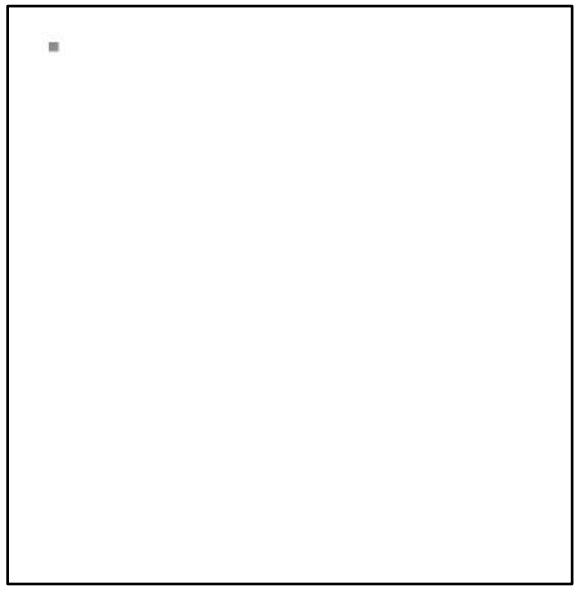
Slides not Covered in Class

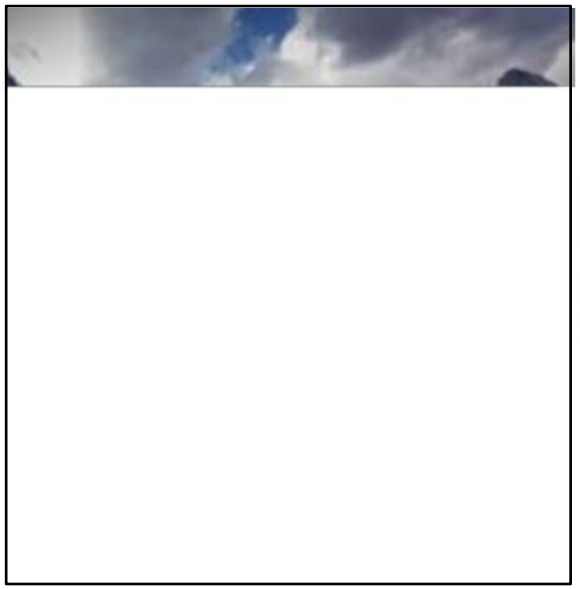
- Last week, the last uploaded slides (Optimal Quantization) were not covered in class.
- These slides will not be included in the material for next exam.
- This material may be included later…

Image Representation (~10K pixels)

Pixel by Pixel



1 pixel

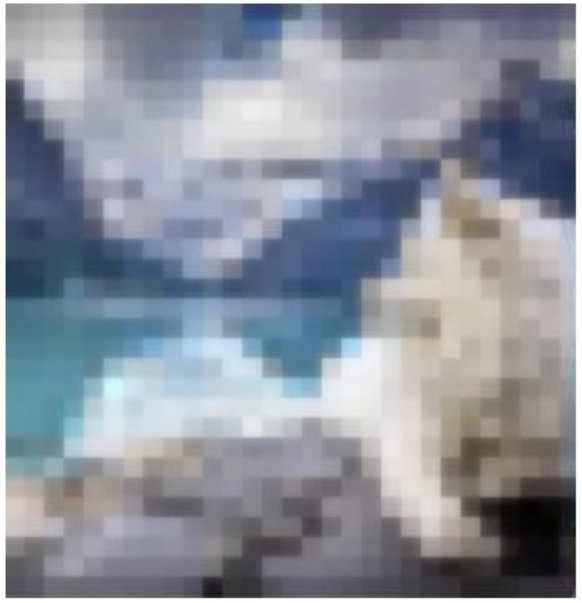


1K pixels

5K pixels

Blurred to Sharp

Image average



Basis of a Vector Space

- Every vector in a vector space is a linear combination of basis vectors

$$\begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} = 2\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + 1\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + 1\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + 0\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

- Standard (natural) basis – local representation – 1 pixel
- In a *k*-dimensional vector space
  - Every set of *k* <u>independent</u> vectors forms a basis
- <u>Orthogonal</u> Basis: every two basis vectors are orthogonal
- <u>Orthonormal</u> Basis: the absolute value of all basis vectors is 1

Transforms: Change of Basis

Natural Basis: (Local, 1 pixel)

$$\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} = 3\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + 0\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + 0\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + 1\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Hadamard Basis (Global, Orthonormal):

$$\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} = 2\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} + 1\begin{bmatrix} 1/2 & -1/2 \\ 1/2 & -1/2 \end{bmatrix} + 1\begin{bmatrix} 1/2 & 1/2 \\ -1/2 & -1/2 \end{bmatrix} + 2\begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1.5 & 0.5 \\ 1.5 & 0.5 \end{bmatrix}$$

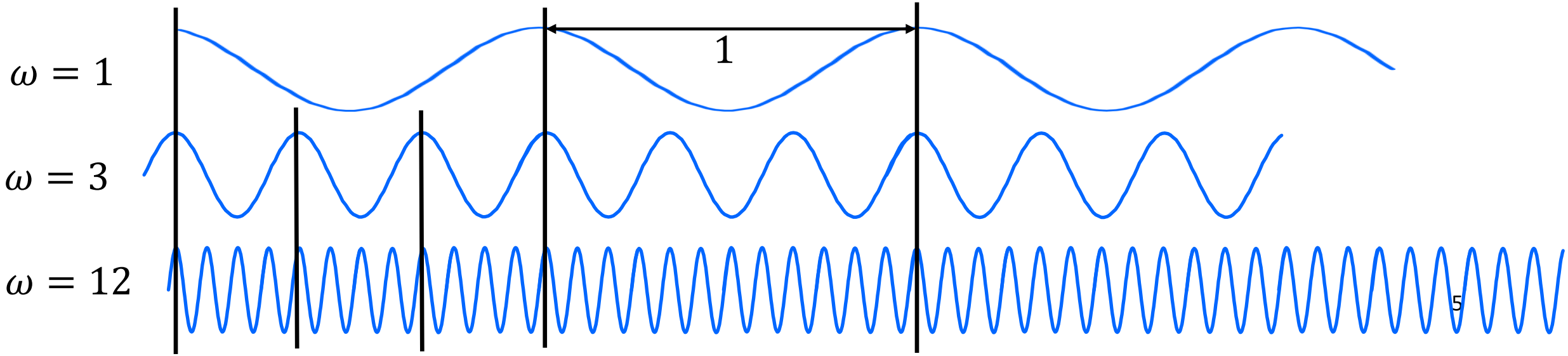
$$\begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$$

Fourier Transform

Easy *Equations but* ***Hard*** *to Understand*

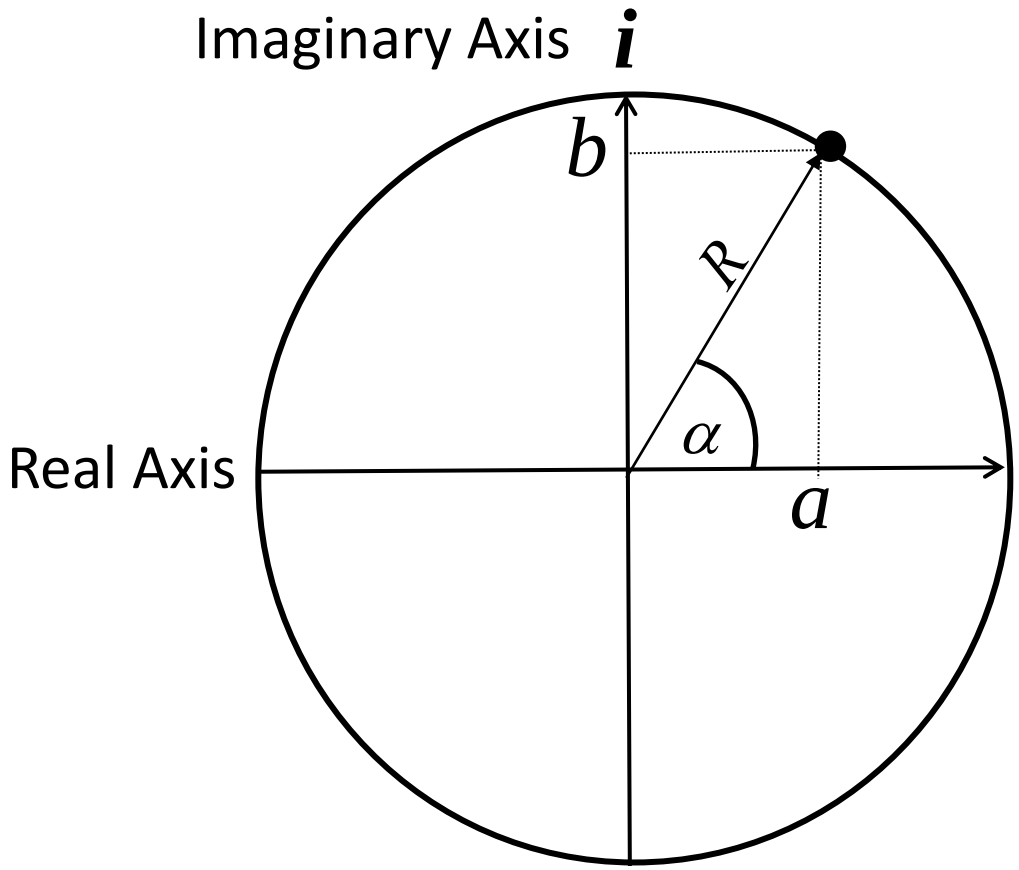
- A representation of signals and images
- Original representation: Collection of samples, collection of pixels
  - One value gives the grey level at one pixel (**Local** Representation)
- Fourier representation: weighted **Sum** of sine waves (<u>frequency</u> ω):
  - Each wave <u>ω</u> is assigned an <u>amplitude</u> and a <u>phase</u> (**Global**)



Jean Baptiste Joseph Fourier (1768-1830)

had crazy idea (1807):

- ***Any*** *periodic function can be rewritten as a weighted sum of sines and cosines of different frequencies.*
- Believe it?
  - Neither did Lagrange, Laplace, Poisson etc.
  - Not translated into English until 1878!
- But it's true!
  - Called Fourier Series
- Are pictures periodic?
  - If we tile them…



Complex Numbers

$$i^2 = -1$$

(a, b) $\quad a + bi = R \cdot e^{i\alpha}$

(R, α) $\quad e^{i\alpha} = \cos(\alpha) + i\sin(\alpha)$

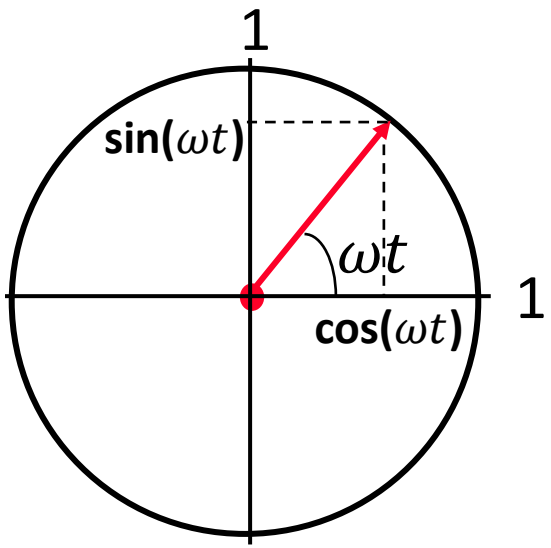
Euler's formula

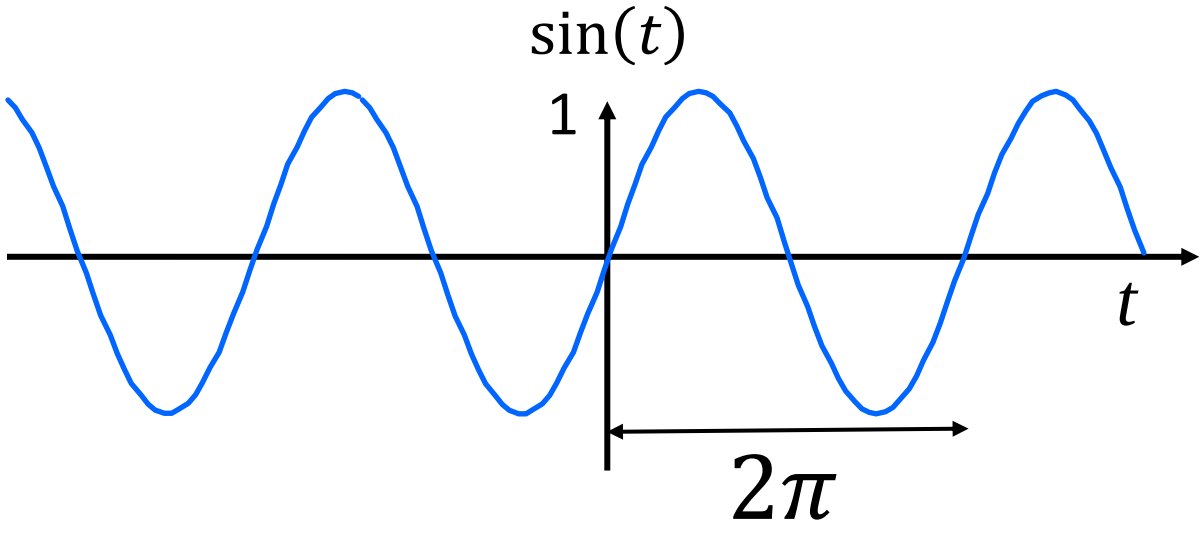
Absolute Value: $R = \sqrt{(a^2 + b^2)}$

Phase: $\alpha = \tan^{-1}\left(\frac{b}{a}\right)$

Multiplication: $R_1 e^{i\alpha_1} \cdot R_2 e^{i\alpha_2} = R_1 \cdot R_2 \cdot e^{i(\alpha_1 + \alpha_2)}$

Rotating Radius Wavelength and Frequency (Hertz)



