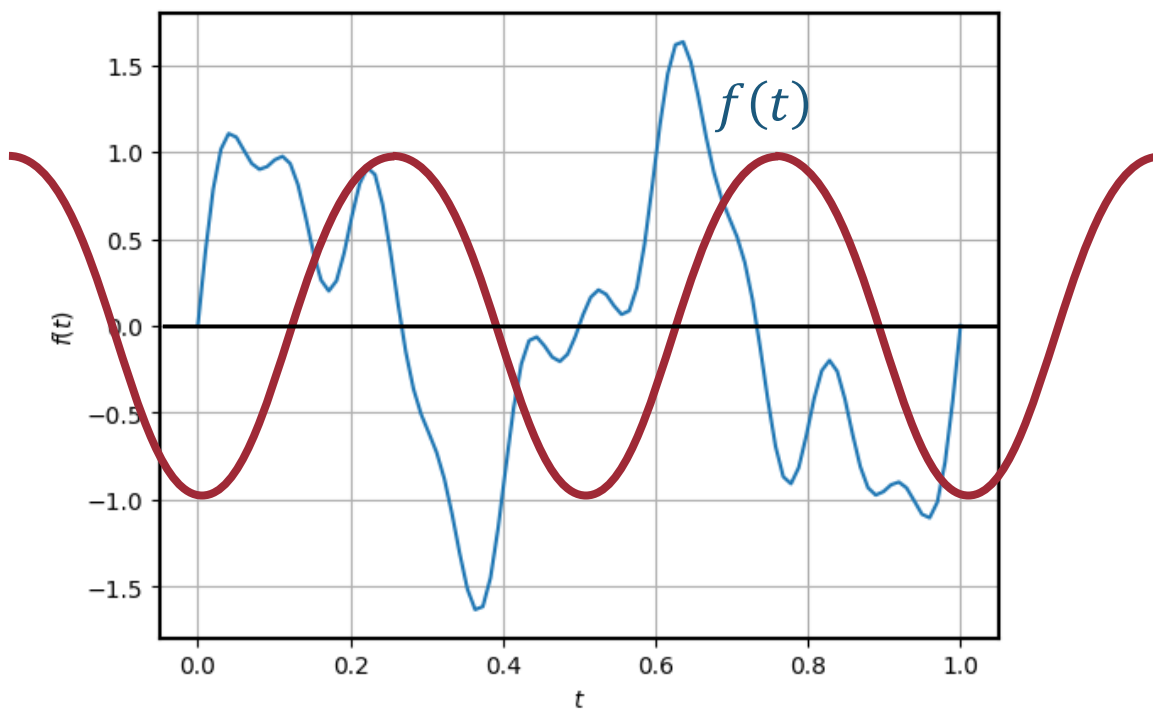


Demystifying Fourier Transform

Real part

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \cos(-\omega t) dt + j \int_{-\infty}^{\infty} f(t) \sin(-\omega t) dt$$

Imaginary part

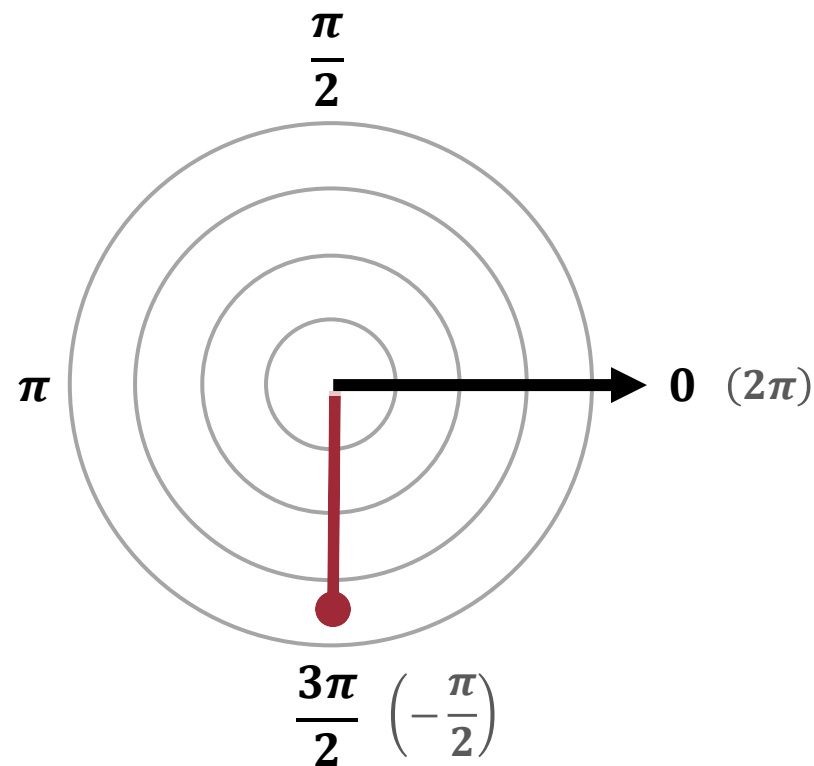
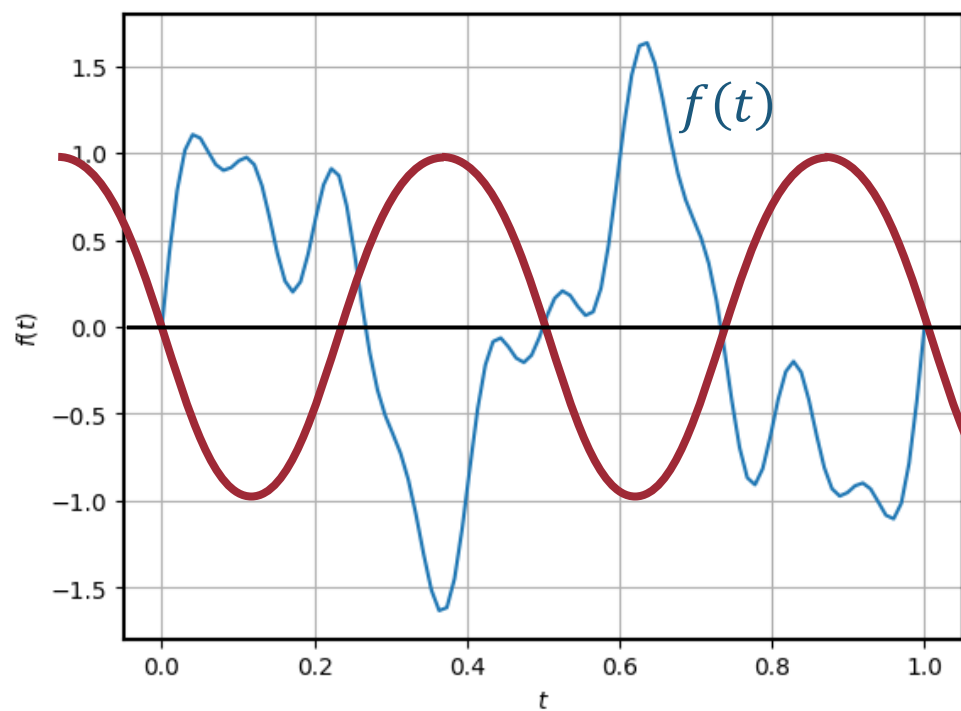


Demystifying Fourier Transform

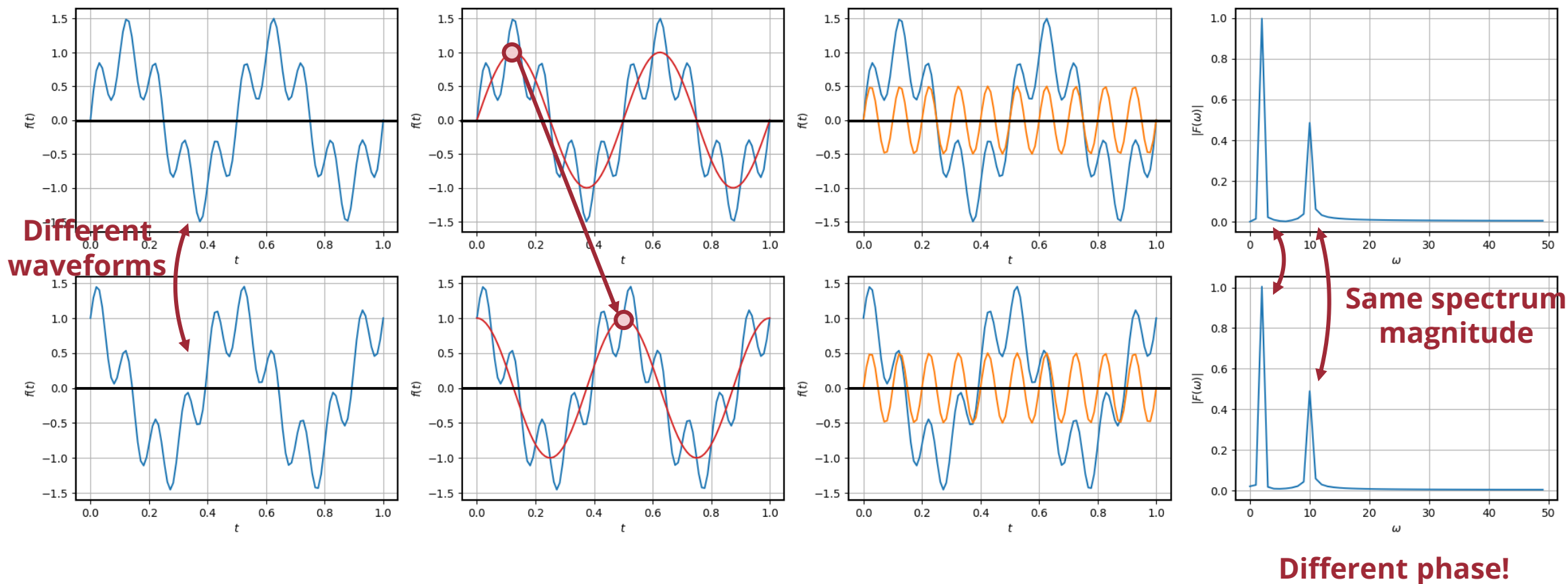
Real part

$$F(\omega) = \int_{-\infty}^{\infty} f(t) \cos(-\omega t) dt + j \int_{-\infty}^{\infty} f(t) \sin(-\omega t) dt$$