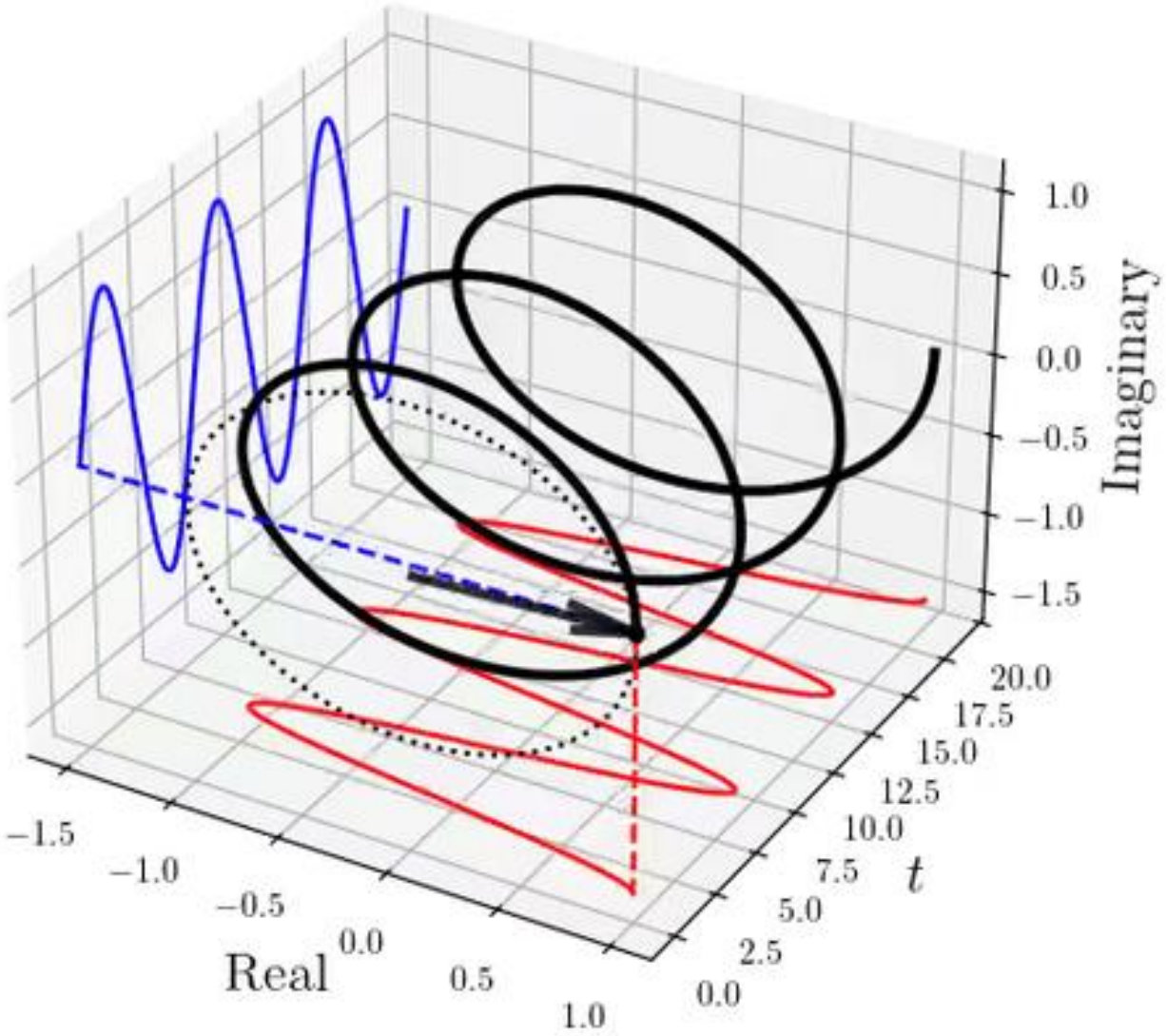


– The wavelength of $sin(t)$ is 2π . Wavelength of $sin(2t)$ is π

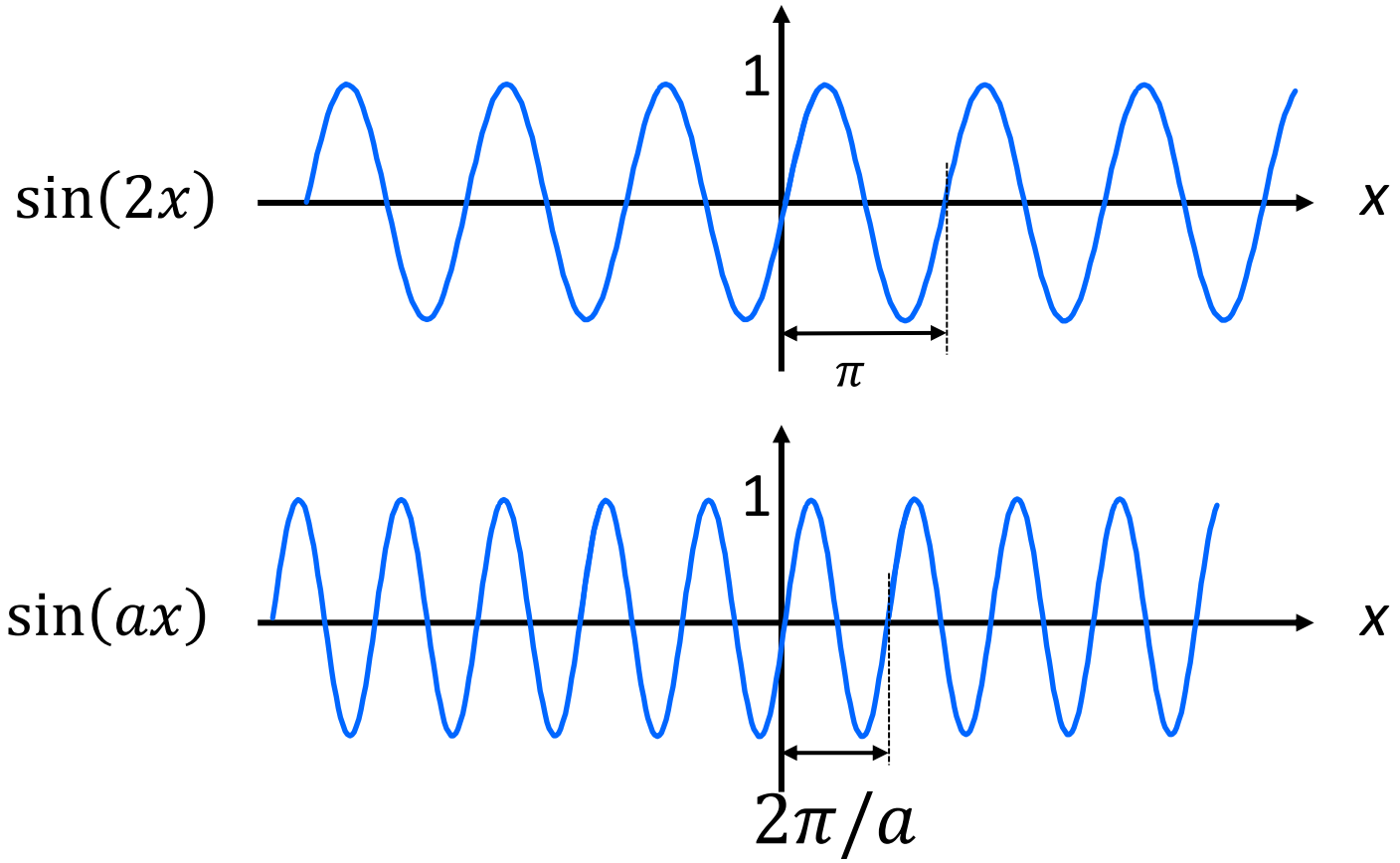
– The frequency of $sin(t)$ is $1/(2\pi)$ Hz. Frequency $sin(2t)$ is $1/\pi$ Hz

- Frequency = the number of waves between 0 and 1

Animation of Euler Formula



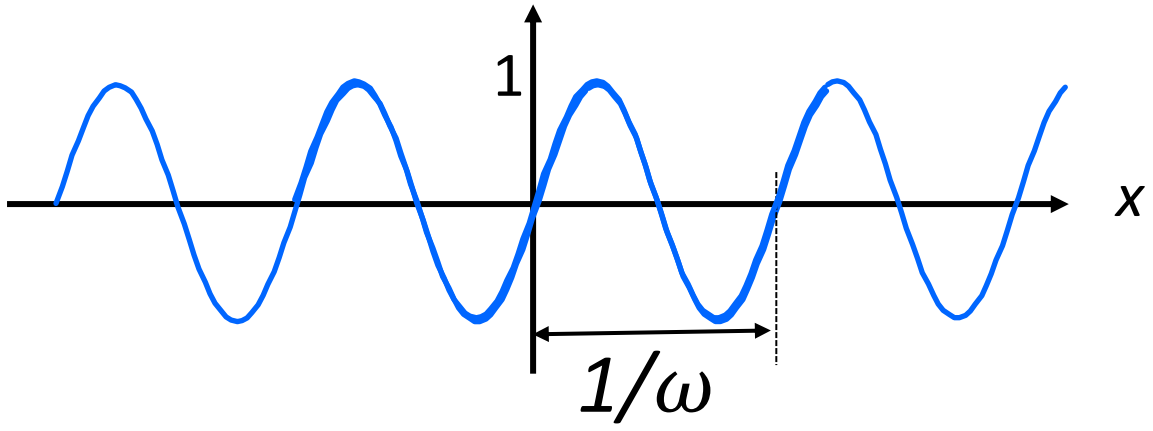
Wavelength and Frequency (Hertz)



- The wavelength of *sin(ax)* is $2\pi/a$
- The frequency of *sin(ax)* is $a/(2\pi)$ *Hz*

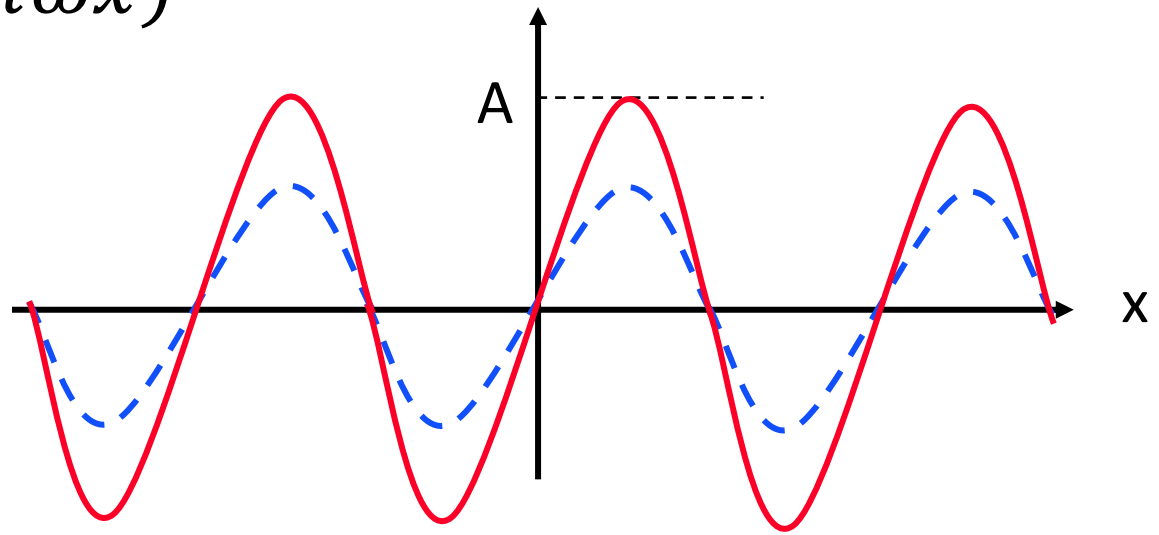
Wavelength and Frequency (Hertz)

$$\sin(2\pi\omega x)$$



- The wavelength of $\sin(2\pi\omega x)$ is $1/\omega$
- The frequency of $\sin(2\pi\omega x)$ is ω Hz

– Changing Amplitude: $A\sin(2\pi\omega x)$

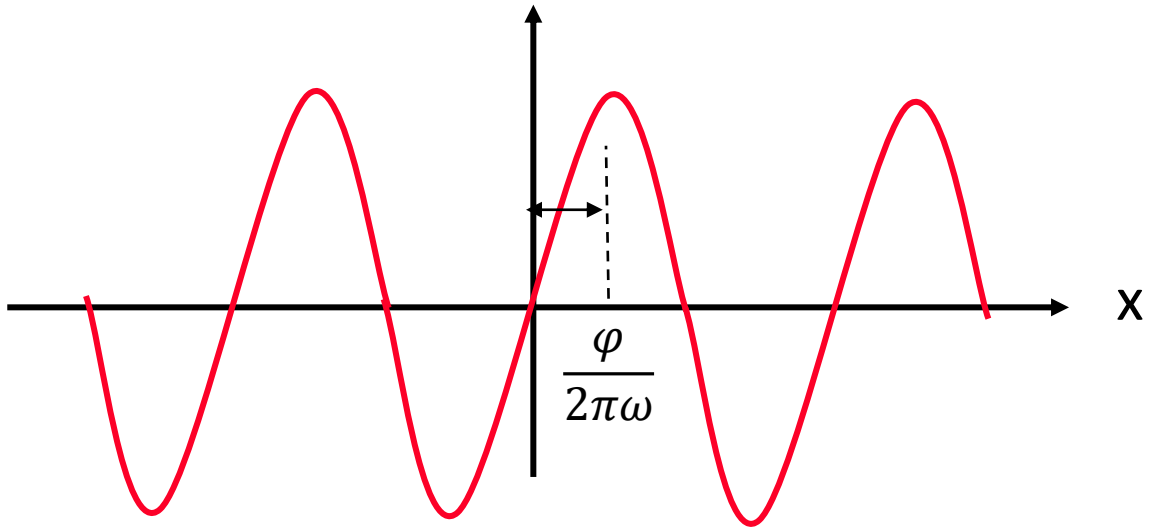


– Changing Phase: $A\sin(2\pi\omega x - \varphi)$

$$\sin(2\pi\omega x - \varphi) = 0$$

$$2\pi\omega x - \varphi = 0,\ 2\pi, \ldots$$

$$x = \frac{\varphi}{2\pi\omega}$$



1-D Discrete Fourier Transform

$$\boldsymbol{f}(\boldsymbol{x}) \mid \big(f(0), f(1), \ldots, f(N-1)\big) \quad \Rightarrow \quad \boldsymbol{F}(\boldsymbol{u}) \mid (F(0), F(1), \ldots, F(N-1))$$

1D Fourier Transform

$$F(u) = \frac{1}{N} \sum_{x=0}^{N-1} f(x) e^{\frac{-2\pi i u x}{N}}$$

$$F(0) = \frac{1}{N} \sum_{x=0}^{N-1} f(x) e^{0} = \bar{f}$$

1D Inverse Fourier Transform

$$f(x) = \frac{1}{1} \sum_{u=0}^{N-1} F(u) e^{\frac{2\pi i u x}{N}}$$

f is a weighted sum of sines and cosines

Complexity: $O(N^2)$ $\Rightarrow$ $(10^6 \quad 10^{12})$

FFT (Fast Fourier Transform): $O(N \log N)$ $\Rightarrow$ $(10^6 \quad 10^7)$

Fourier Basis Vectors

- Computing f from F

$$f(x) = \sum_{u=0}^{N-1} F(u) e^{\frac{2\pi i u x}{N}}$$

- Fourier Basis

$$e^{\frac{2\pi i u x}{N}} = \cos\left(\frac{2\pi u x}{N}\right) + i \sin\left(\frac{2\pi u x}{N}\right)$$

- For each frequency $0 \le u \le N-1$ we have the <u>basis vector</u> above $(x = 0, 1, \ldots N-1)$

Examples: Signal (N=100) Fourier **Spectrum (abs)**

$\omega = 5$

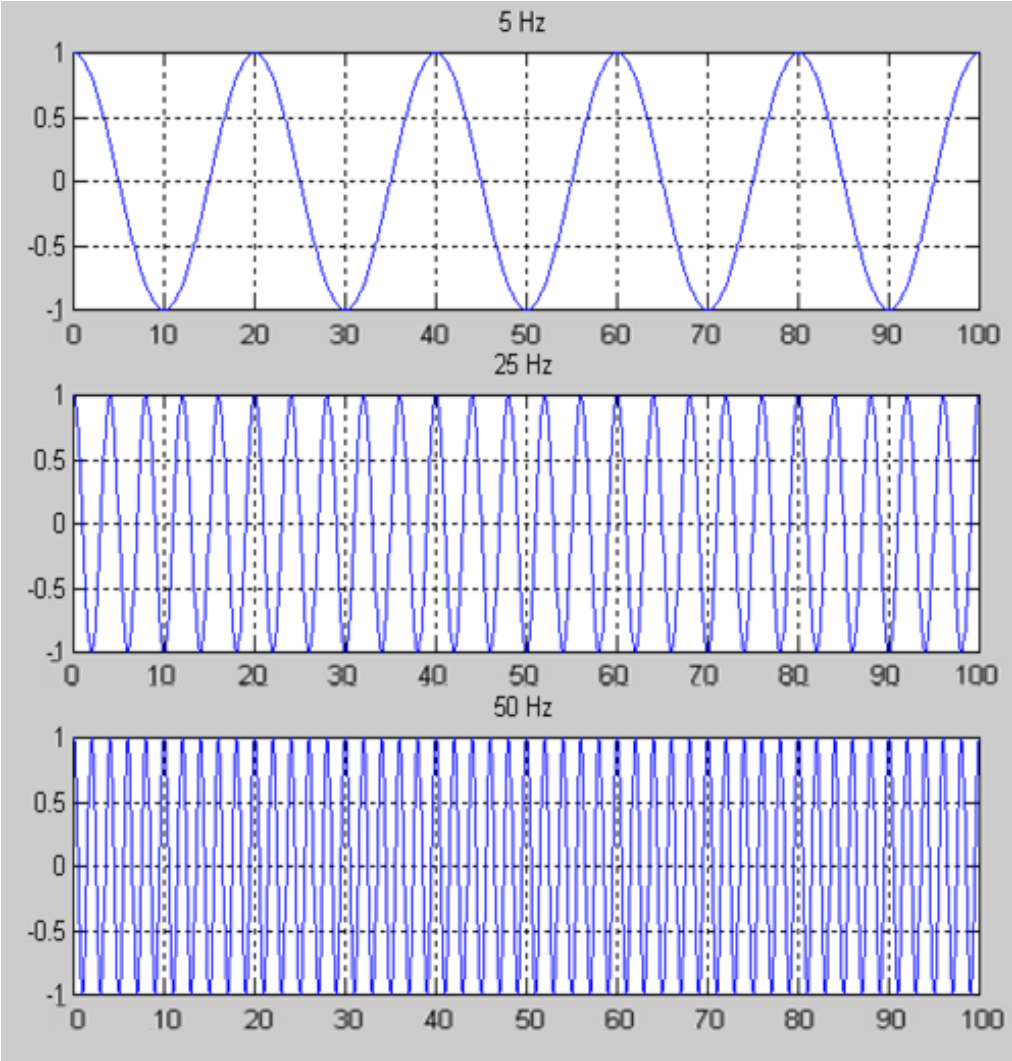
$$\cos\left(\frac{2\pi \cdot 5t}{100}\right)$$

$\omega = 25$

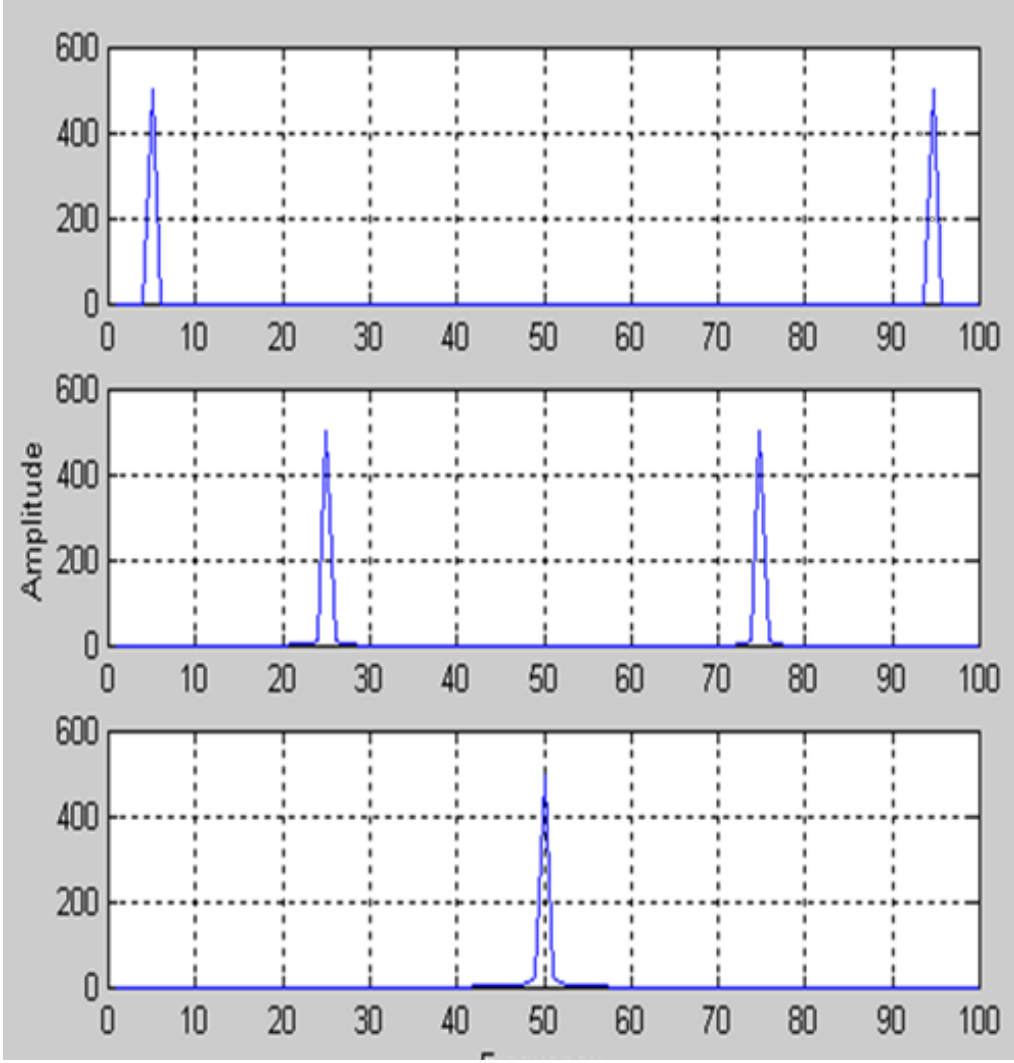
$$\cos\left(\frac{2\pi \cdot 25t}{100}\right)$$

$\omega = 50$

$$\cos\left(\frac{2\pi \cdot 50t}{100}\right)$$



$t \longrightarrow$



$u \longrightarrow$

Periodicity & Symmetry (For a Real Signal f)

$F(u) = F(u + N)$ $\qquad N = 256 \quad \Longrightarrow \quad F(6) = F(262)$

$F(u) = F^*(-u) = F^*(N - u)$ $\qquad \rightarrow F(6) = F^*(-6) = F^*(250)$

$(a + bi)^* = (a - bi)$

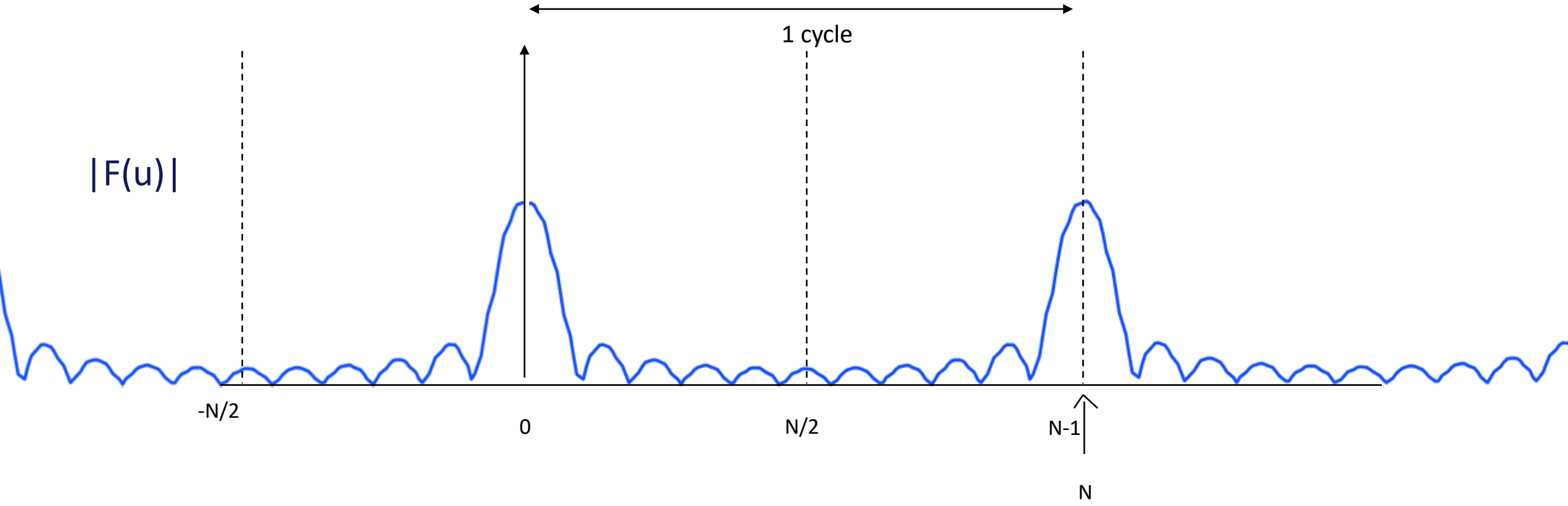
$|F(u)| = |F(-u)|$ $\qquad \rightarrow |F(6)| = |F(-6)| = |F(250)|$

- Above properties are immediate from Fourier definition
- Fourier of N real numbers gives N complex Fourier Coefficients ($2N$ real numbers) – redundancy!
- Given $f(0), f(1), \ldots, f(N-1)$ we only need $\frac{N}{2}$ coefficients:

$F(0), F(1), \ldots, F(\frac{N}{2} - 1)$ $\qquad$ because $F(N - u) = F^*(u)$

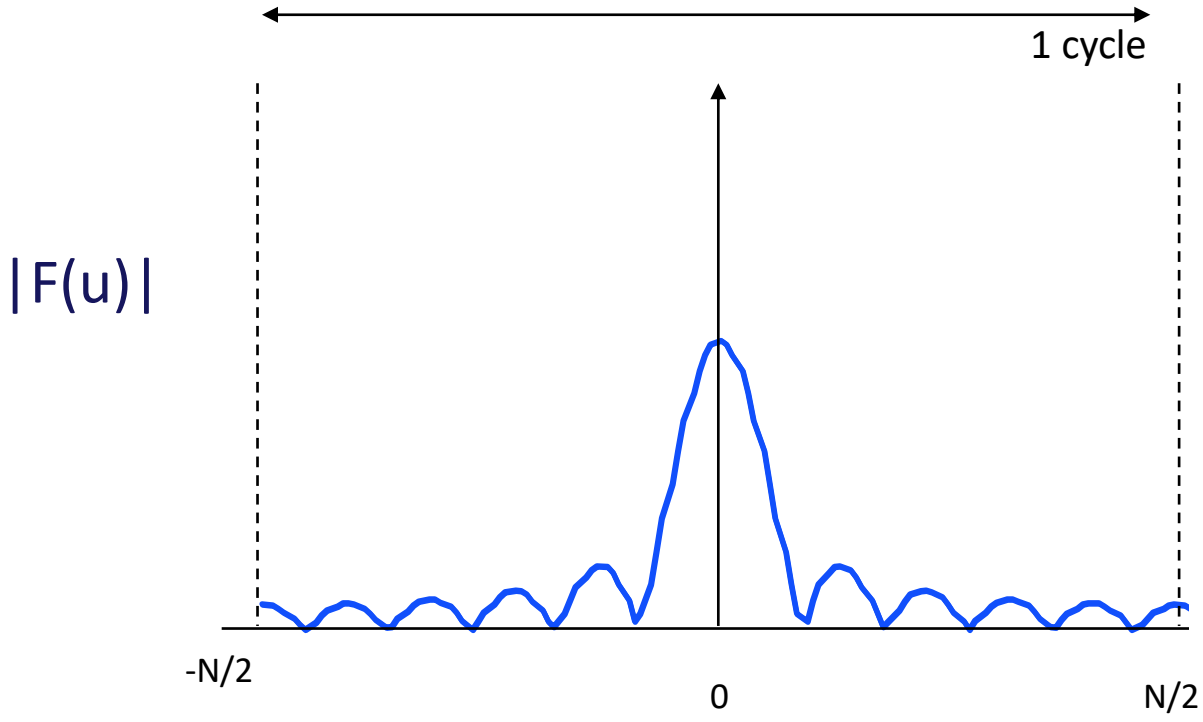
Periodicity & Symmetry (1D)

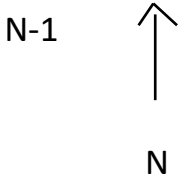
$$|F(u)| = |F(N-u)| = |F(-u)|$$



Periodicity & Symmetry (1D)

$$|F(u)| = |F(N-u)| = |F(-u)|$$





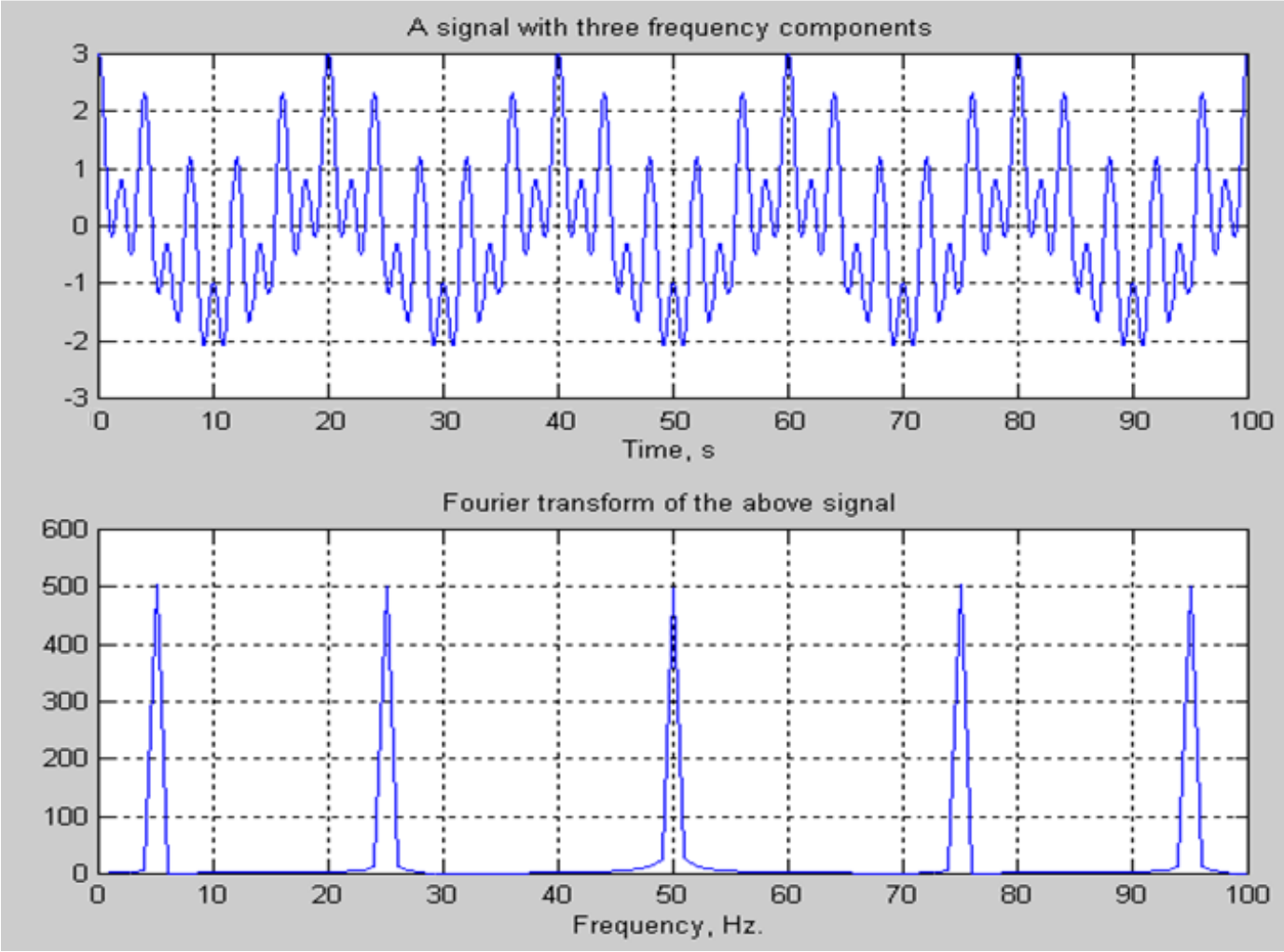
Zero-centered representation

Fourier Analysis – Examples (cont'd)