Imaginary part

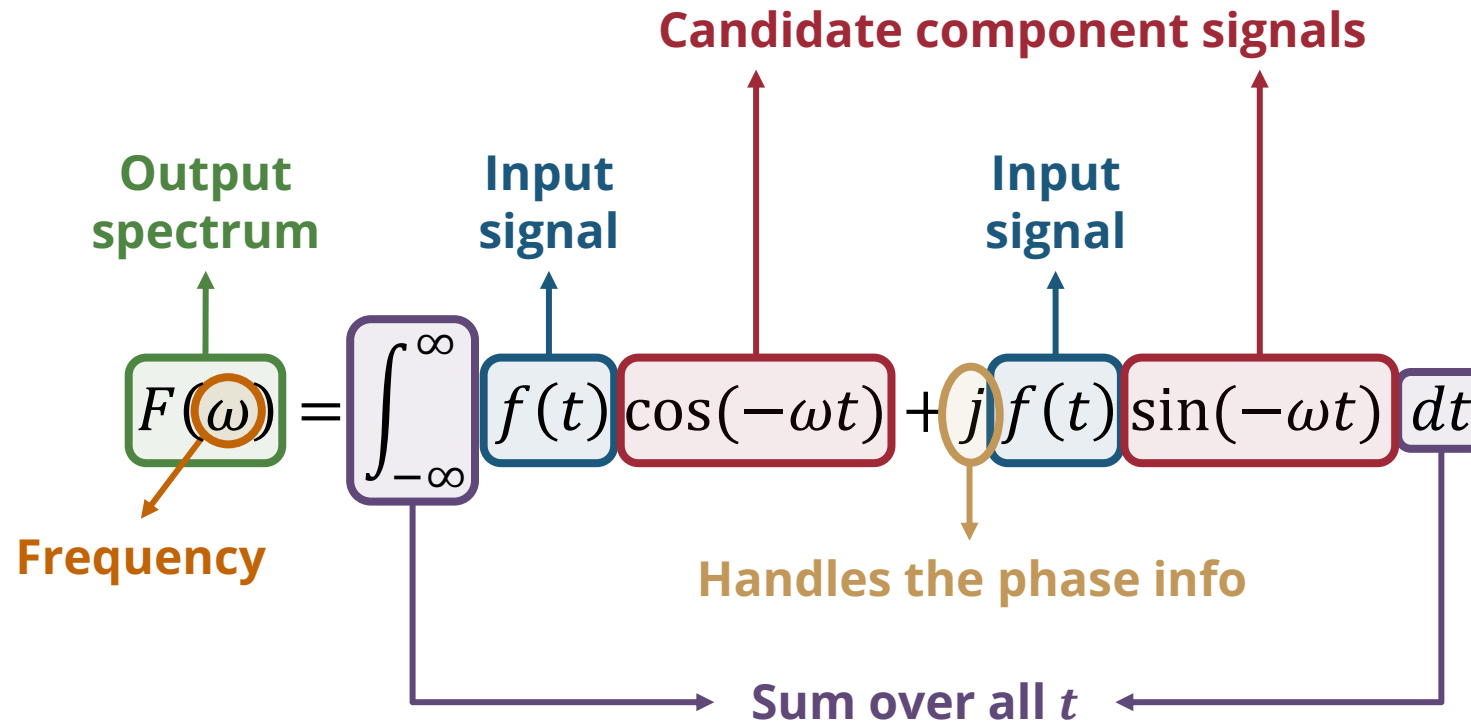


Magnitude & Phase



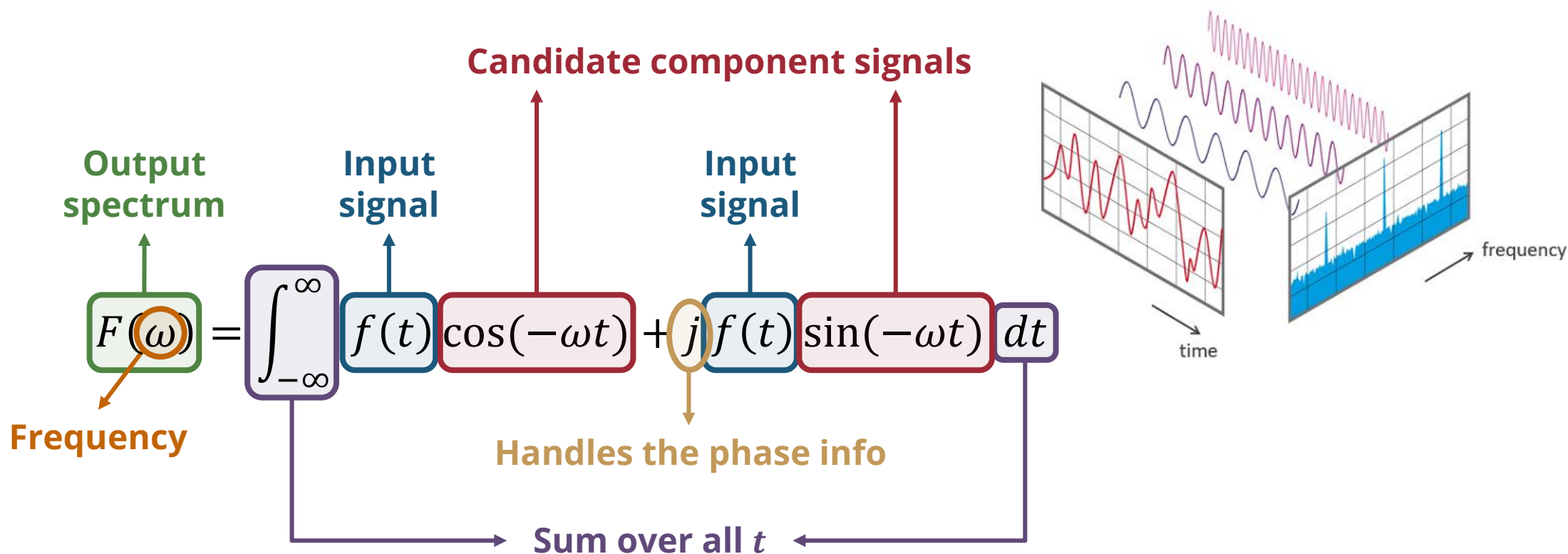
Fourier Transform

- **Intuition:** Decompose time-domain signals into **frequency components**
- Math formulation:



Fourier Transform

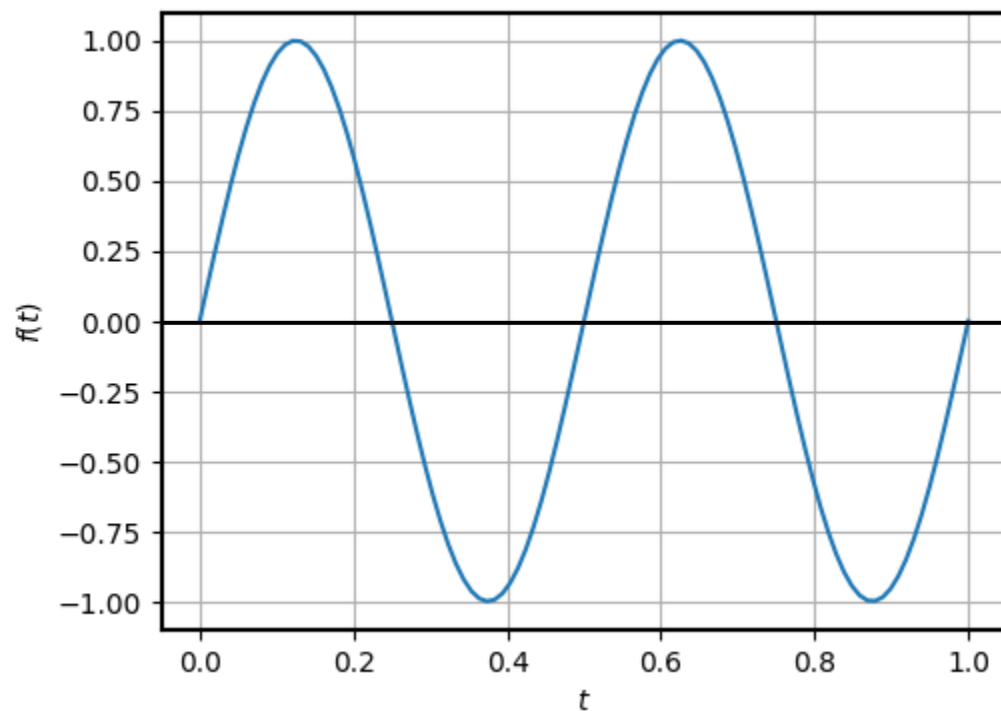
- **Intuition:** Decompose time-domain signals into **frequency components**
- Math formulation:



Example: A 2Hz Sine Wave

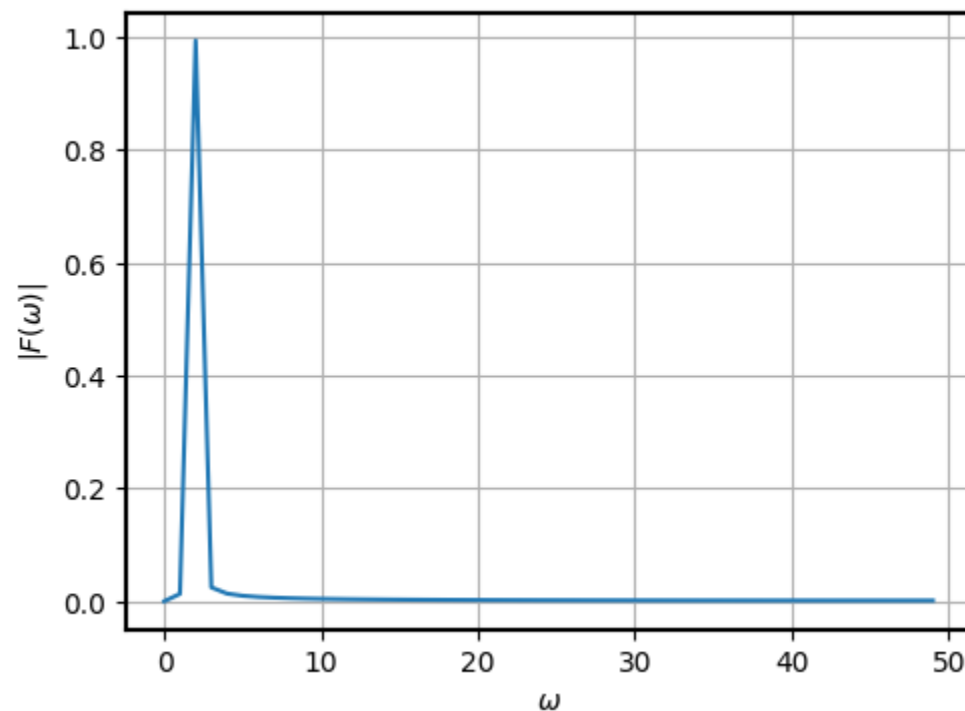
Signal

(time-domain)



Spectrum

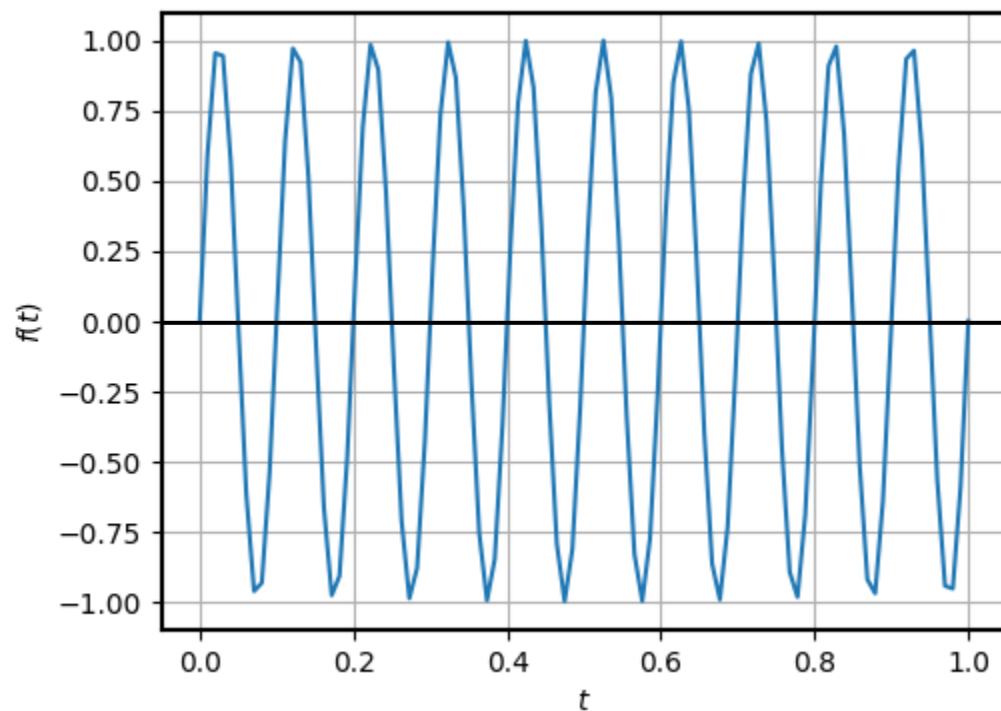
(frequency-domain)



Example: A 10Hz Sine Wave

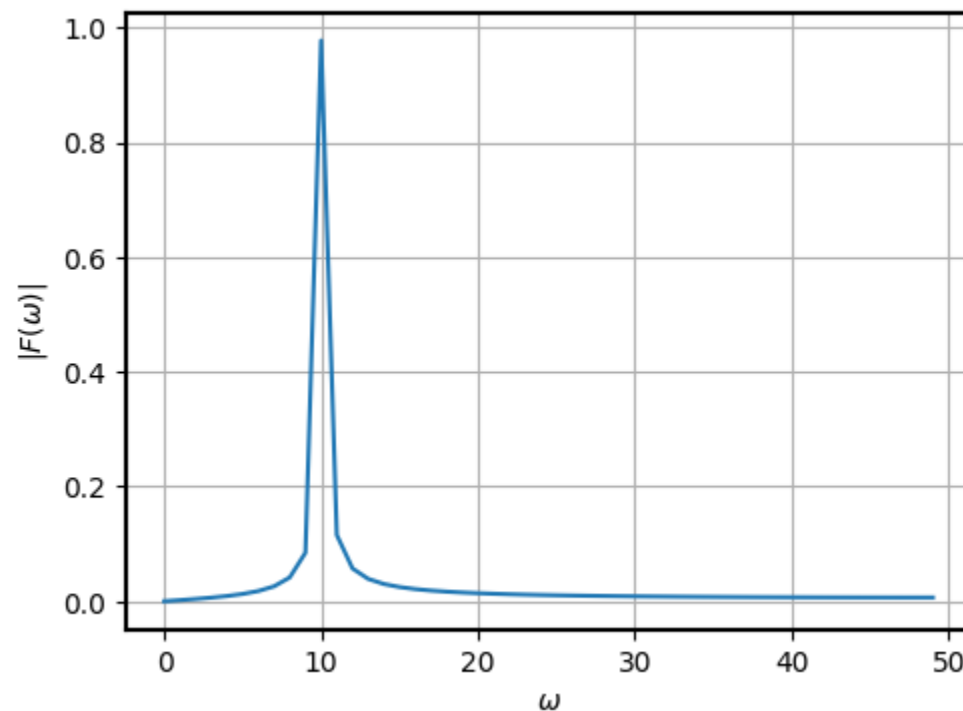
Signal

(time-domain)



Spectrum

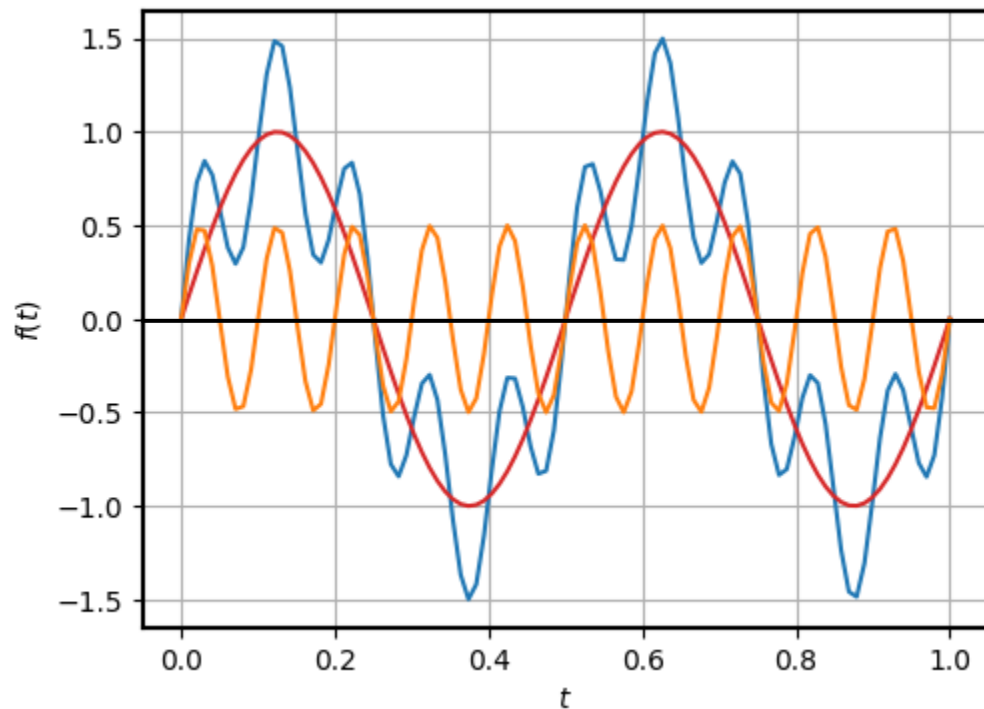
(frequency-domain)