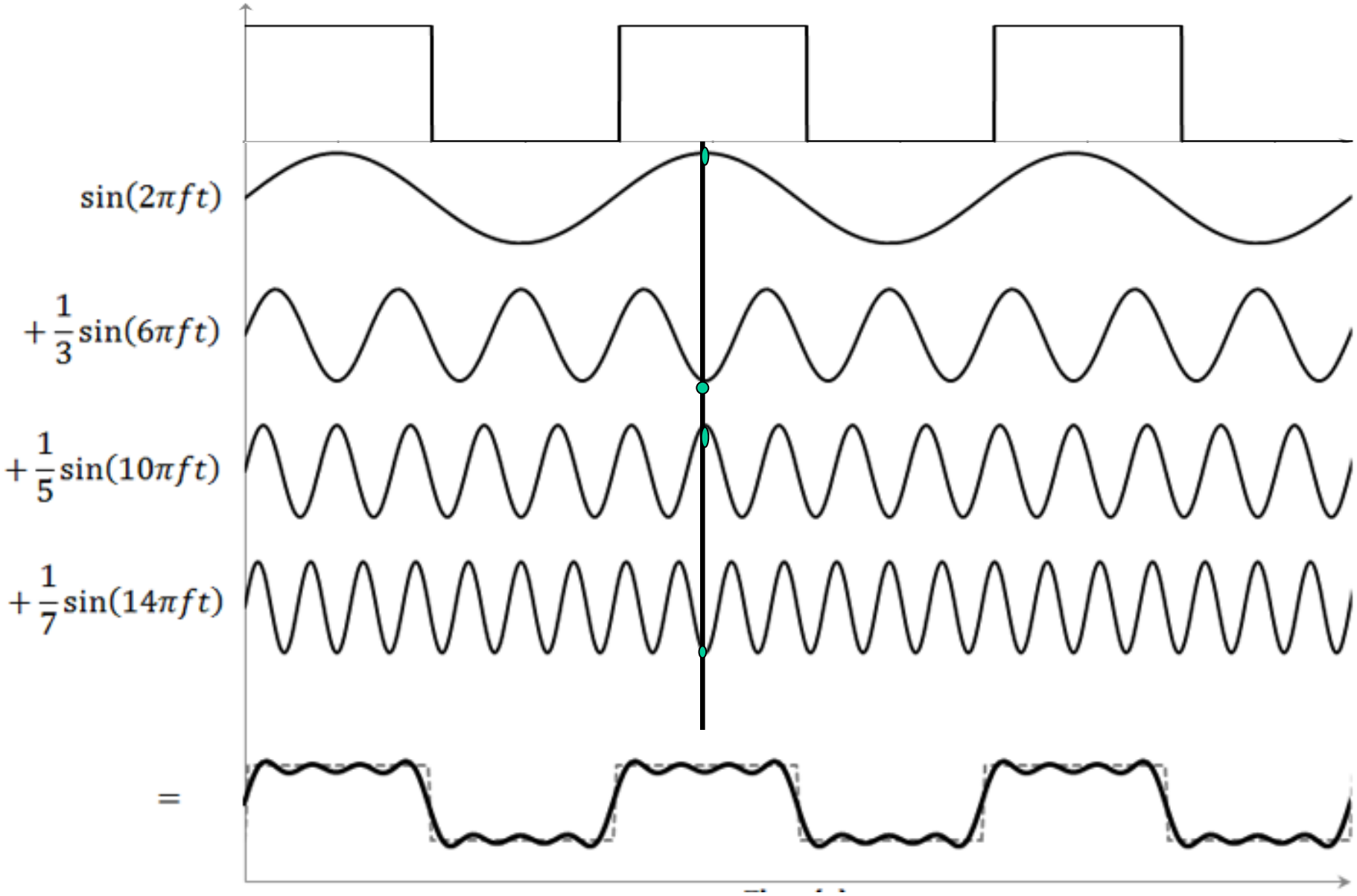
$$f(t) = \cos\left(\frac{2\pi \cdot 5t}{100}\right)$$

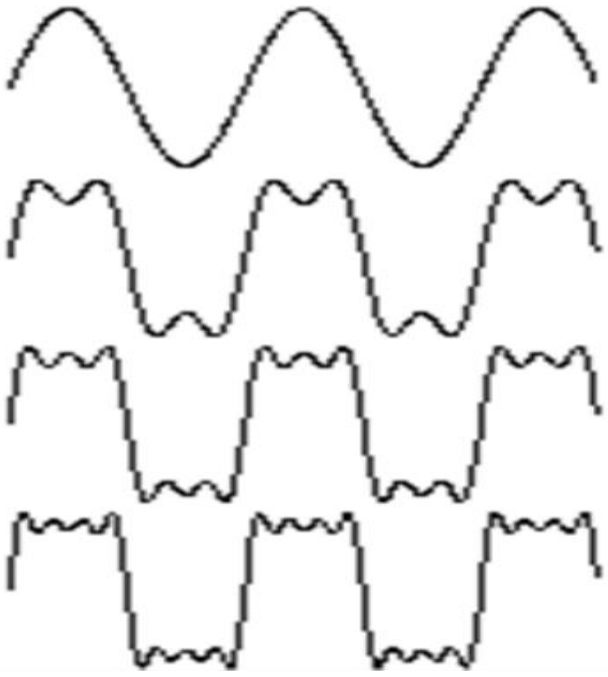
$$+ \cos\left(\frac{2\pi \cdot 25t}{100}\right) + \cos\left(\frac{2\pi \cdot 50t}{100}\right)$$

Fourier provides localization in the frequency domain but no localization in the **time** domain.

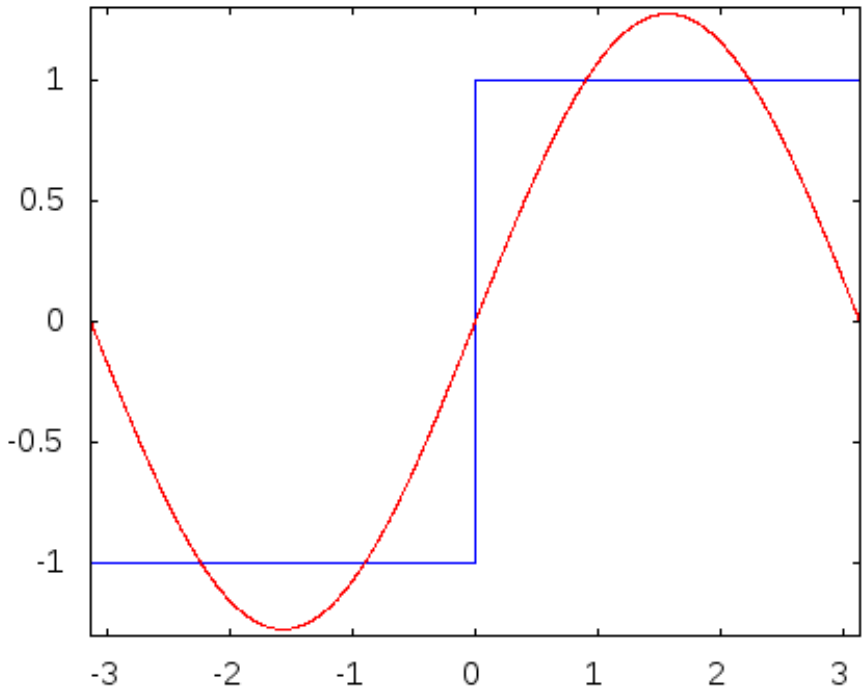


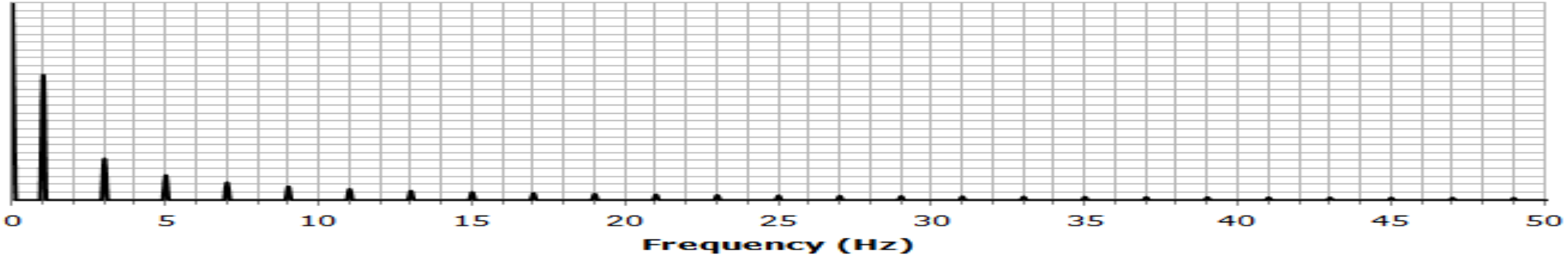
Continuous Square Wave





Animation: Continuous Step-Edge



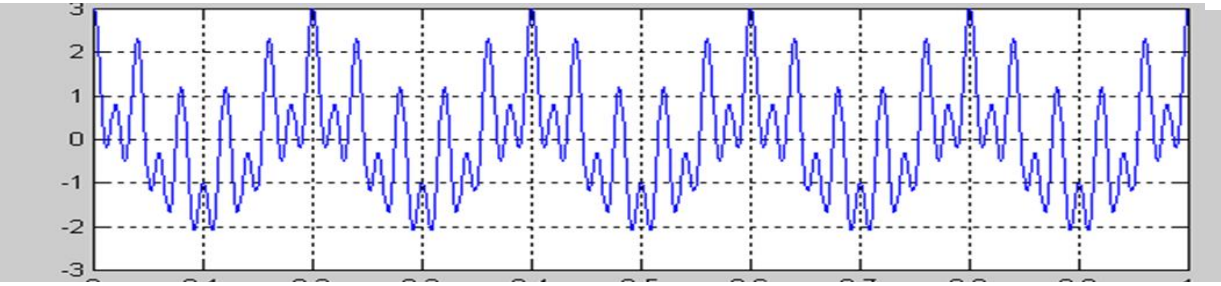


Transform Fourier is (0, 1, 0, 1/3, 0, 1/5, 0, 1/7, …)

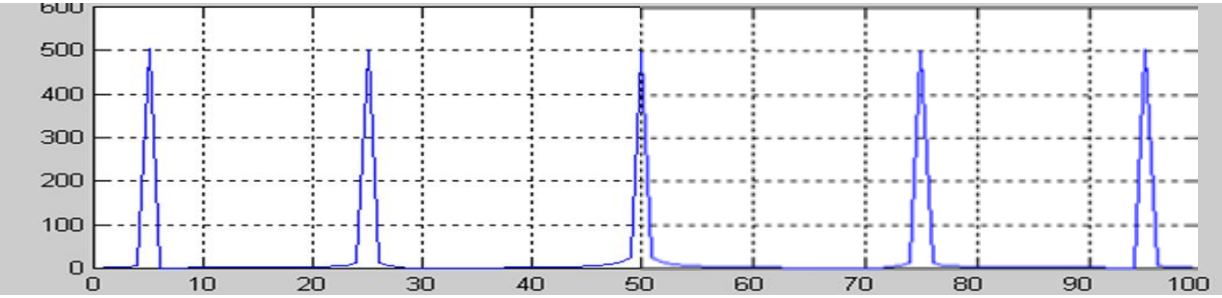
Stationary vs. Non-Stationary Signal

Fixed (Stationary) vs. Variable (Non-Stationary) <u>Characteristics</u> along time:
Mean, Variance, Frequency, etc. (/N ignored in graphs below…)

$\cos(2\pi \cdot 5t) + \cos(2\pi \cdot 25t) + \cos(2\pi \cdot 50t)$

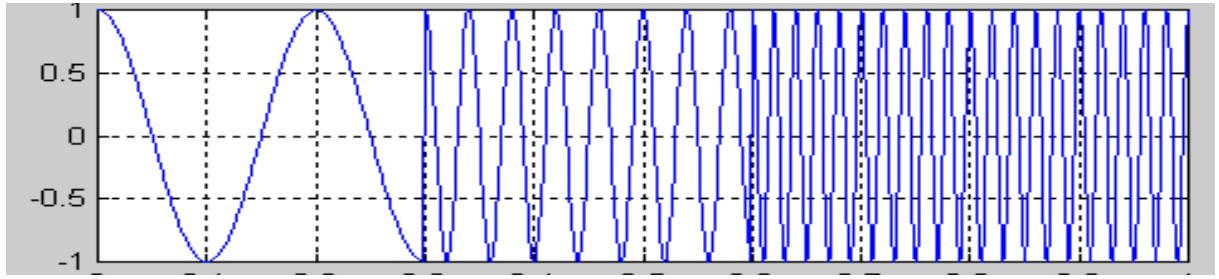


$t \longrightarrow$ Time

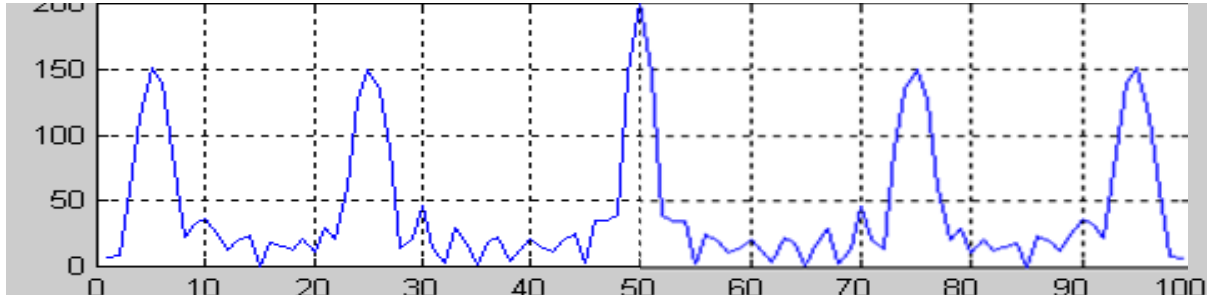


Frequency (Hz)

$\cos(2\pi \cdot 5t)$ **than** $\cos(2\pi \cdot 25t)$ **than** $\cos(2\pi \cdot 50t)$



Time

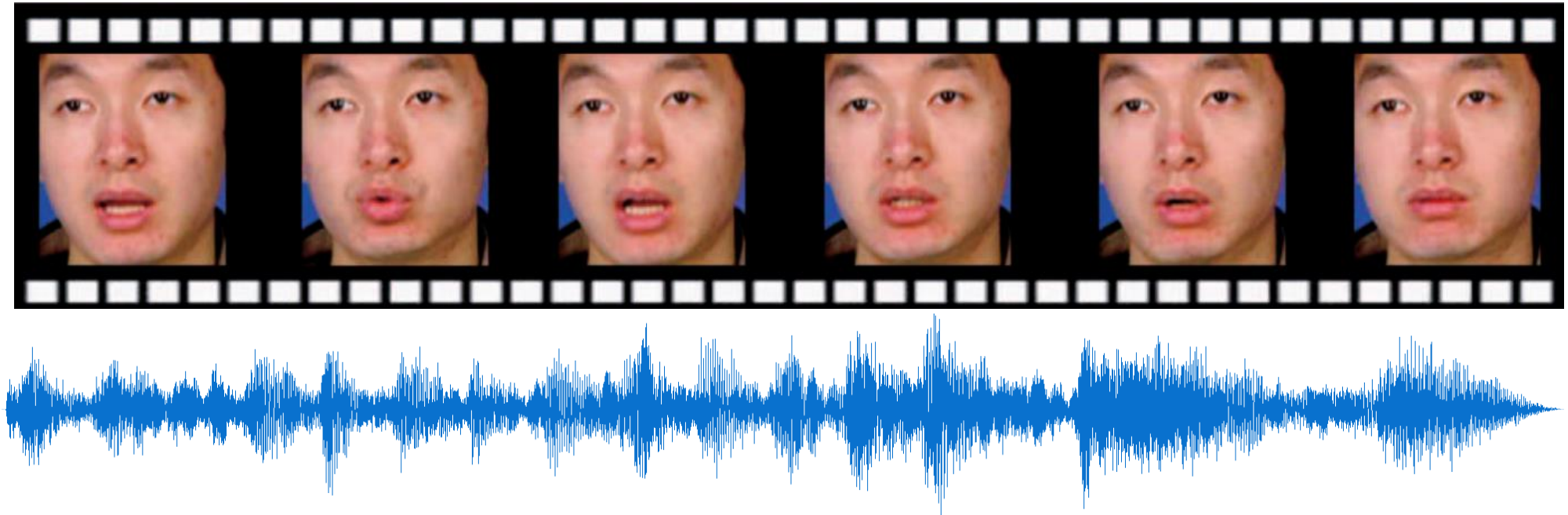


Frequency (Hz)

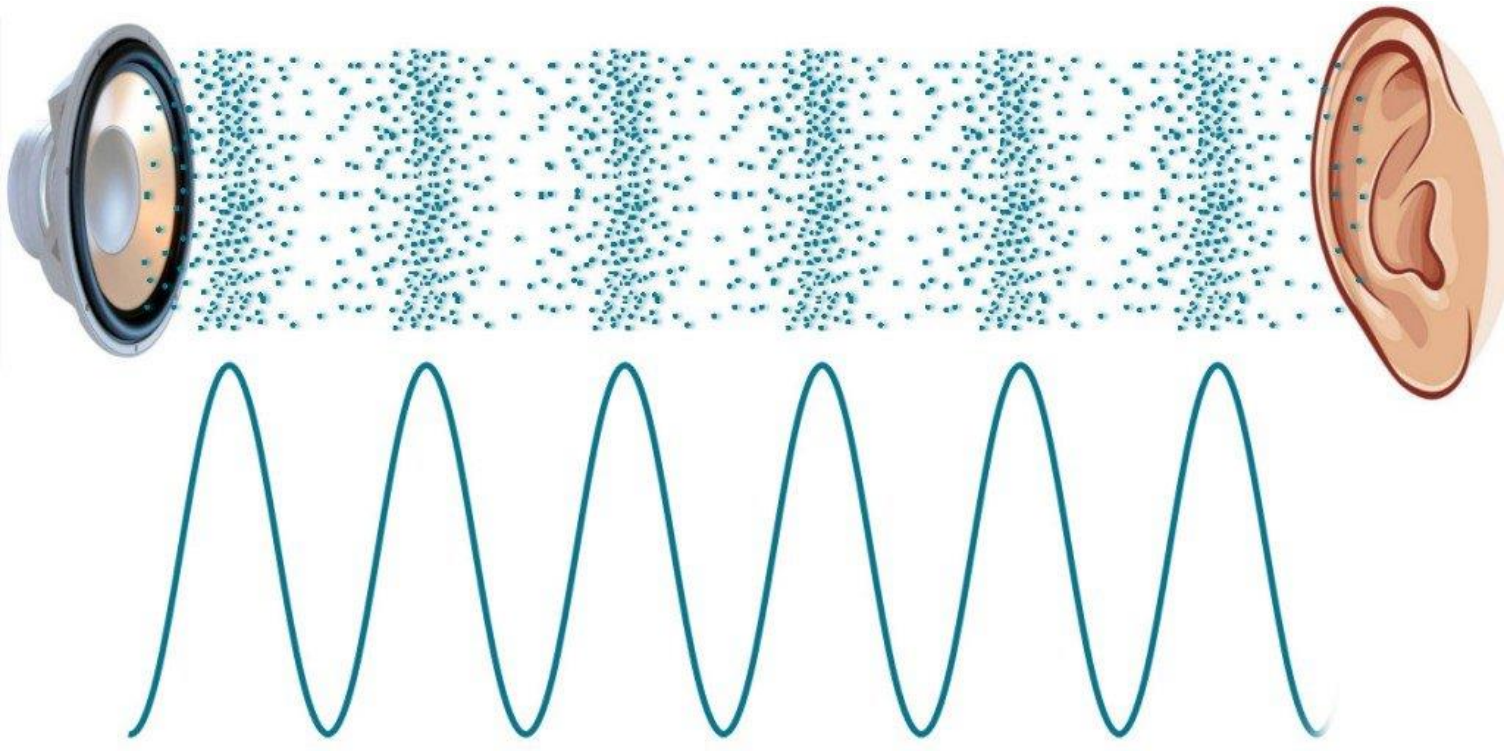
Sound – A Non-Stationary Signal

Why in IMPR? Most video has both frames and sound

- Most sound analysis methods are based on Fourier Transform
- Sound is generated by vibrations creating waves, so Fourier analysis is a natural tool



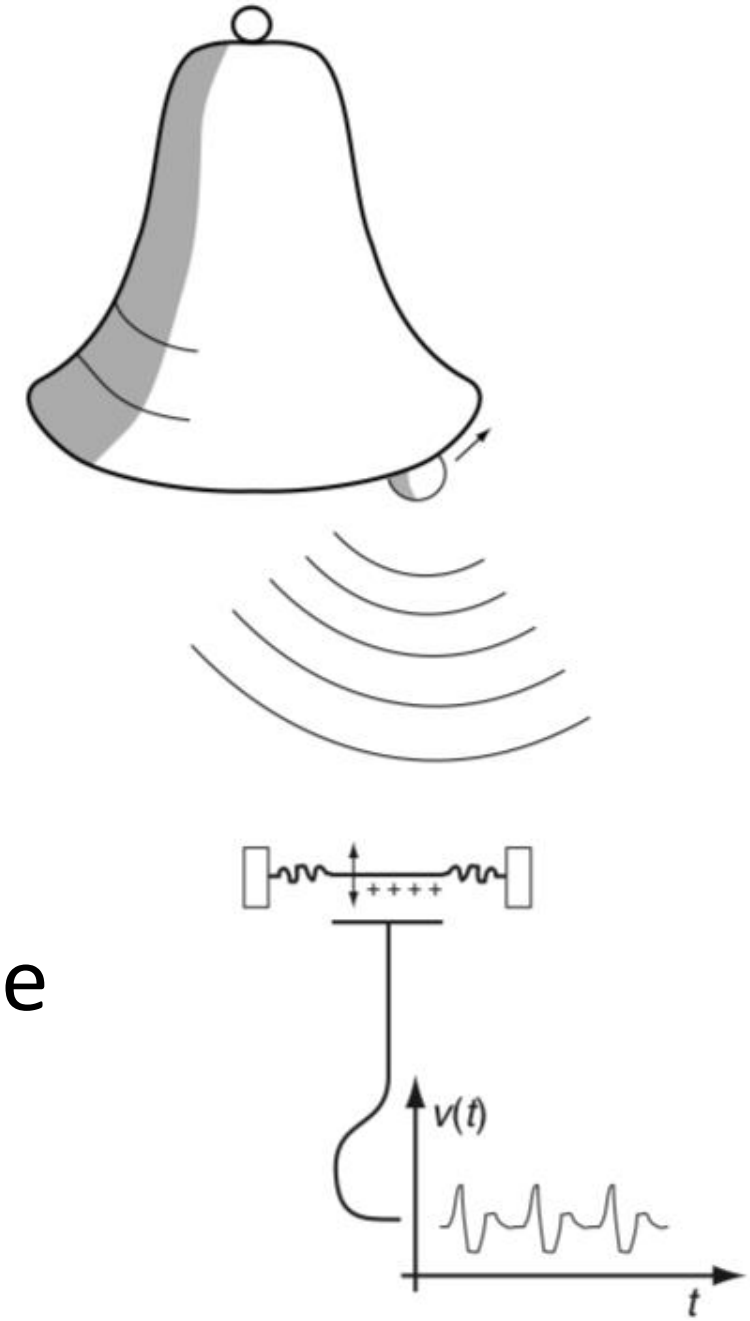
What is sound?



Rapid, tiny variations in air pressure detected by the eardrums

Hearing Range: 50 – 20,000 Hz (Baby. Most adults lose high freq.)

What is sound?



Mechanical vibration

↓

Pressure waves in air

↓

Motion of sensor

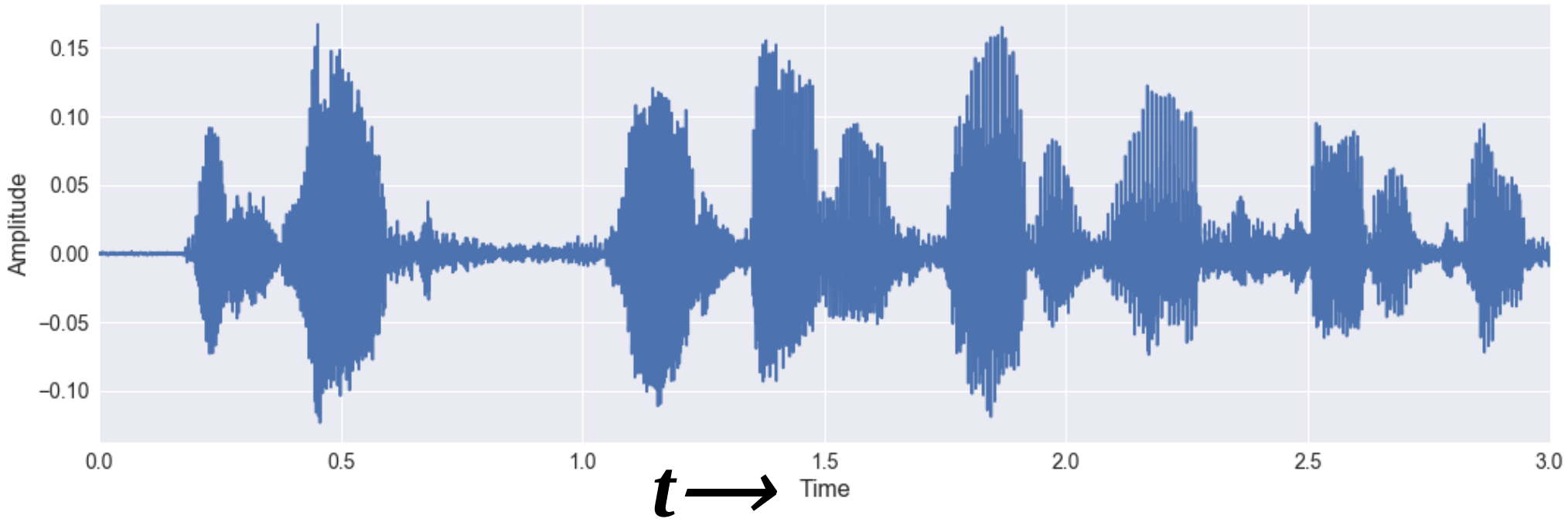
↓

Time-varying voltage

Transducers convert air pressure into voltage

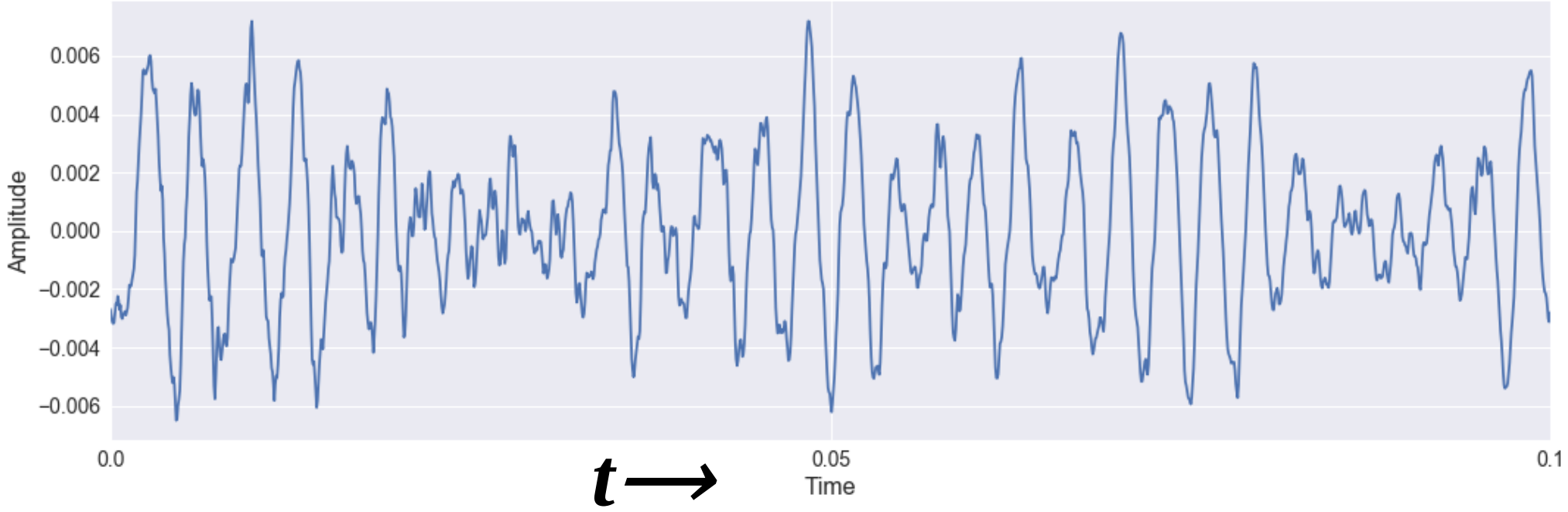
Digitizers convert voltage into a sequence of numbers

Waveform



Amplitude as a function of time (1D signal)

Waveform – zoom in



Amplitude as a function of time (1D signal)

320Hz is clearly dominant – 32 waves in 0.1 Sec.