Example: Sum of 2Hz & 10Hz Sine Waves

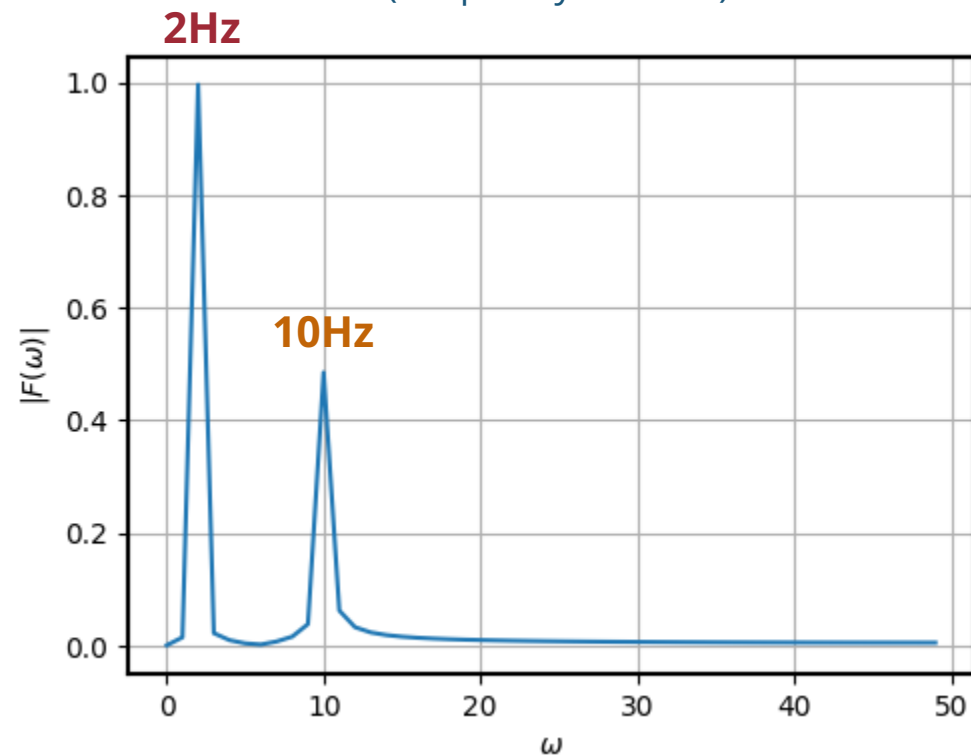
Signal

(time-domain)



Spectrum

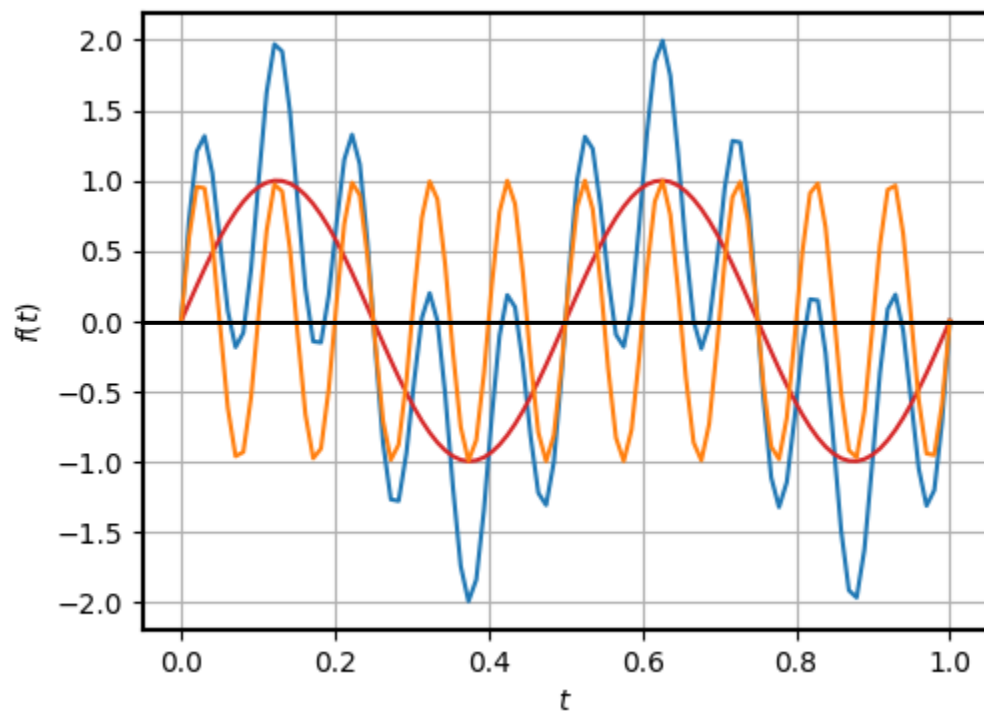
(frequency-domain)



Example: Sum of 2Hz & 10Hz Sine Waves

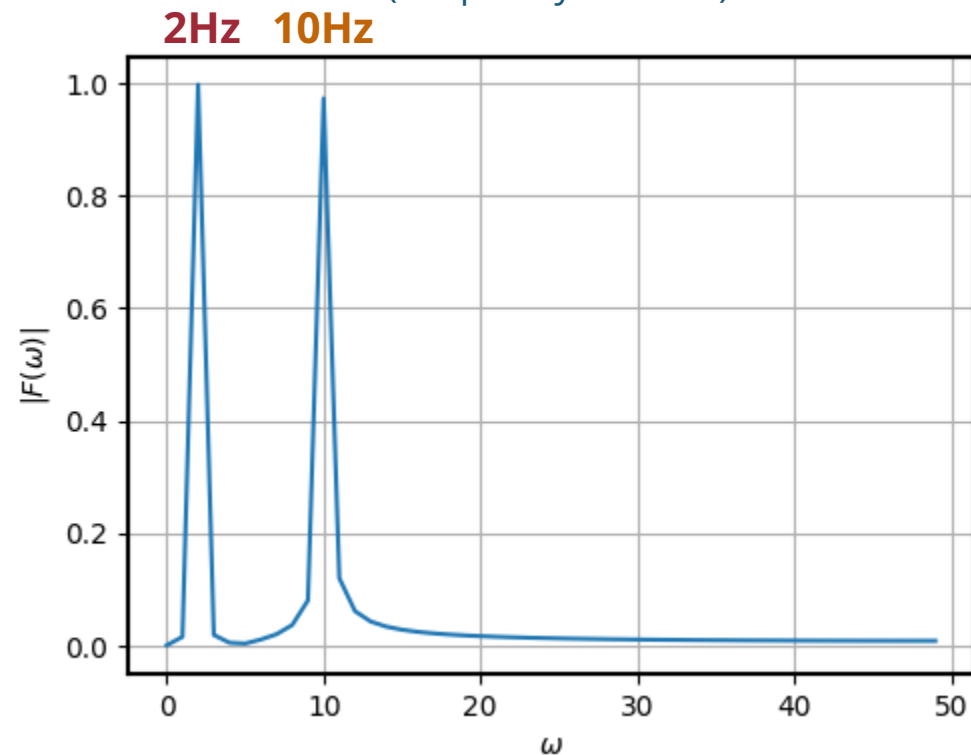
Signal

(time-domain)



Spectrum

(frequency-domain)



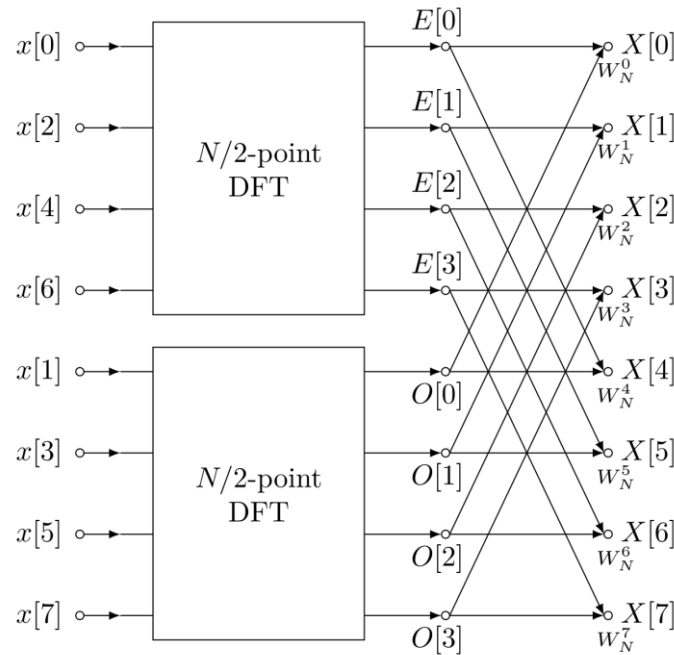
Discrete Fourier Transform (DFT)

- **Intuition:** Fourier transform with **discrete time and frequency**
 - Used for **digital audio** → we cannot achieve an infinite sampling rate...
- Math formulation:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-j2\pi \frac{k}{N}n}$$

In Practice: Fast Fourier Transform (FFT)

- An efficient implementation of discrete Fourier transform
 - Reduce the complexity from $O(n^2)$ to $O(n \log n)$



(Source: Yangwenbo99)

Top 10 algorithms from the 20th century

Computing
in **SCIENCE & ENGINEERING**



IEEE

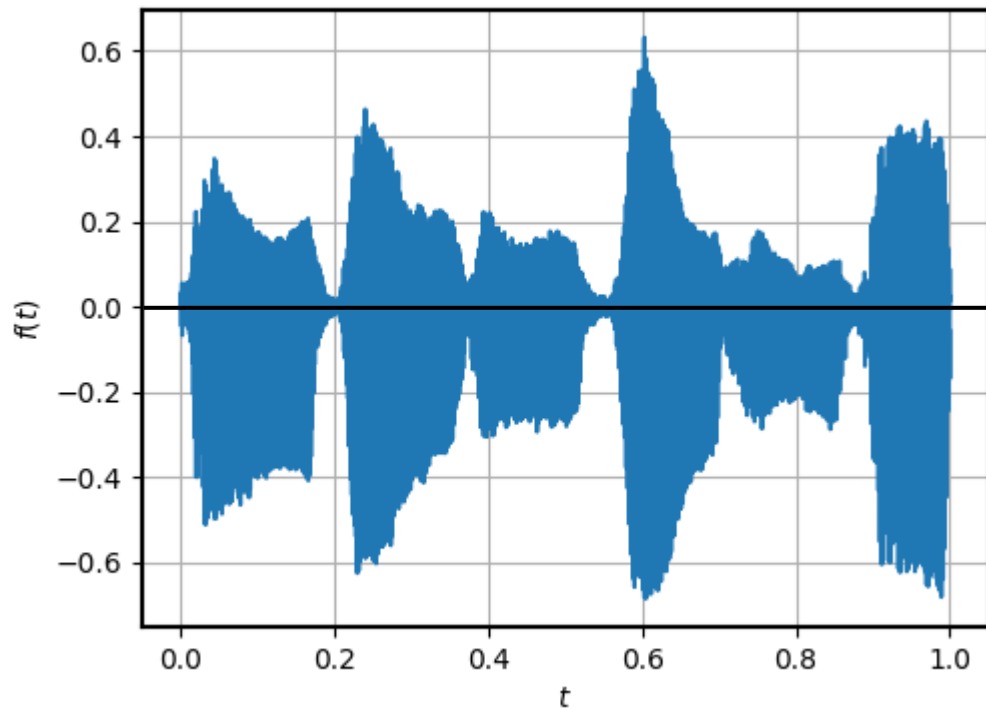
IEEE COMPUTER
SOCIETY
www.computer.org/ieee

Time-Frequency Analysis

Fourier Transform of a Trumpet Sound

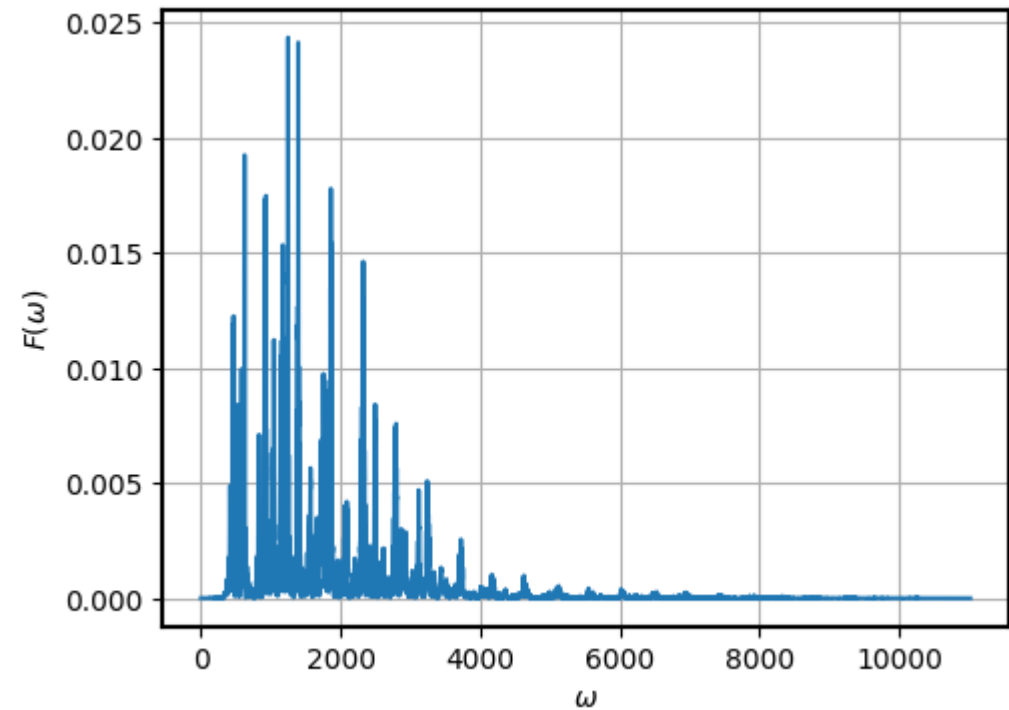
Signal

(time-domain)



Spectrum

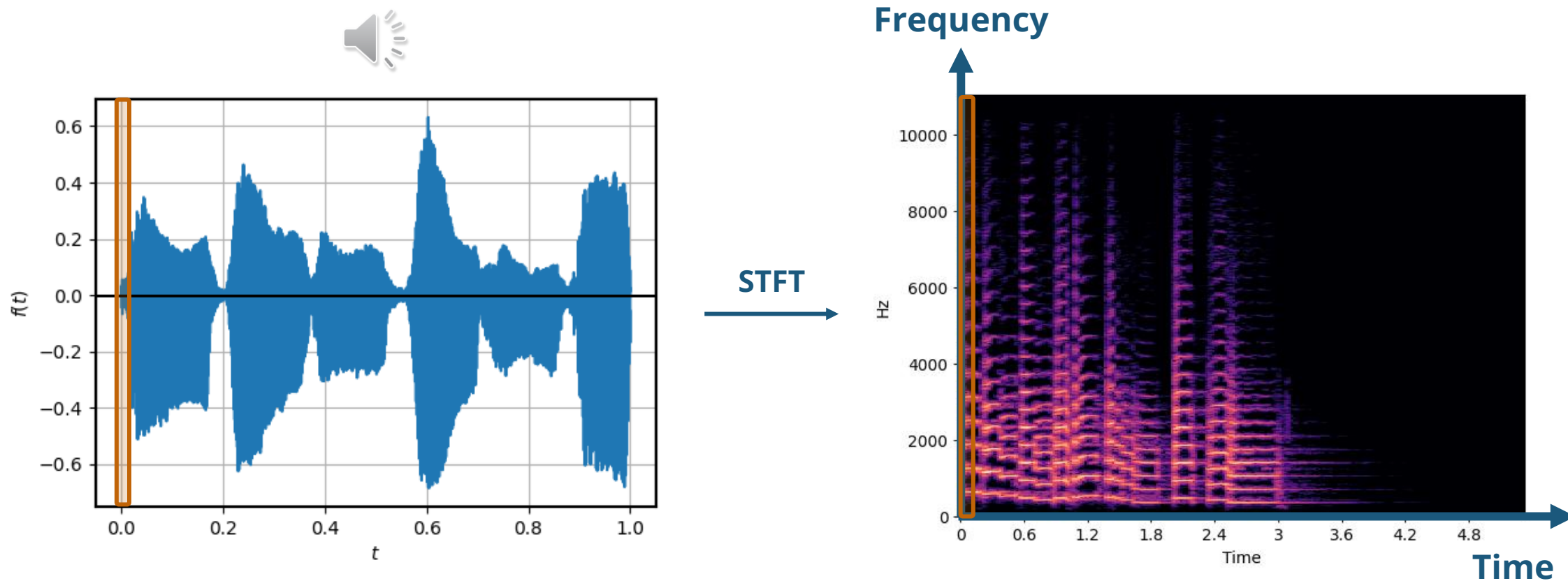
(frequency-domain)



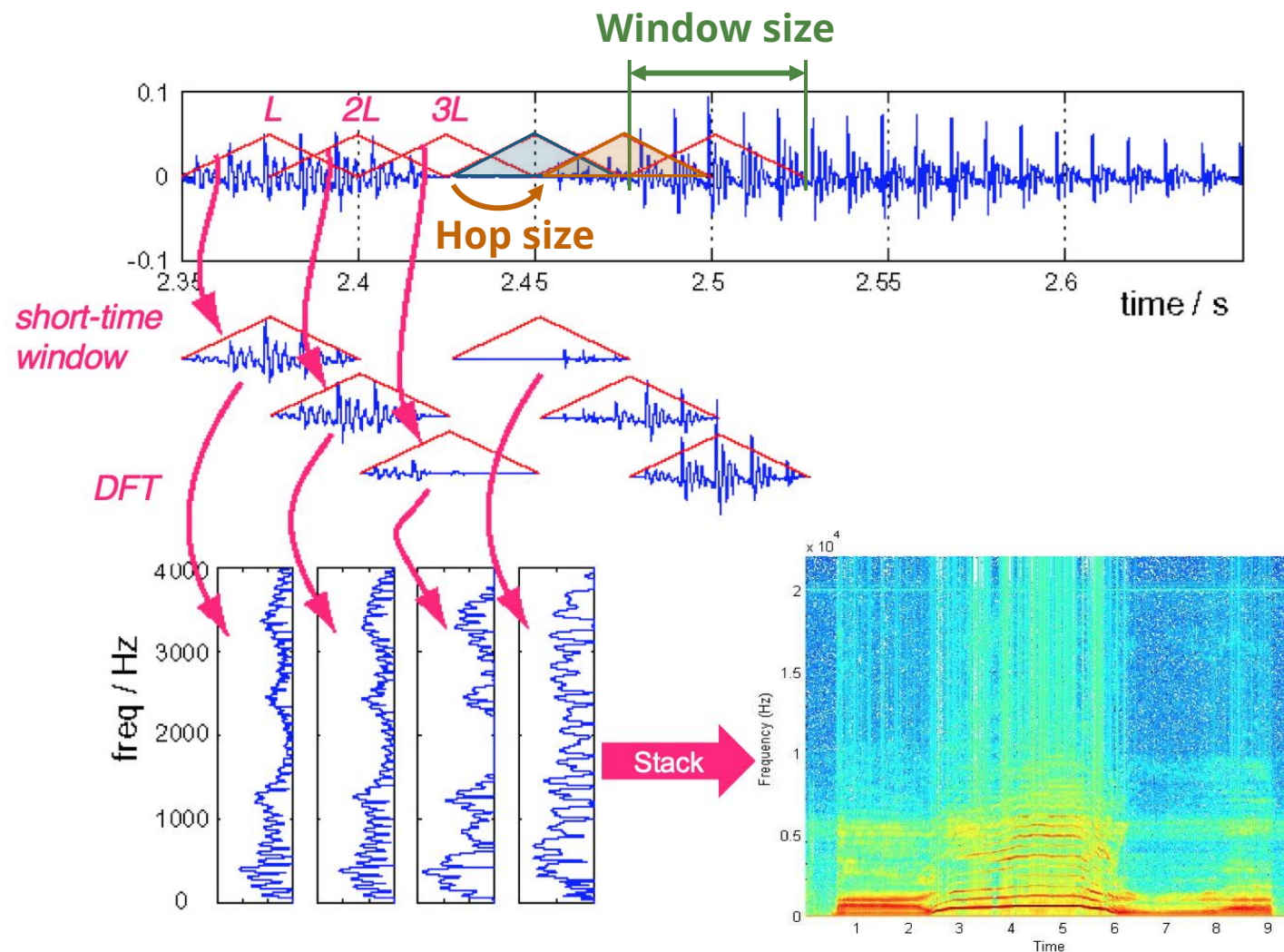
Fourier Transform cannot locate! 🙄

Short-Time Fourier Transform (STFT)

- **Intuition:** Slice the audio into chunks and apply Fourier transform

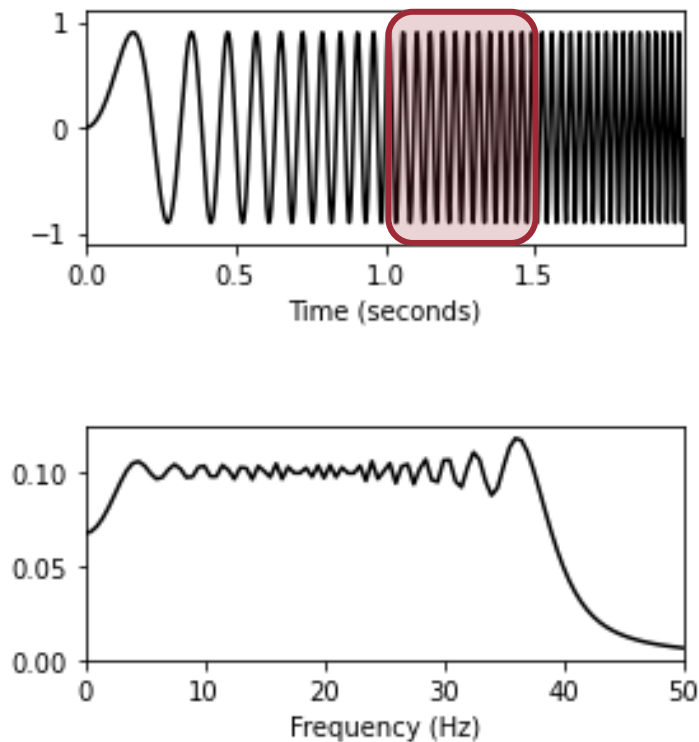


Short-Time Fourier Transform (STFT)

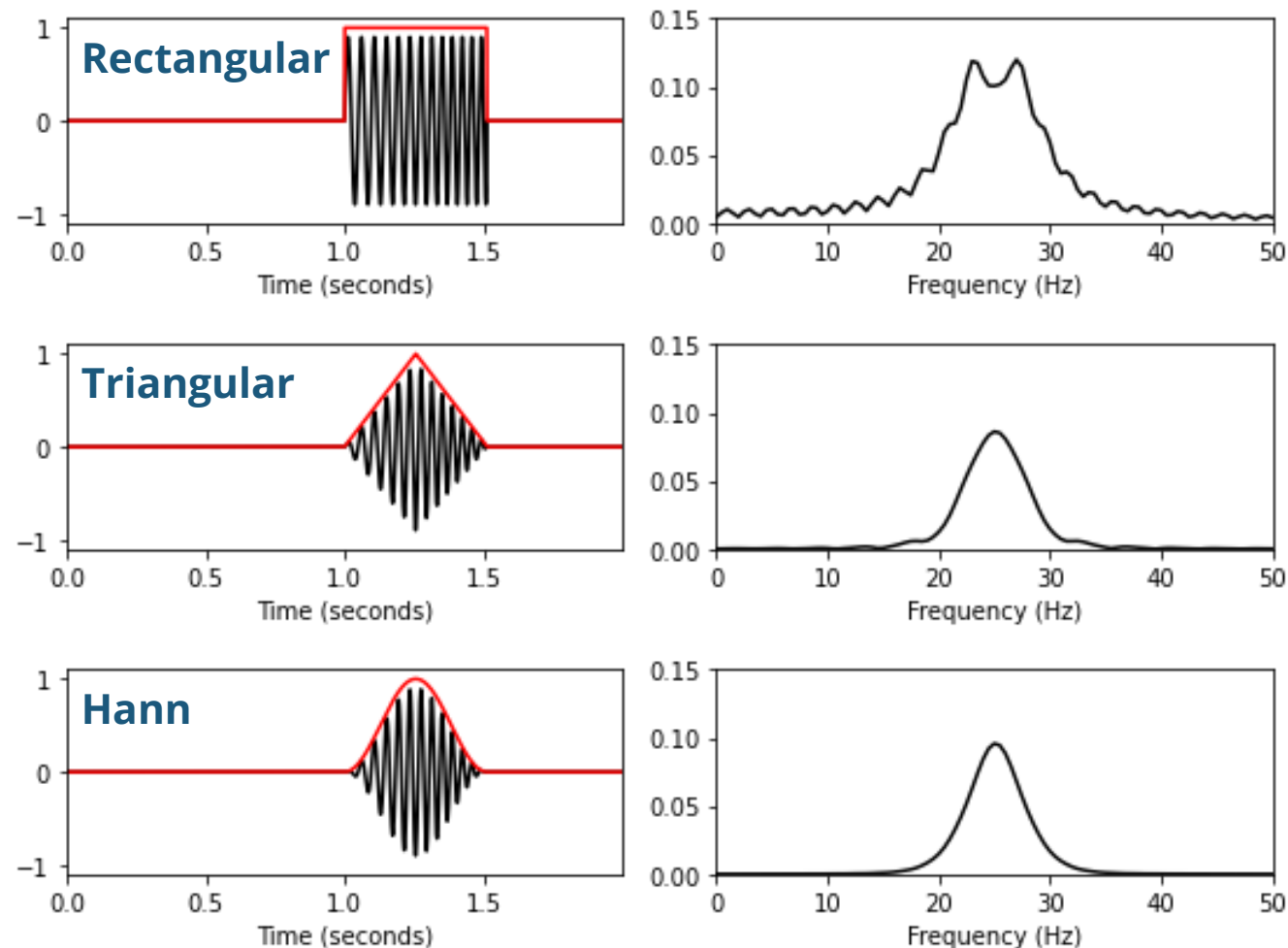


(Source: Bryan Pardo)