Waveform – Sampling Rate (SR)

Image equivalent: pixels per line / bits per pixel

- Sound quality depends on sampling rate (samples/second)
- SR should be at least double the maximum frequency (why?)
  - (20 kHz is maximum audible frequency for humans)
- Common sampling rates:
  - Telephone – 8 kHz, Each sample 8-bit [-128 to +127], Max 4kHz
  - AM radio – 22.05 kHz, Max sound frequency 11 kHz
  - Audio CDs – 44.1 kHz, Each sample 16-bit [-32,768 to + 32,767], 22kHz
- Audio CD uses 11 times more bits than Telephone

Waveform – file format and code

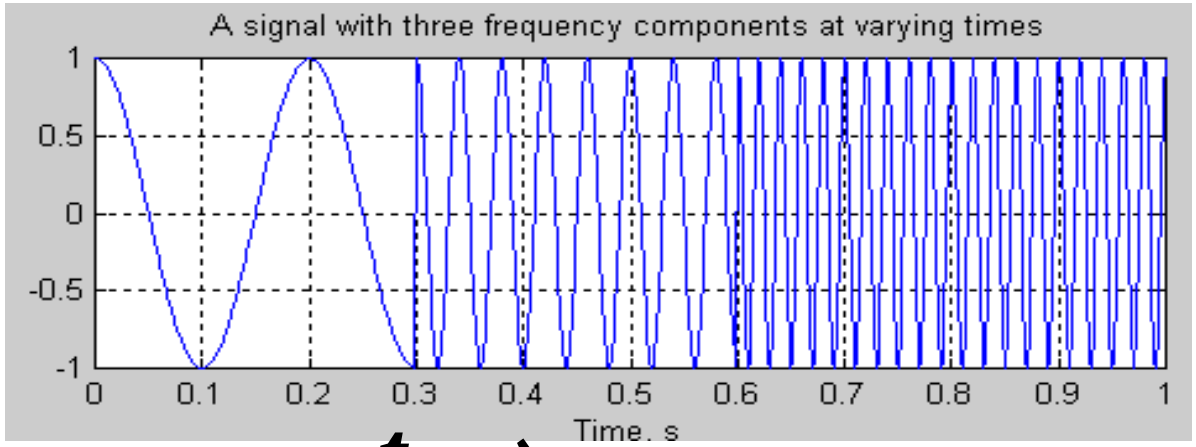
- Common file format: ‘.wav’ with PCM (lossless)
- Can be represented as integers or as floats in [0,1]
- Python libraries:
  - **scipy.io.wavfile** for io
  - **matplotlib** for visualization
  - **librosa** for io, processing, visualization

Audio vs Images vs Video

	Audio (waveform)	Images	Video
Physical phenomenon	Vibrations in air pressure	Light variation across image	
Sensor	Microphone diaphragm	Camera sensor	
Temporal resolution	Sampling rate usually 22.05/44.1 kHz	--	Frame rate usually 24-30 FPS
“Spatial” dimension	1D (time)	2D (space)	3D (Space-time)
Amplitude resolution	“Bit depth” usually 8/16/32-bit	8 bit (bw) / 24 bit (color)	
Dynamic Range	Less of an issue (No human adaptation to sound level)	Very Important (Strong human adaptation to light level, e.g. by changing pupil diameter)	
Common formats	‘.wav’ ‘.mp3’	‘.png’ ‘.jpg’	‘.avi’ ‘.mp4’

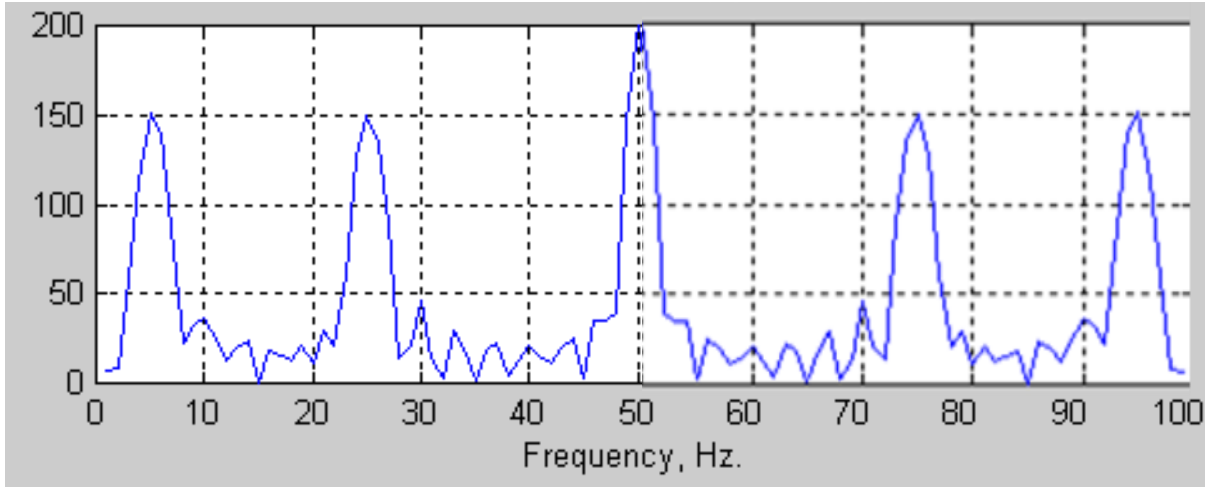
Non-Stationary Signals

Non-stationary signal:
characteristics, e.g. frequency, change along time



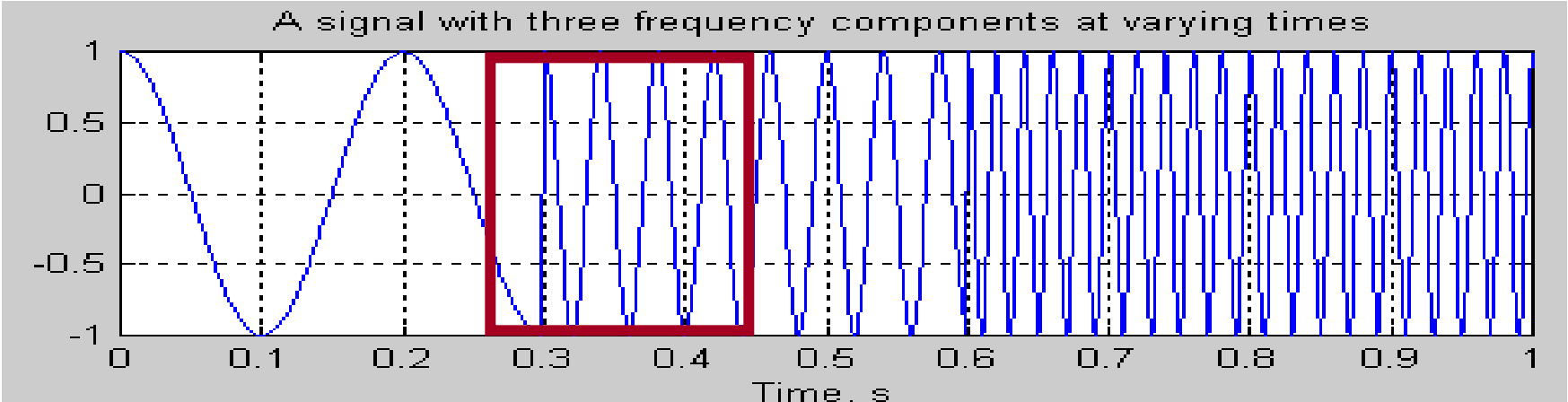
$t \longrightarrow$

Fourier:
Three frequency components, No temporal localization. Very similar to stationary case



Short Time Fourier Transform (STFT)

- Compute the FT at narrow time intervals (i.e., narrow enough to be considered stationary). “window”
- Each FT provides the spectral information of a separate time-slice of the signal, providing simultaneous time (window location) and frequency (FT in window) information.



STFT - Steps

(1) Choose a window $w(n)$ of finite length

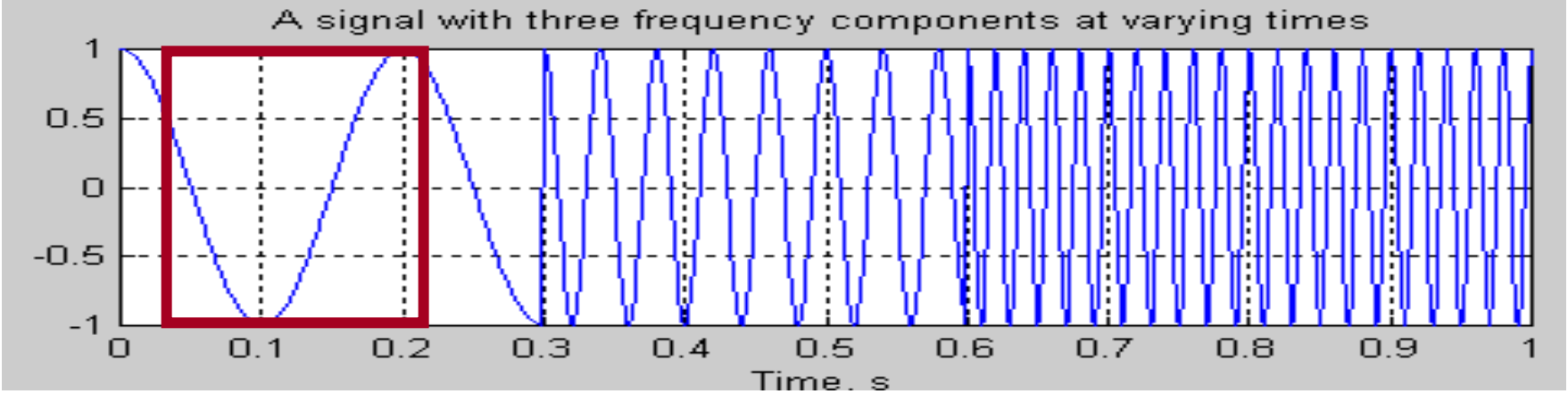
(2) Place the window on top of the signal at t=0

(3) Truncate the signal by multiplication with this window

(4) Compute the FT of the truncated signal, save results.

(5) Incrementally slide the window to the “right”

(6) Go back to step 3, until window reaches the end of the signal



Fourier Transform and STFT

- Fourier $$F(u) = \frac{1}{N}\sum_{x=0}^{N-1} f(x)\, e^{\frac{-2\pi i u x}{N}} \qquad f(x) = \sum_{u=0}^{N-1} F(u)\, e^{\frac{2\pi i u x}{N}}$$

- STFT (for every n) $$F(n,u) = \sum_{m=0}^{N-1} f(m + nH)\, w(m) e^{\frac{-2\pi i u m}{N}}$$

  <u>where</u>: n is time; N is number of frequencies; H is hop size.

- ISTFT (Inverse)

$$f(n) = K \sum_{u=0}^{N-1} F(n,u) e^{\frac{2\pi i u n}{N}}$$

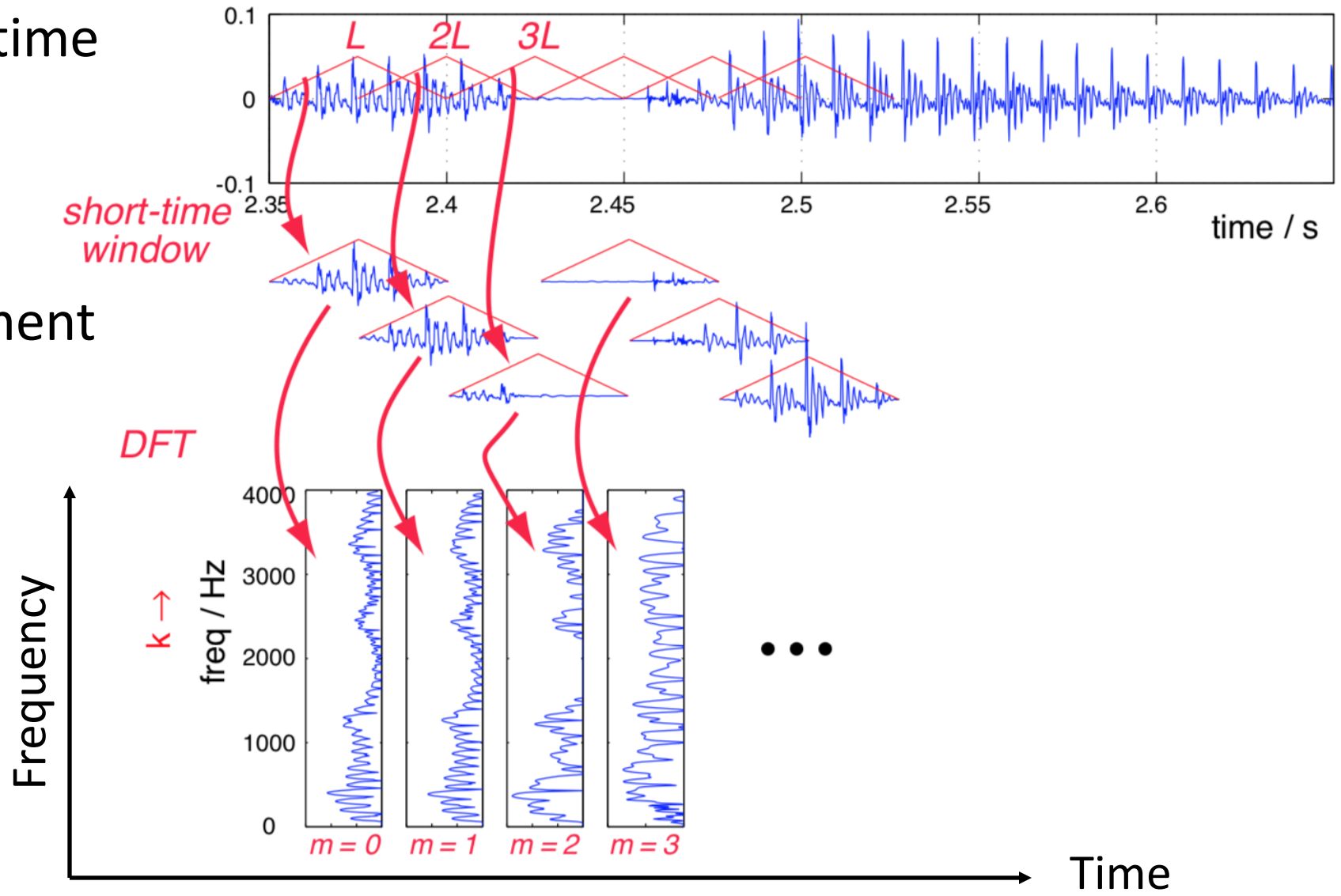
For a carefully selected window w, hop H, *and normalization* K.