Window Functions in STFT

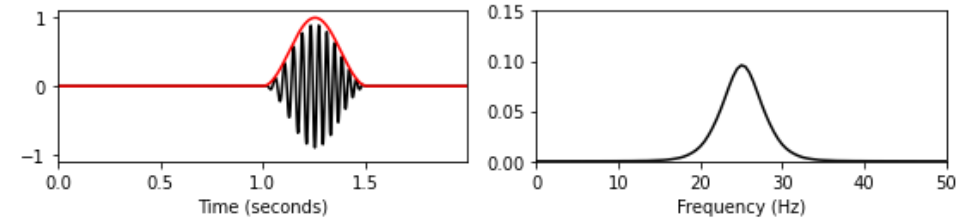
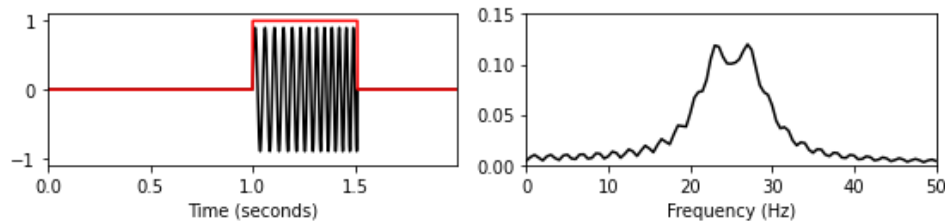
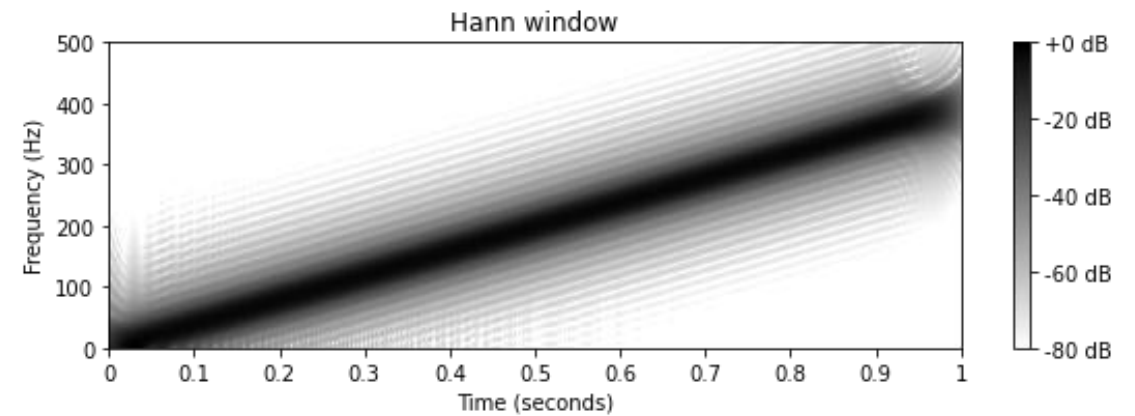
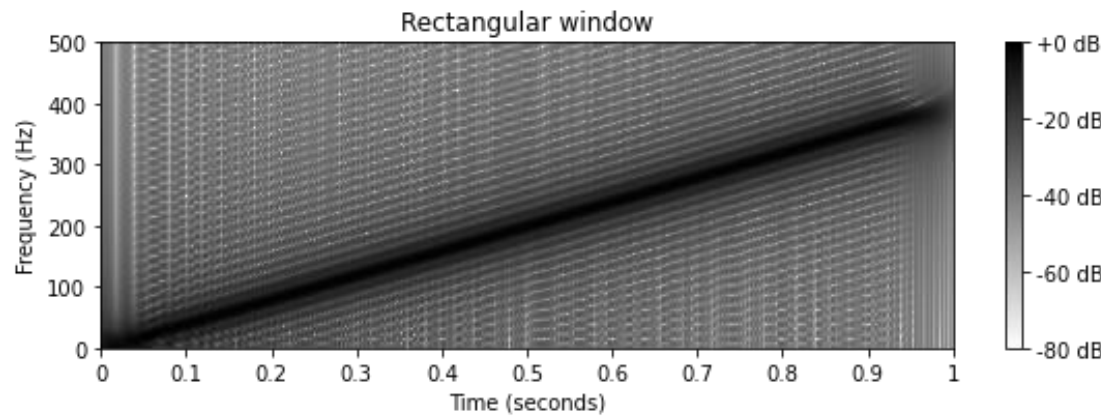
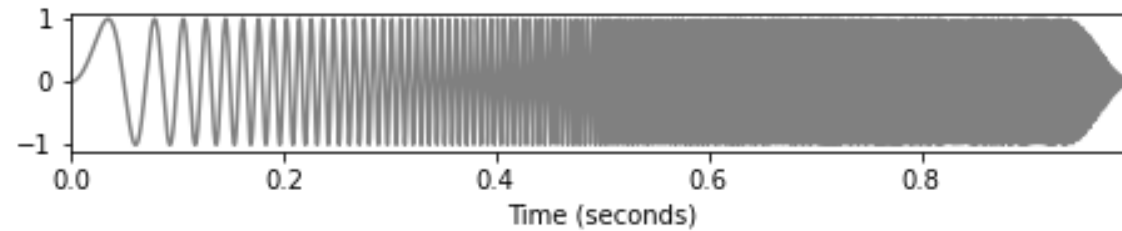


(Source: Müller et al., 2021)



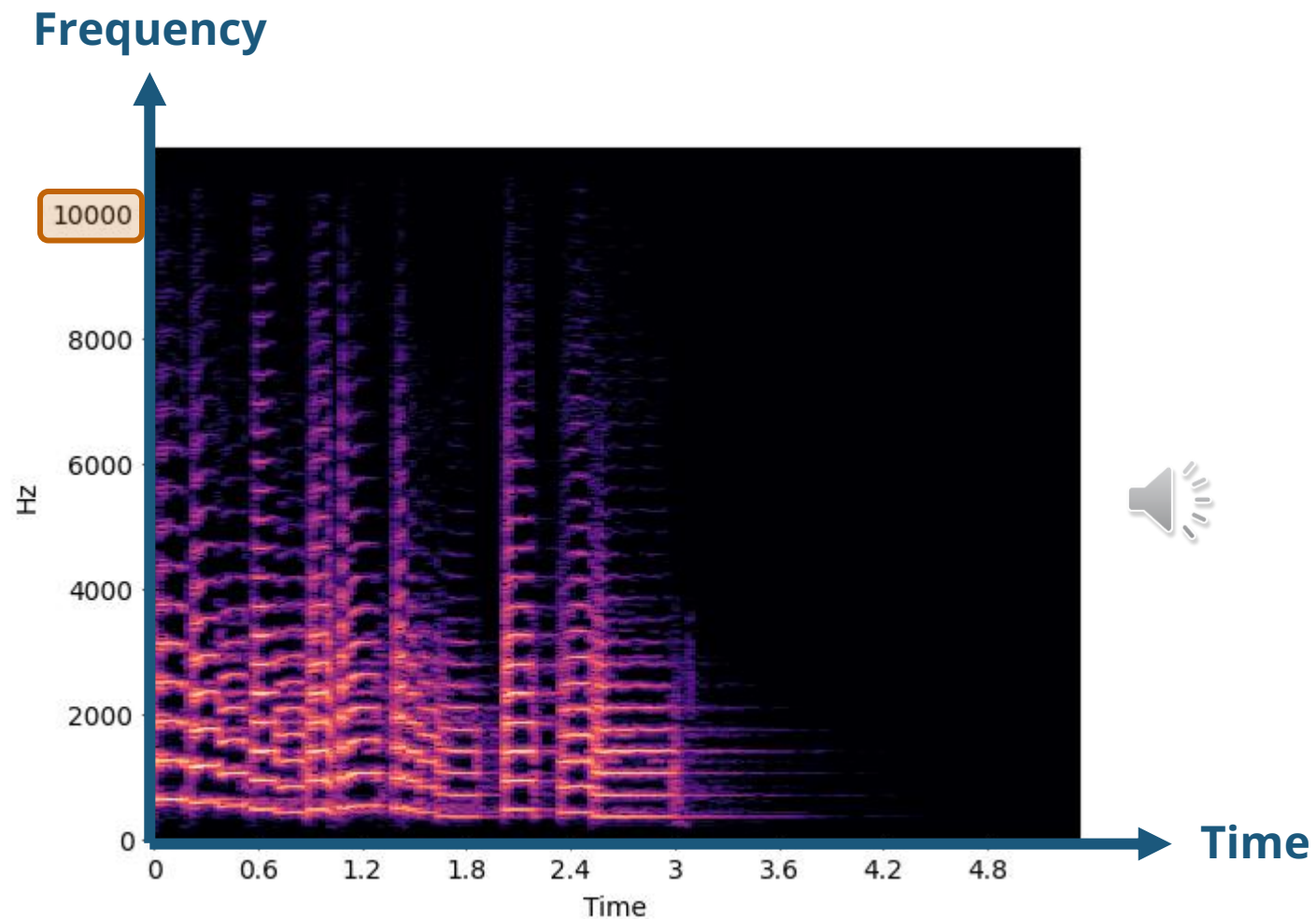
(Source: Müller et al., 2021)

STFT of a Chirp Signal

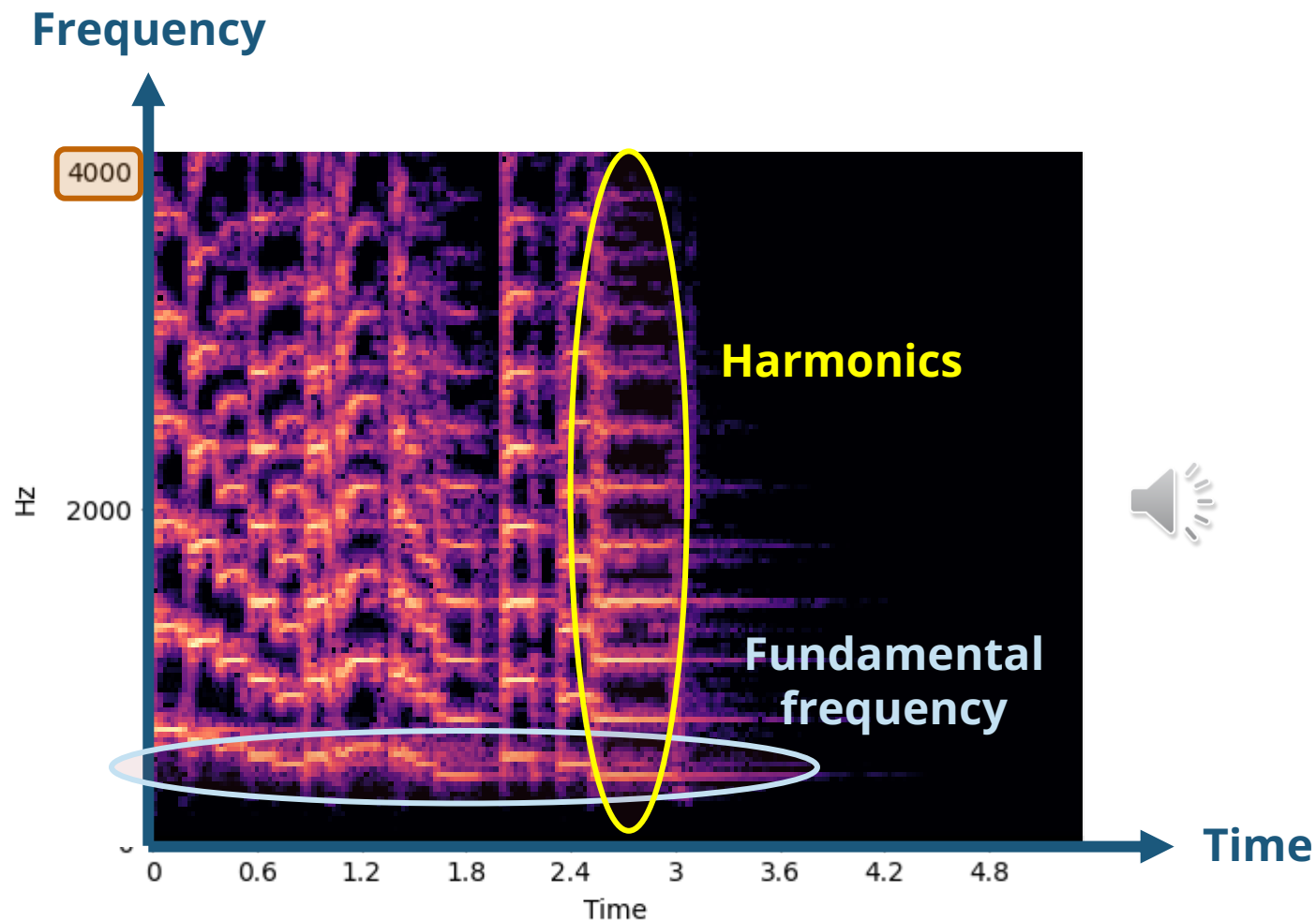


(Source: Müller et al., 2021)

Spectrogram of a Trumpet Sound



| Spectrogram of a Trumpet Sound



🤔 Instrument Timbres

Trumpet

Violin

Flute

Piano

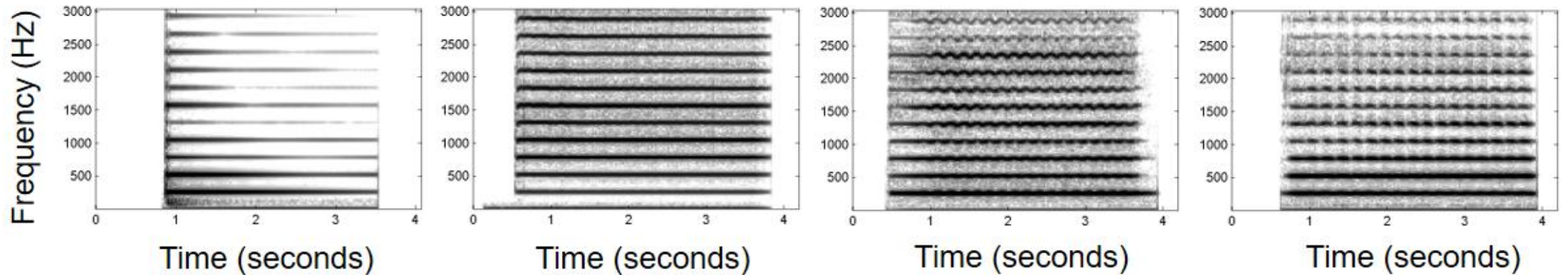


Figure 1.23 from [Müller, FMP, Springer 2015]

(Source: Müller et al., 2021)

🤔 Instrument Timbres

Trumpet

Violin

Flute

Piano

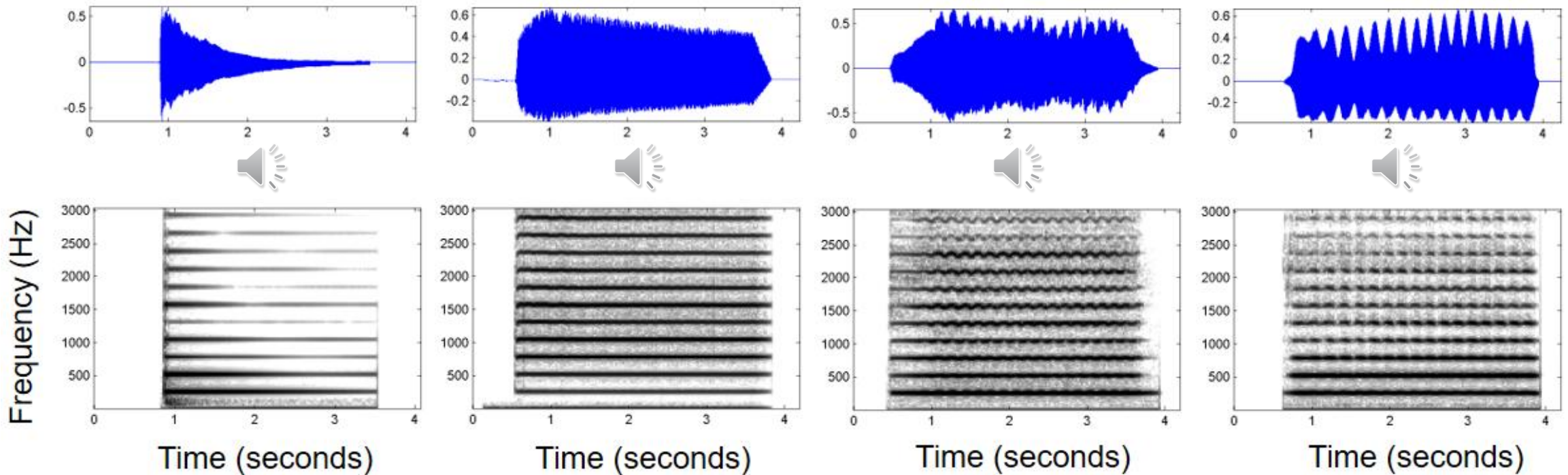


Figure 1.23 from [Müller, FMP, Springer 2015]

(Source: Müller et al., 2021)

🤔 Instrument Timbres

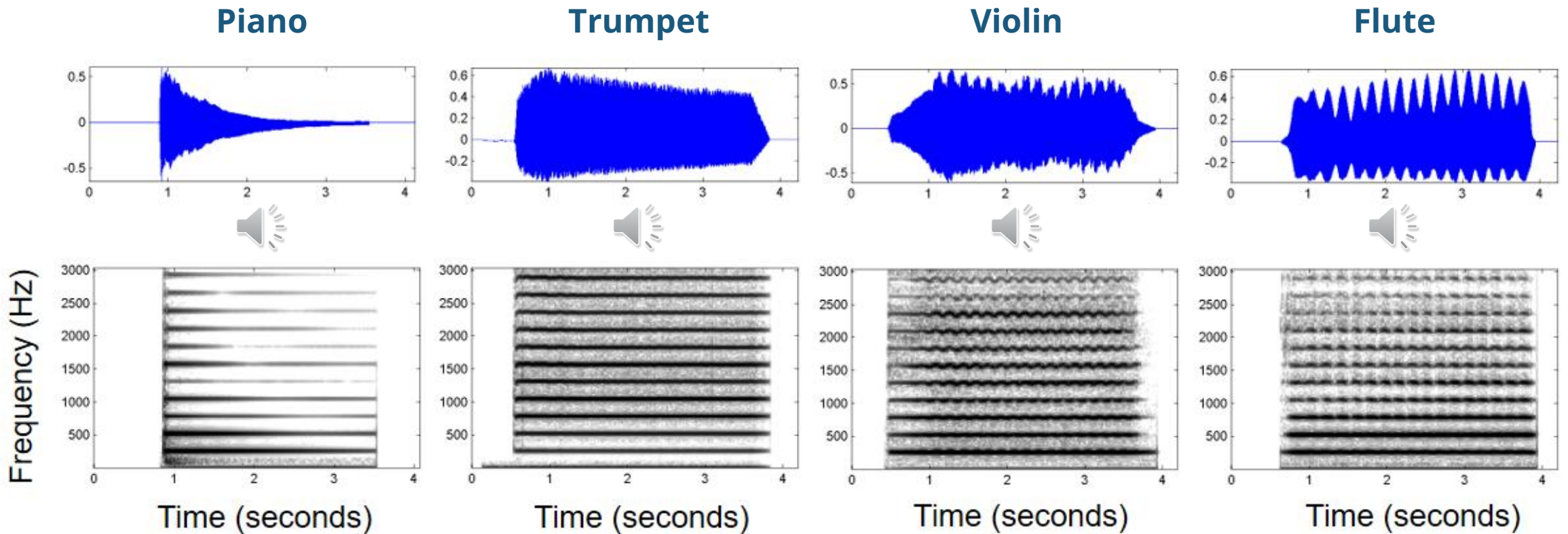
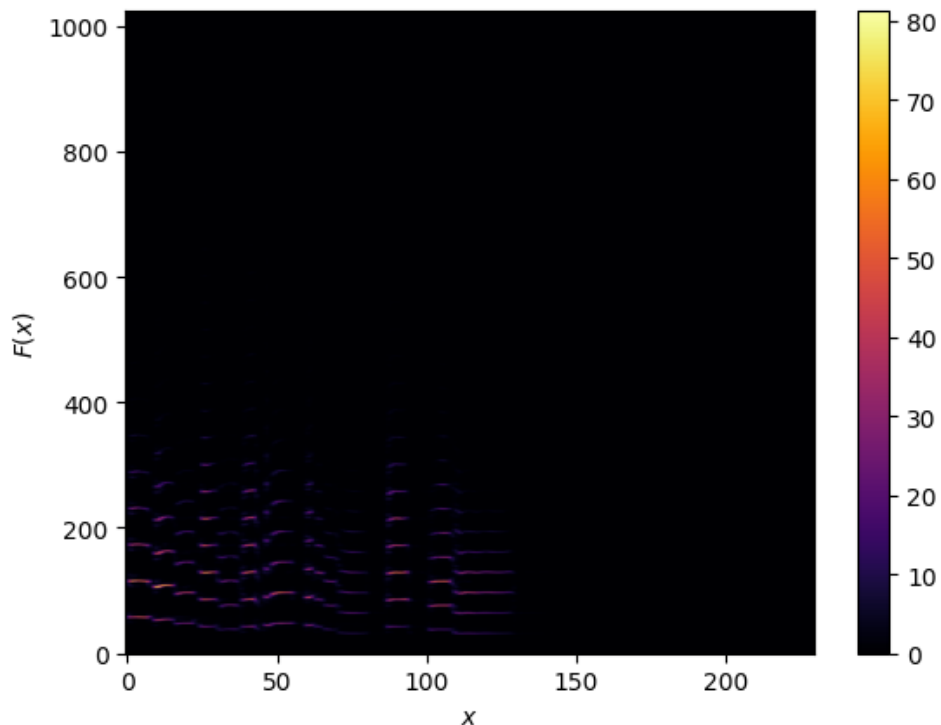


Figure 1.23 from [Müller, FMP, Springer 2015]

(Source: Müller et al., 2021)

Example: `librosa.stft`



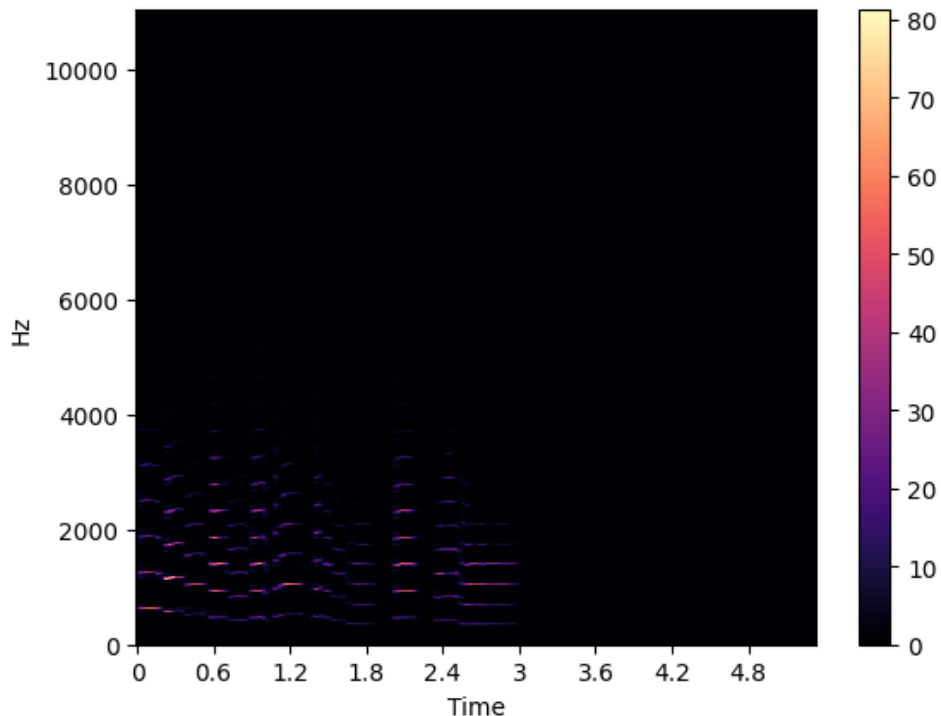
```
# Load the example audio in librosa
y, sr = librosa.load(librosa.example("trumpet"))

# Compute the spectrogram
S = np.abs(librosa.stft(y))

# Plot the spectrogram
im = plt.imshow(S, cmap="inferno", aspect="auto",
                origin="lower")

plt.colorbar(im)
plt.xlabel("Time (sec)")
plt.ylabel("Frequency (Hz)")
plt.show()
```

Example: `librosa.display.specshow`



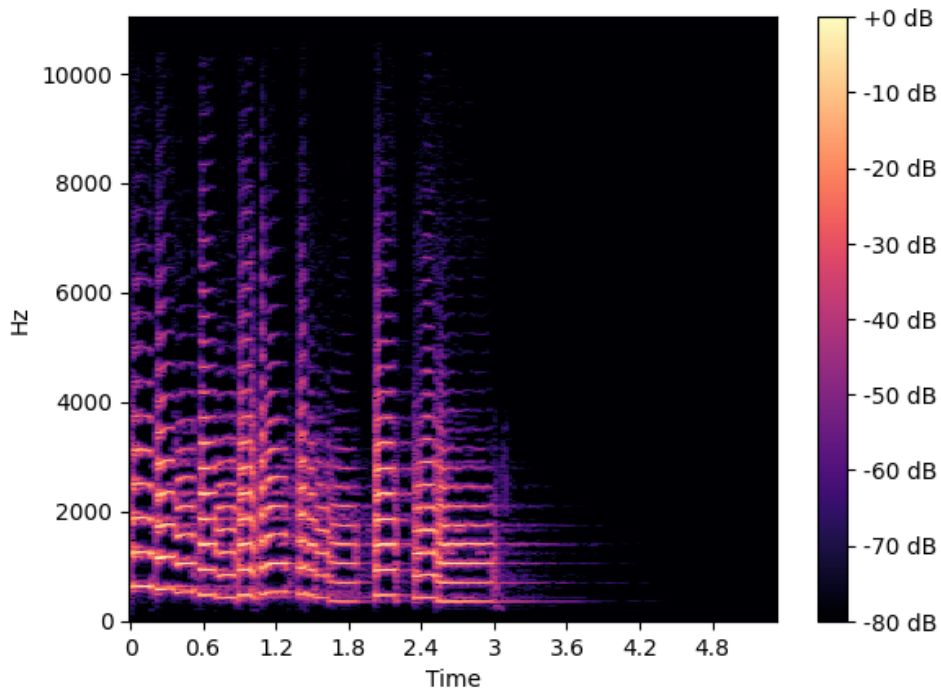
```
# Load the example audio in librosa
y, sr = librosa.load(librosa.example("trumpet"))

# Compute the spectrogram
S = np.abs(librosa.stft(y))

# Plot the spectrogram
im = librosa.display.specshow(S, x_axis="time",
                               y_axis="linear")

plt.colorbar(im)
plt.show()
```

Example: `librosa.amplitude_to_db`



```
# Load the example audio in librosa
y, sr = librosa.load(librosa.example("trumpet"))

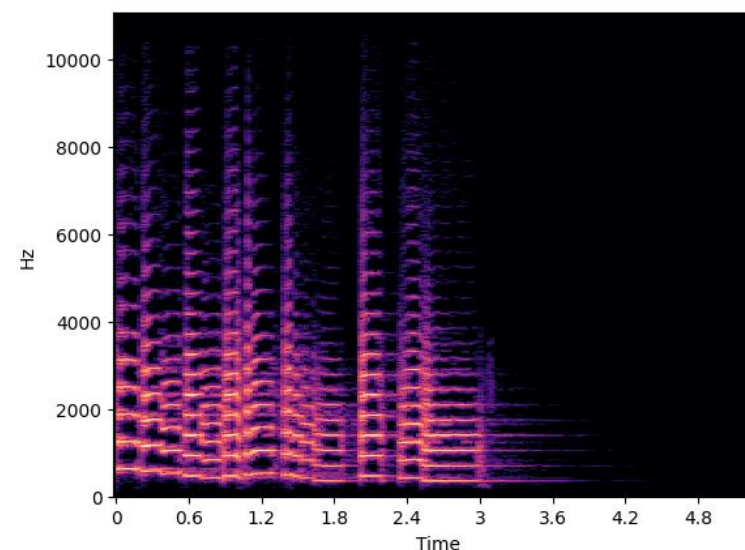
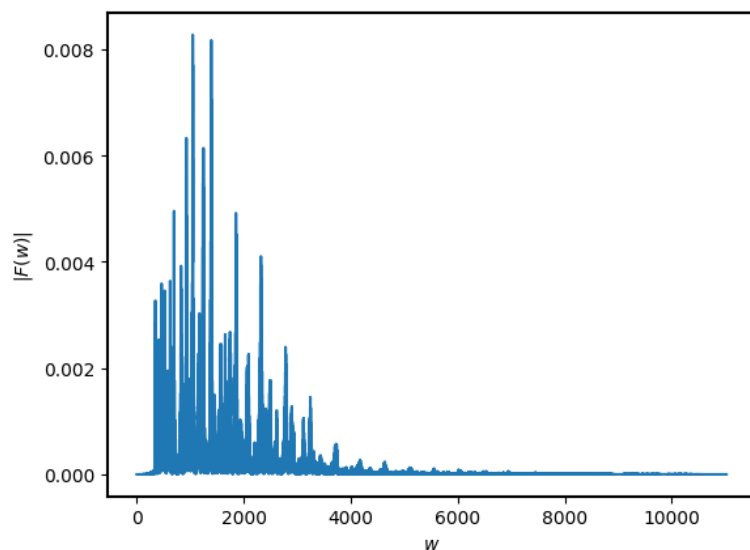
# Compute the spectrogram
S = np.abs(librosa.stft(y))
S_db = librosa.amplitude_to_db(S, ref=np.max)

# Plot the spectrogram
im = librosa.display.specshow(S_db, x_axis="time",
                               y_axis="linear")

plt.colorbar(im, format="%+2.0f dB")
plt.show()
```