Short-time Fourier Transform (STFT)

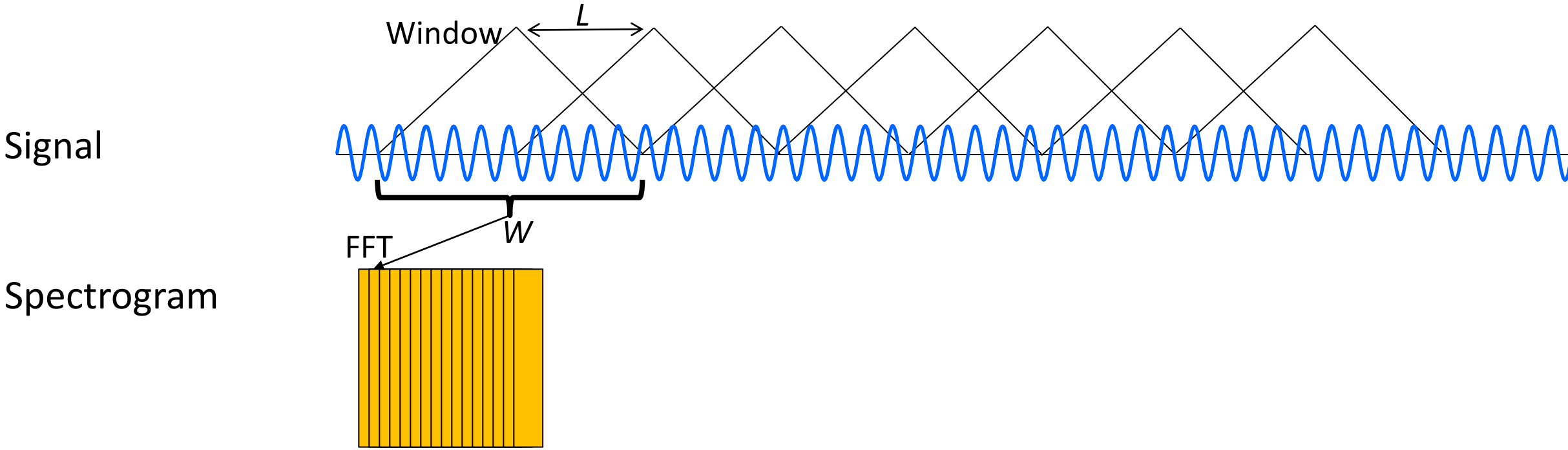
Break sound into short-time segments

Multiply each time segment by window function

Calculate DFT on each time segment

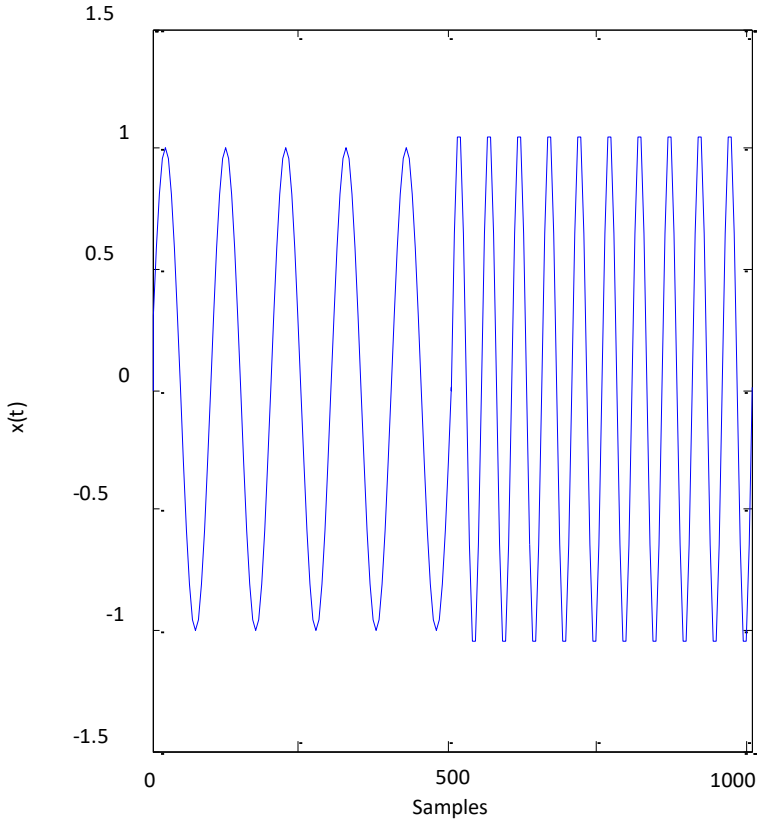


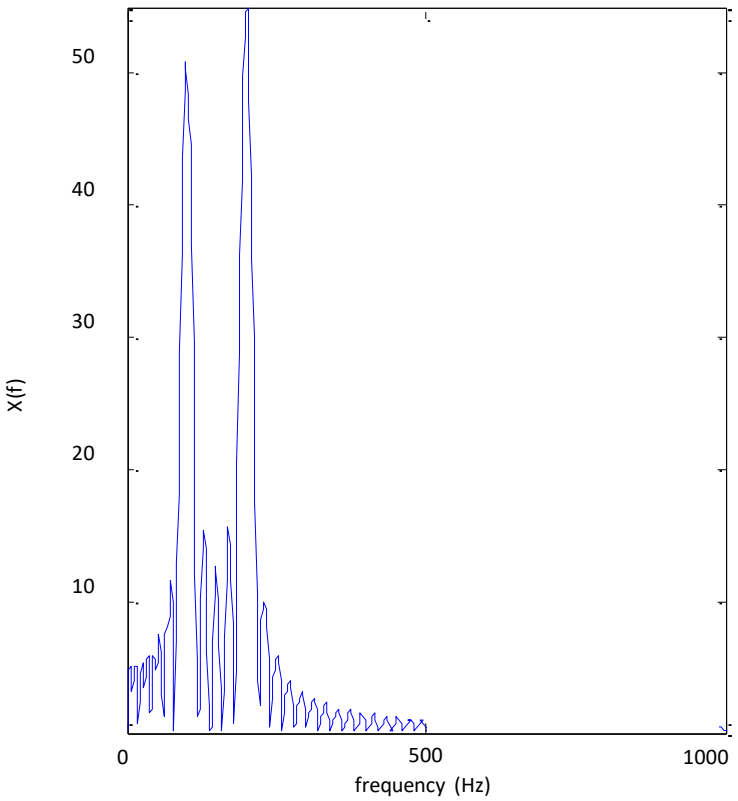
Windowing

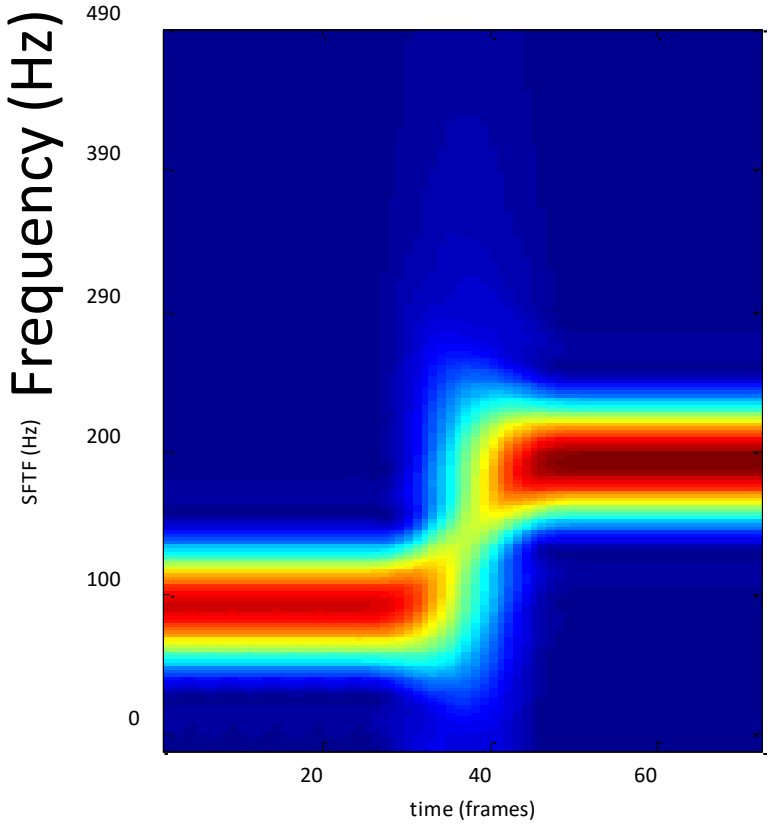


- Window Shape: Symmetric, Unimodal: E.g. Gaussian, Triangle
- Window Length (*W*): The number of non-zero coefficients in *w*
- L – the number of samples between the successive windows
- Window Overlap: *W-L*

Spectrogram







Signal

Low Freq. followed by High Freq

Fourier

Frequencies discovered, no location

Spectrogram

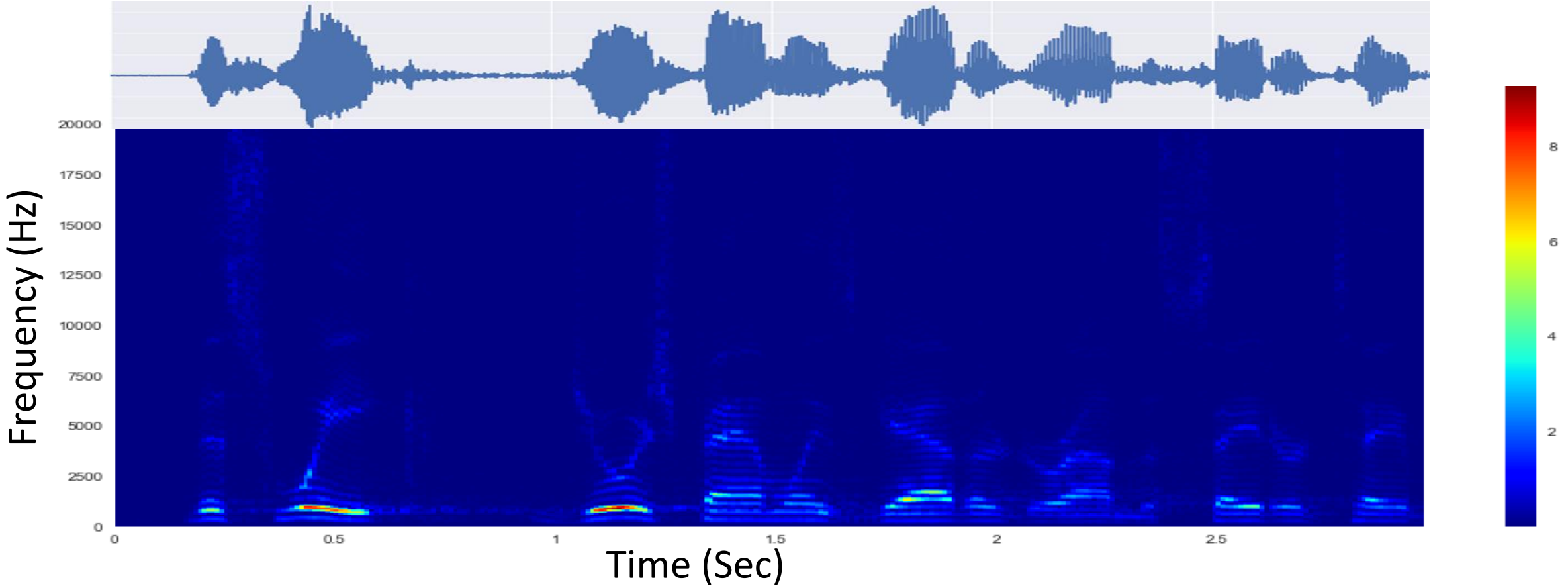
Frequencies discovered at location

Spectrogram

Spectrograms are 2D arrays of complex numbers in Time/Frequency array

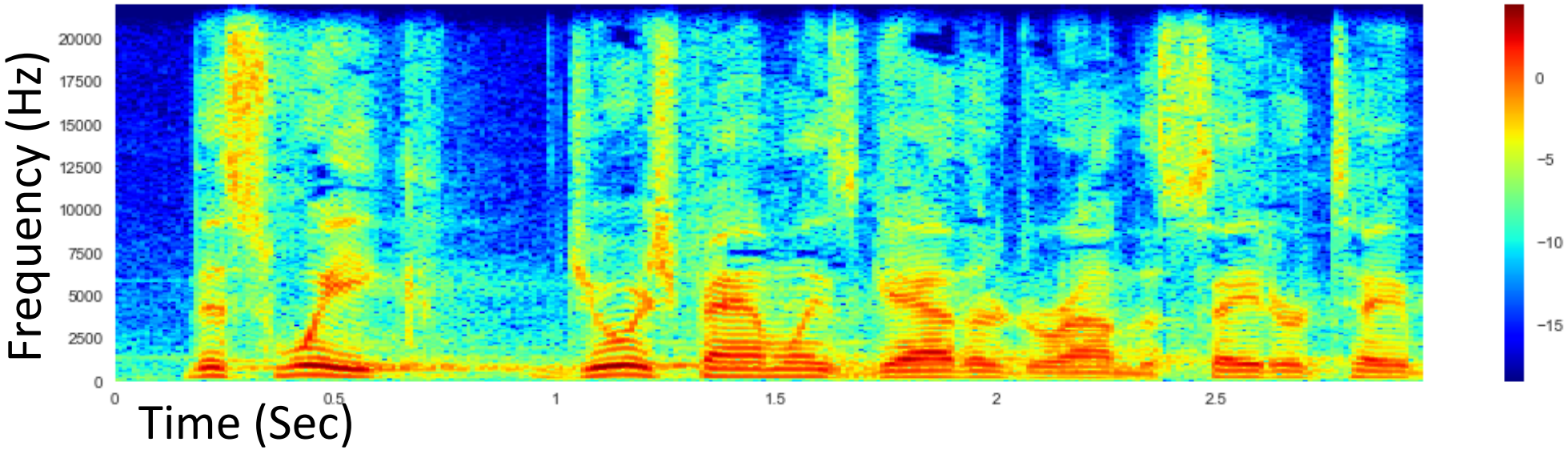
Visualized (and usually processed) using magnitude/energy

Localize energies in time and frequency



Spectrogram **Visualization** – Log Scale

Log scale



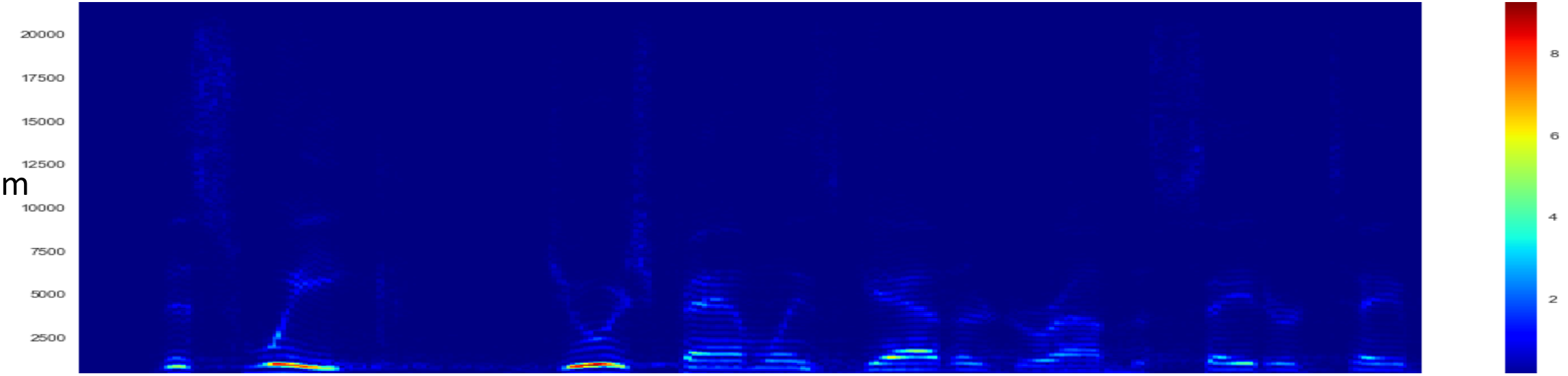
1. Compute $\log(|F(u)| + 1)$
2. Scale to full range