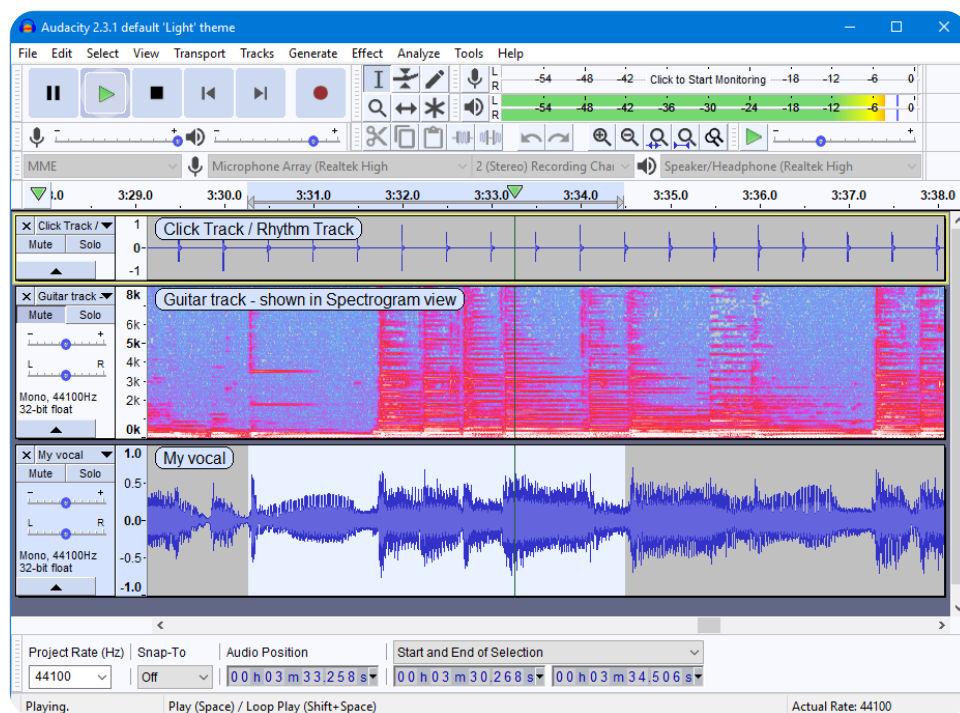
🔥 PA 2: Spectral Analysis

- Instructions will be sent by **emails** and released on the **course website**
- Use **librosa** to process audio files
 - Fast Fourier transform (**FFT**)
 - Short-time Fourier transform (**STFT**)



Software

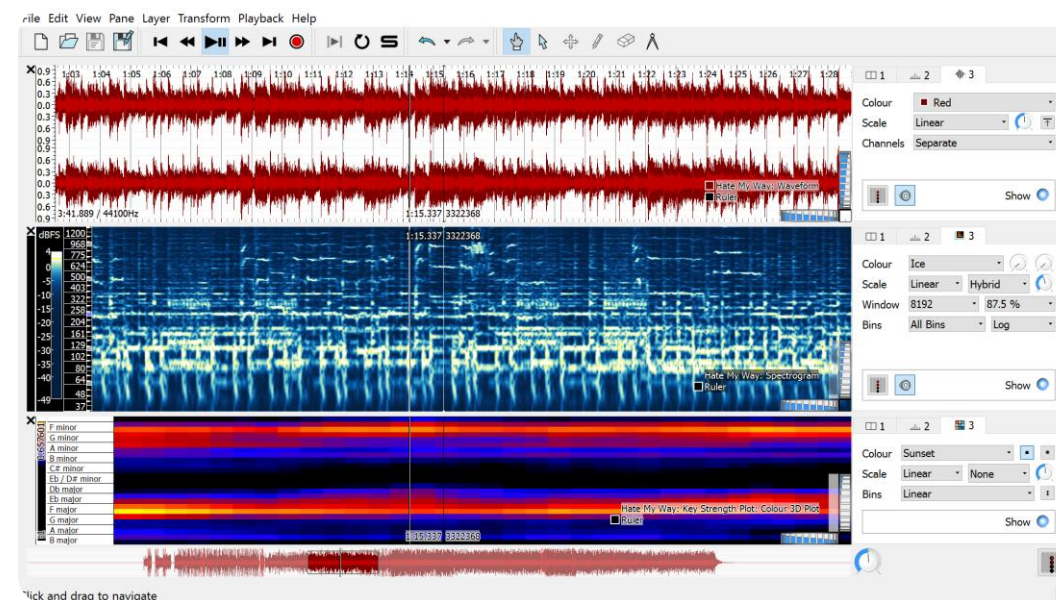
Audacity



(Source: audacity-2.3.1 via Internet Archive)

archive.org/details/audacity-2.3.1
sonicvisualiser.org

Sonic Visualiser

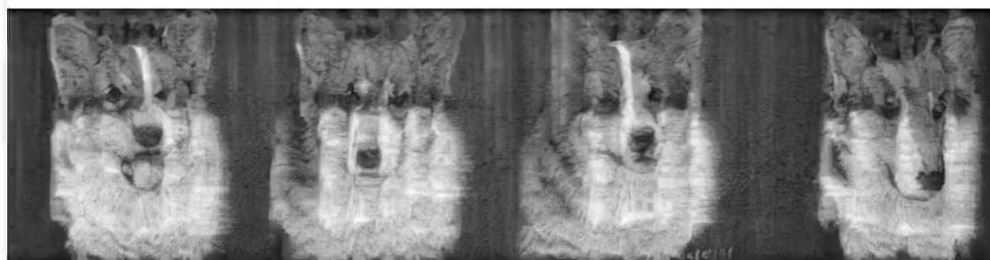


(Source: sonicvisualizer.org)

Images that Sound (Chen et al., 2024)

Using diffusion models to generate visual spectrograms that look like images but can also be played as sound.

Image prompt: a colorful photo of corgis



Audio prompt: dog barking

(Source: Chen et al., 2024)

Image prompt: a colorful photo of tigers



Audio prompt: tiger growling

(Source: Chen et al., 2024)

Images that Sound (Chen et al., 2024)

Using diffusion models to generate visual spectrograms that look like images but can also be played as sound.

Image prompt: a colorful photo of an auto racing game



Audio prompt: a race car passing by and disappearing

(Source: Chen et al., 2024)

Image prompt: a colorful photo of a castle with bell towers



Audio prompt: bell ringing

(Source: Chen et al., 2024)

Recap

Music Representations



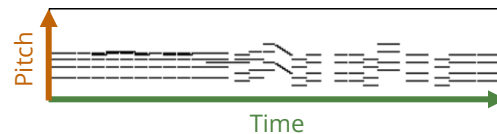
Symbolic music representations

Text-based

Image-based

```
Program_change_0,  
Note_on_60, Time_shift_2, Note_off_60,  
Note_on_60, Time_shift_2, Note_off_60,  
Note_on_76, Time_shift_2, Note_off_67,  
Note_on_67, Time_shift_2, Note_off_67,  
...
```

MIDI



Piano roll



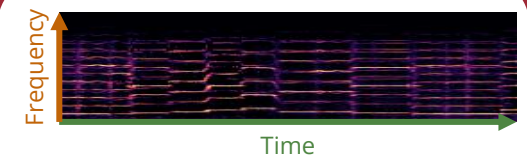
Audio-domain music representations

Time series-based

Image-based



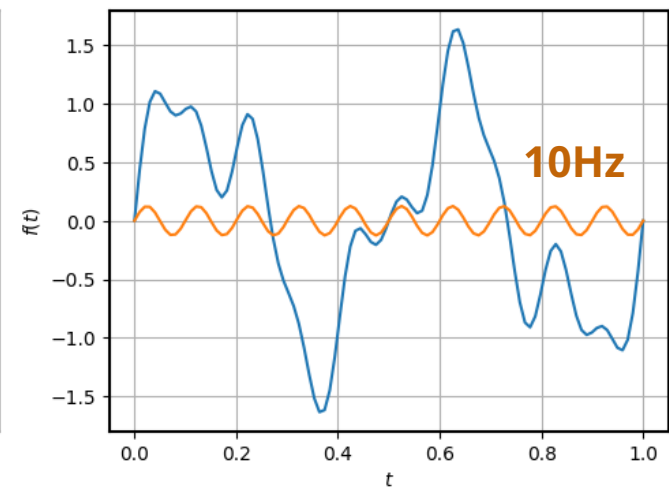
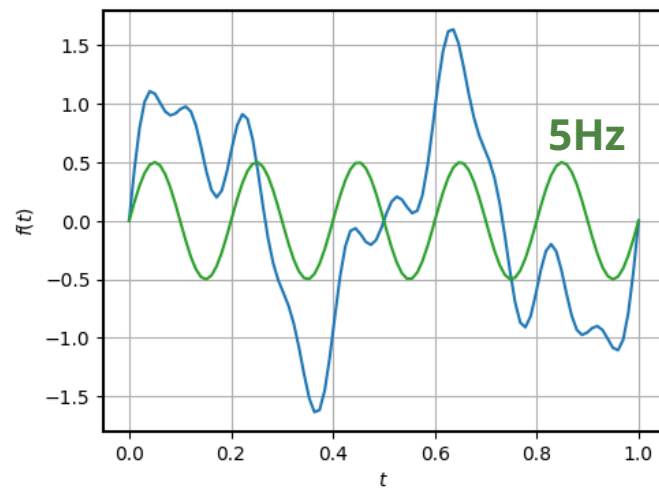
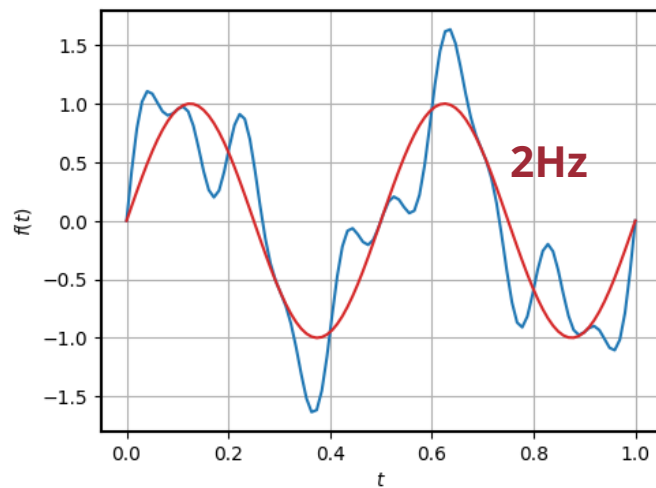
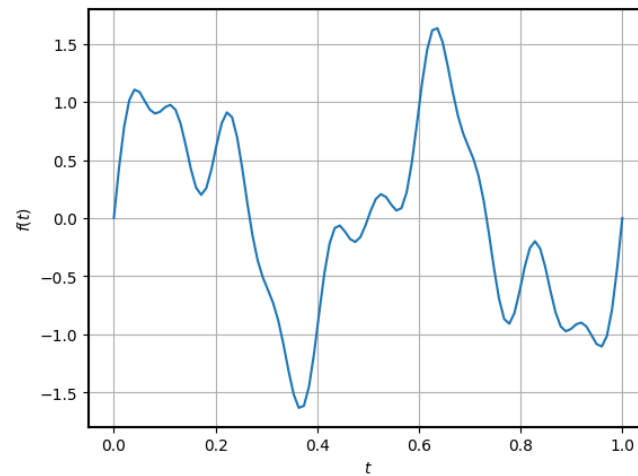
Waveform



Spectrogram

Today's topic!

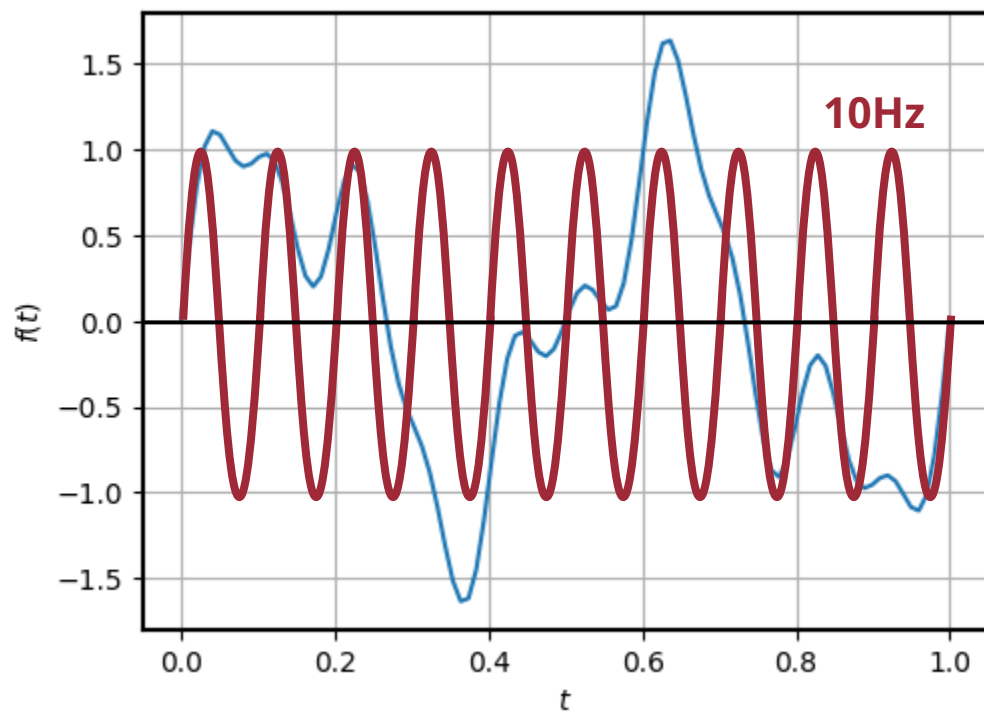
Spectral Analysis



Demystifying Fourier Transform

Signal

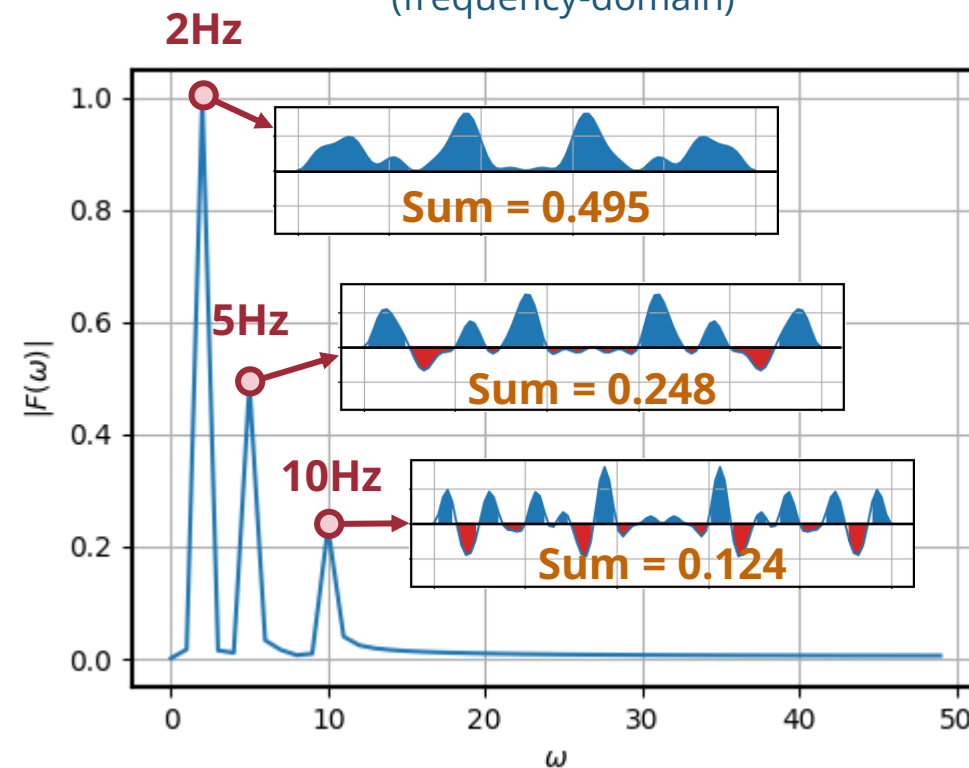
(time-domain)



Fourier Transform
→

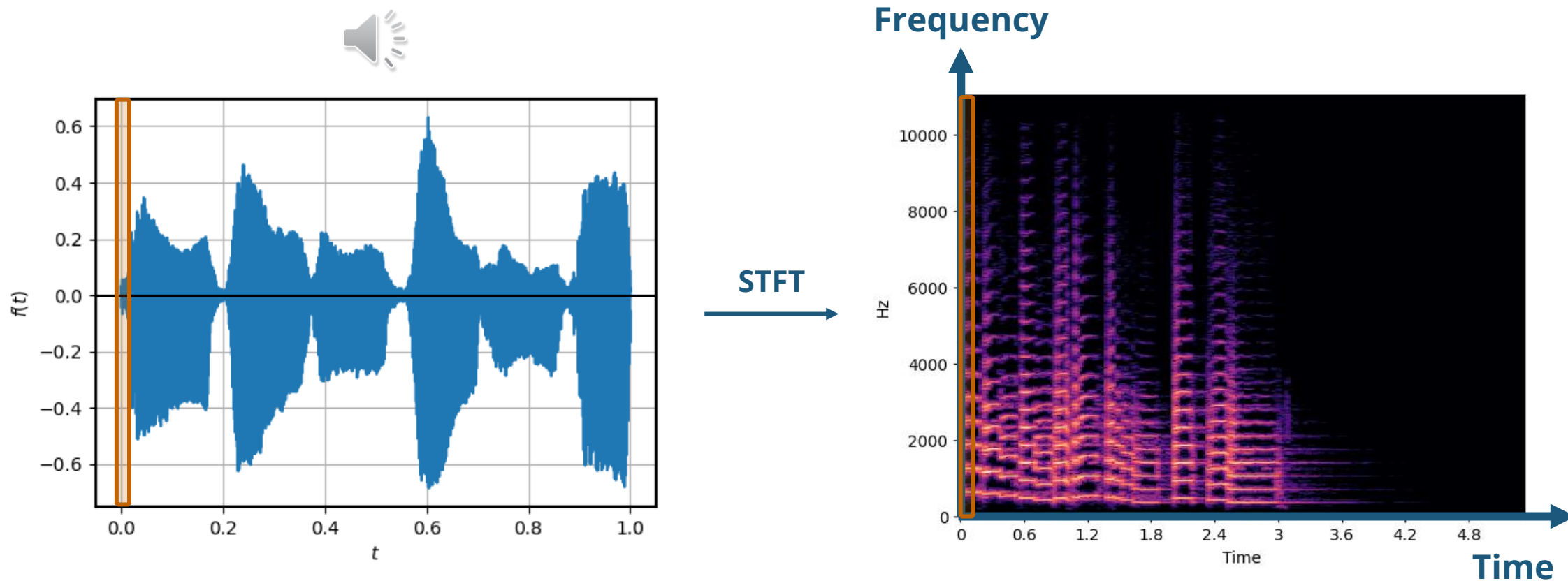
Spectrum

(frequency-domain)

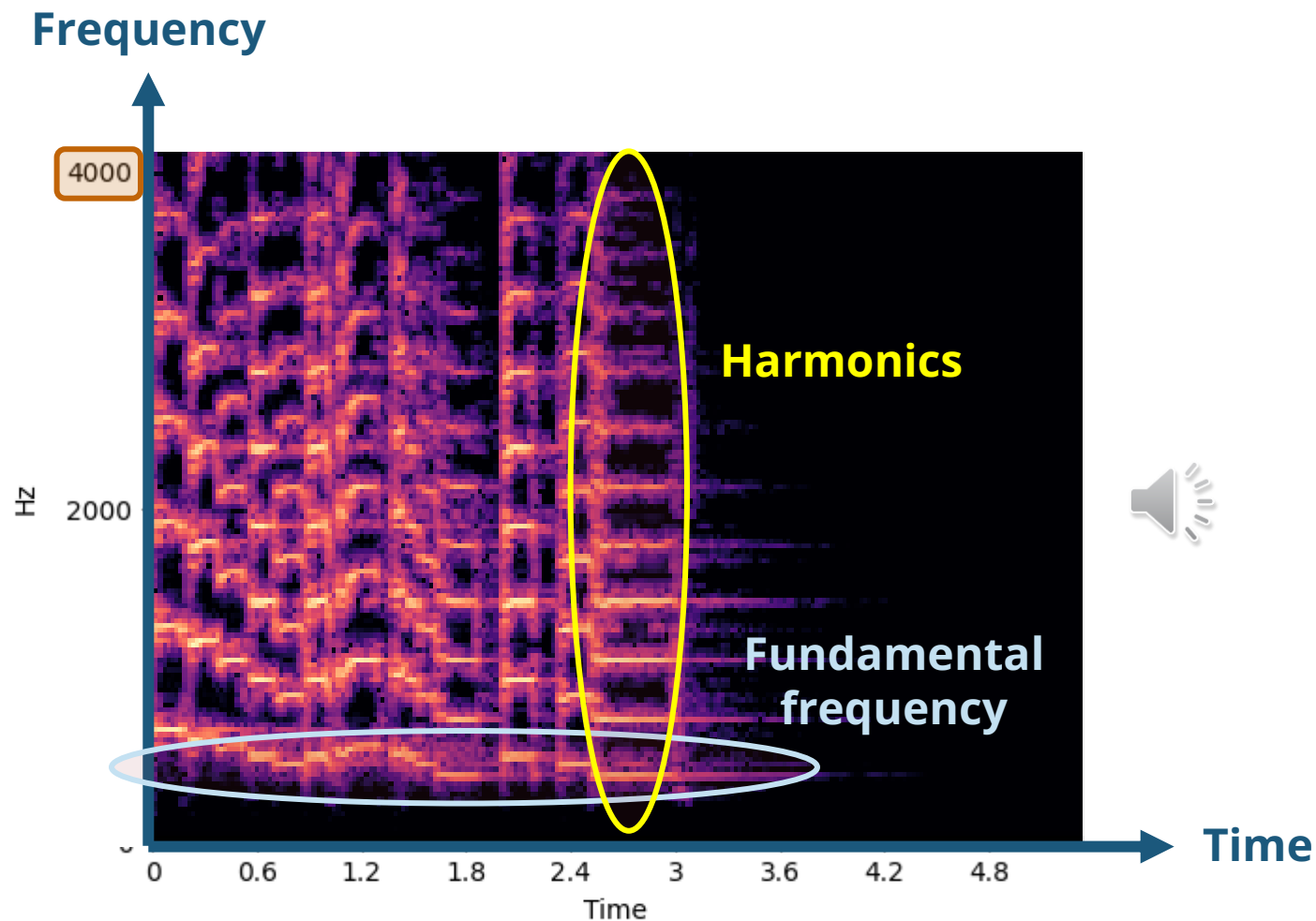


Short-Time Fourier Transform (STFT)

- **Intuition:** Slice the audio into chunks and apply Fourier transform



| Spectrogram of a Trumpet Sound



🤔 Instrument Timbres

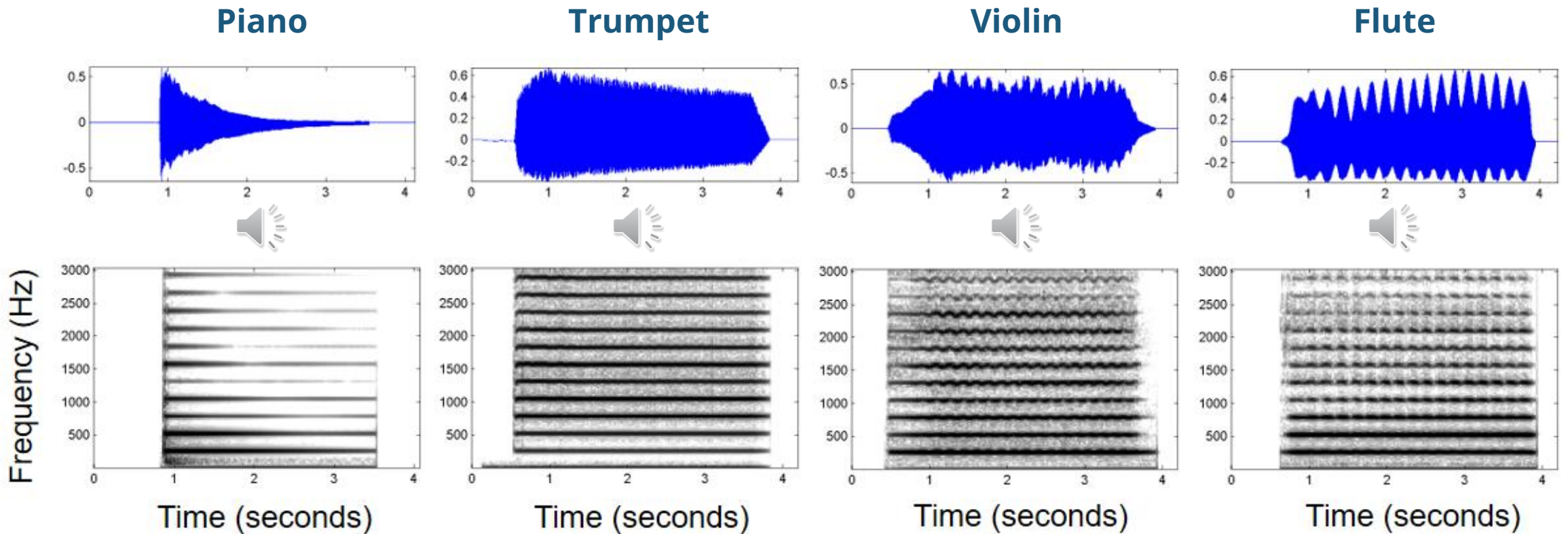


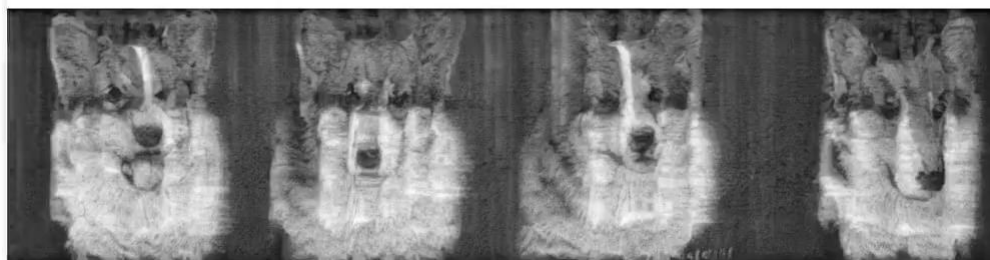
Figure 1.23 from [Müller, FMP, Springer 2015]

(Source: Müller et al., 2021)

Images that Sound (Chen et al., 2024)

Using diffusion models to generate visual spectrograms that look like images but can also be played as sound.

Image prompt: a colorful photo of corgis



Audio prompt: dog barking

(Source: Chen et al., 2024)

Image prompt: a colorful photo of tigers



Audio prompt: tiger growling

(Source: Chen et al., 2024)

Next Lecture

Deep Learning Fundamentals