Log Scale of [0..100]

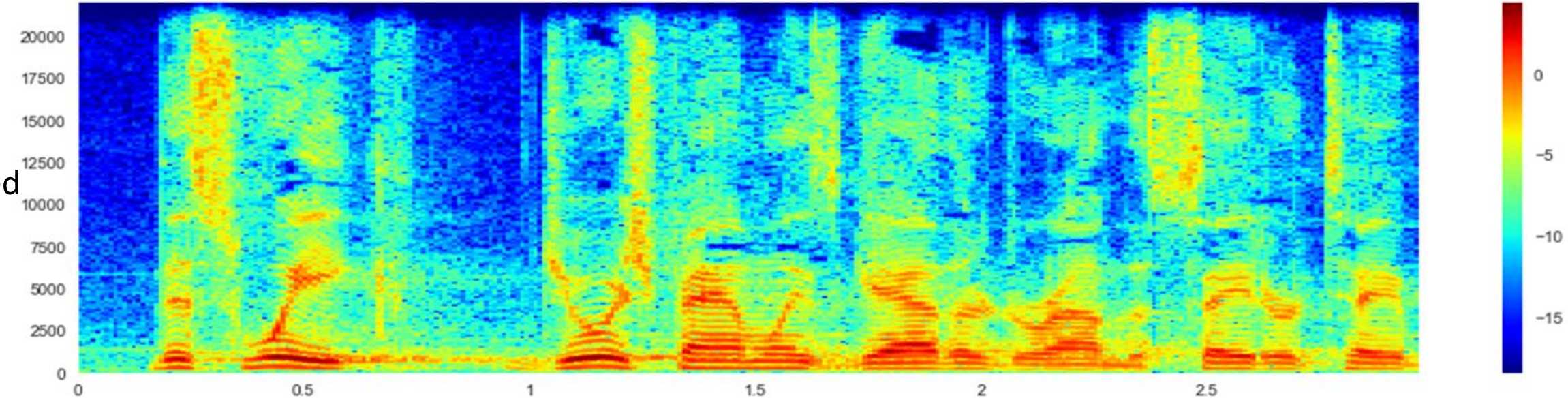
Original *F*	100	4	2	1	0
Log (*1+ F*)	4.62	1.61	1.01	0.69	0
Scaled to 100	100	35	22	15	0

Signal

Original Spectrogram



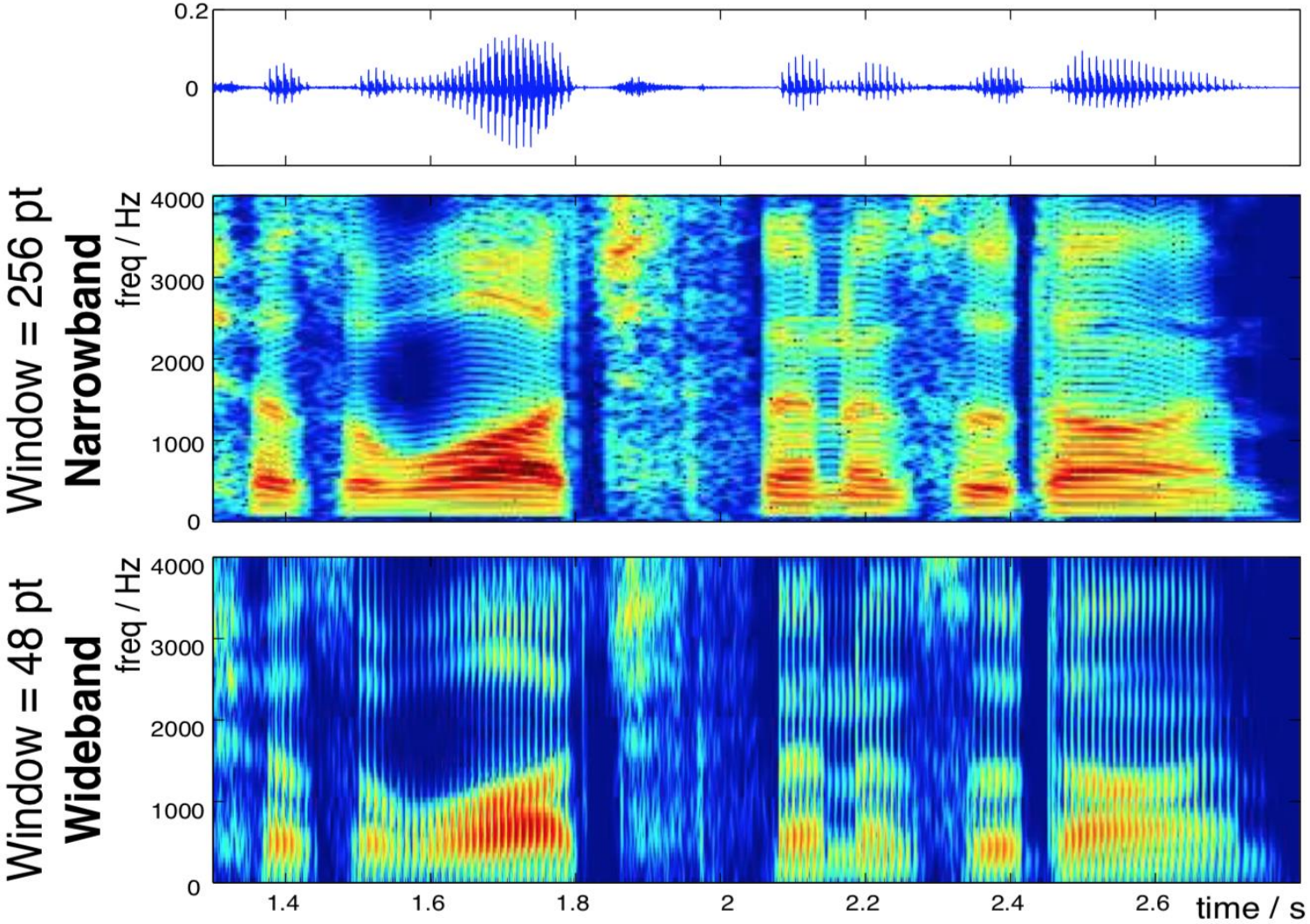
Log compressed



Time-frequency tradeoff

Longer window $w[n]$ gains frequency resolution

at cost of time resolution



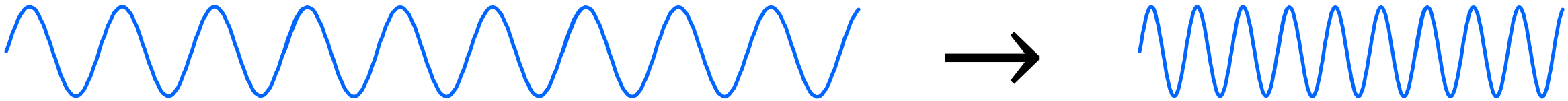
Spectrograms in Practice

- Window length and overlap ("hop length") are hyperparameters
- Common parameters: 25 milliseconds (1/40 Sec) window, 10 ms hop

Sampling KHz	Window length		Hop length	
8	200	25 ms	80	10 ms
22.05	551	25 ms	220	10 ms
44.1	1102	25 ms	441	10 ms

Simple Application – Speech Fast Forward

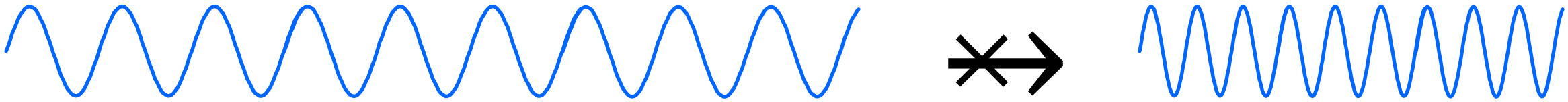
- Naïve Fast Forward of Speech: Play speech at faster speed



2 seconds (5Hz) 1 second (10Hz)

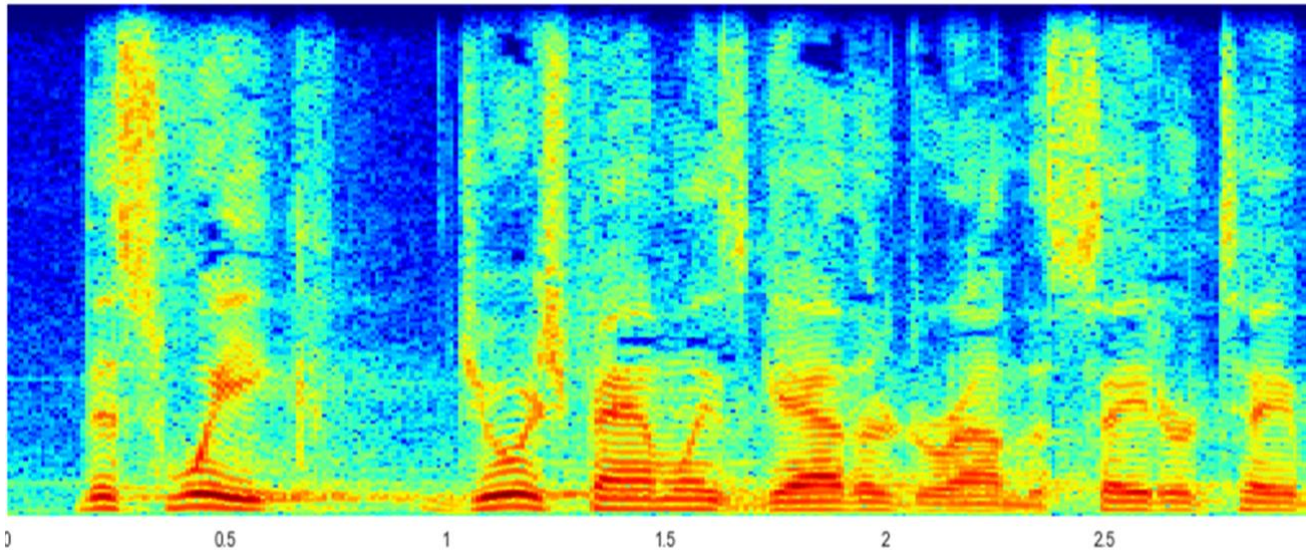
- Side effect - Increase frequency – pitch becomes higher
- Implementation options:
  1. If the original sample rate was 8KHz, tell the system it is 16 KHz. 2 sec of speech will be played at 1 second.
  2. Throw away every other sample point.

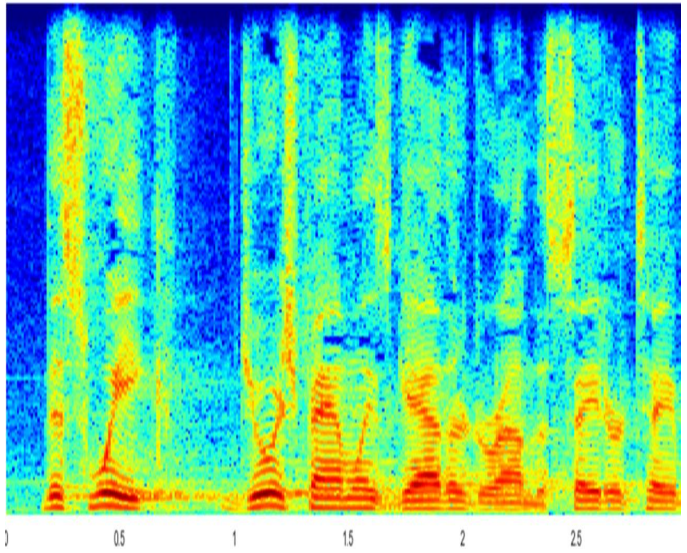
Fast Forward with Spectrogram



2 seconds 1 second

- Create a spectrogram, sample **the spectrogram**, and reconstruct the speech. Frequencies (and pitch) of sound will remain unchanged:

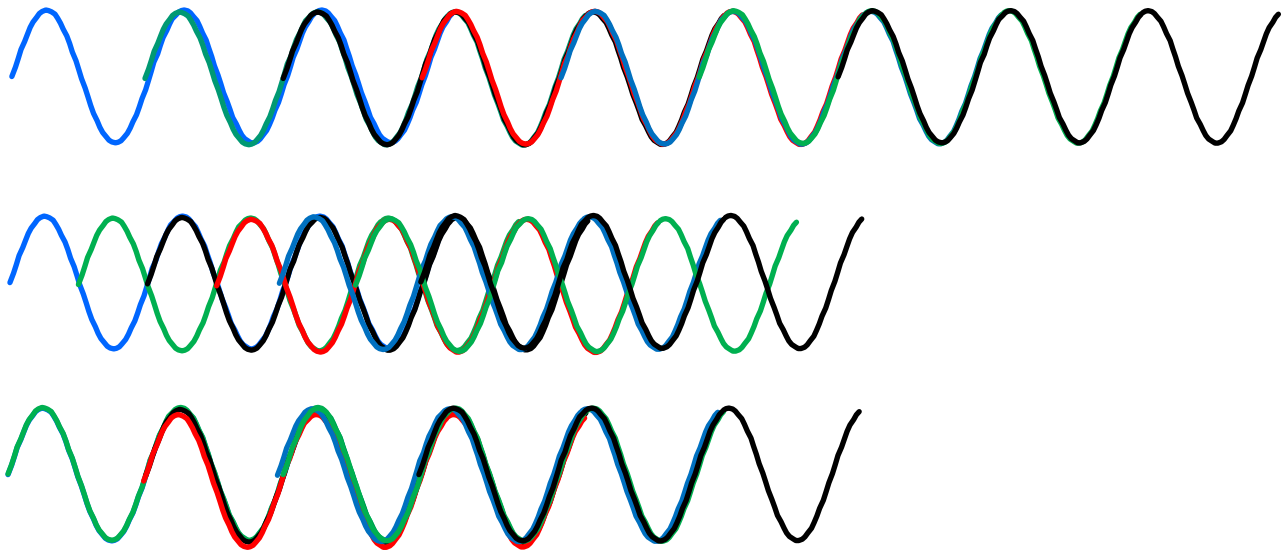




2 seconds 1 second

Spectrogram Scaling – Phase Consistency

- Original Windows
- Windows after Spectrogram scaling
- After phase correction



- What will happen if we do not adjust the phases when windows appear at a different time?
- Green and Red are offset by half wavelength!
- Waves can cancel each other. Need to adjust phase.

2nd Exercise <u>before ChatGPT</u>:

Change Speed of Speech without changing Pitch

<u>If you have spare time:</u> Implement the change of sound speed without changing the pitch.

How does an Equalizer Work?

- Turn sliders into a Column Vector
- Multiply with each spectrogram column
- Log-Scale also the frequencies (Mel Spectrogram)

1.4
1.2
1
0.9
0.7
0.7
0.9
1
1.2
1.4