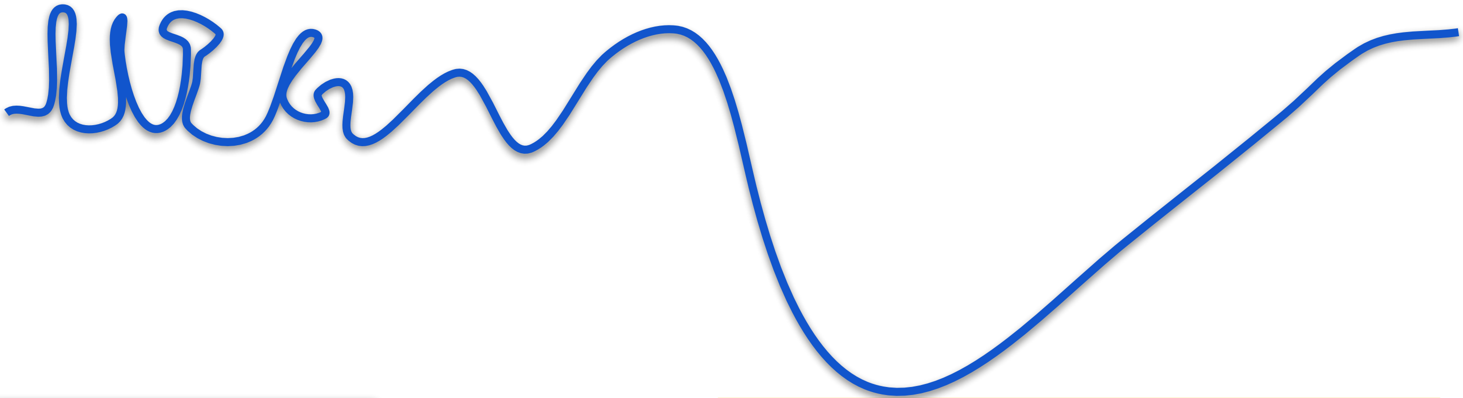
# Computing with Signals



DA 623

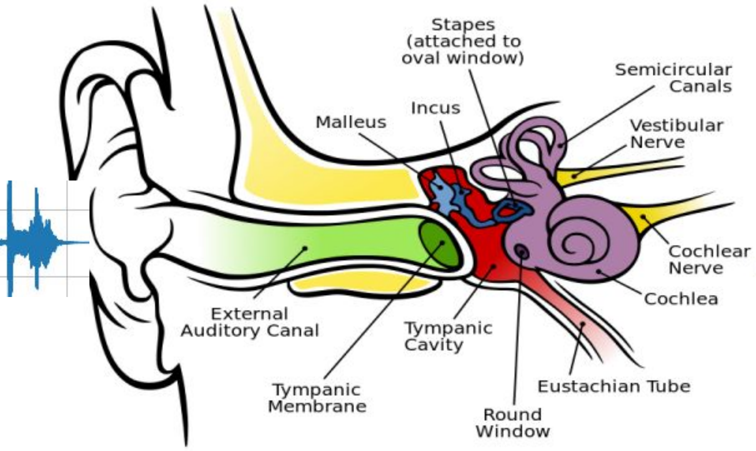
Jan - May 2024

IIT Guwahati

Lecture-21

Instructors: Neeraj Sharma

# Human hearing mechanism




Courtesy: Chittla L. Brockmann

(a)

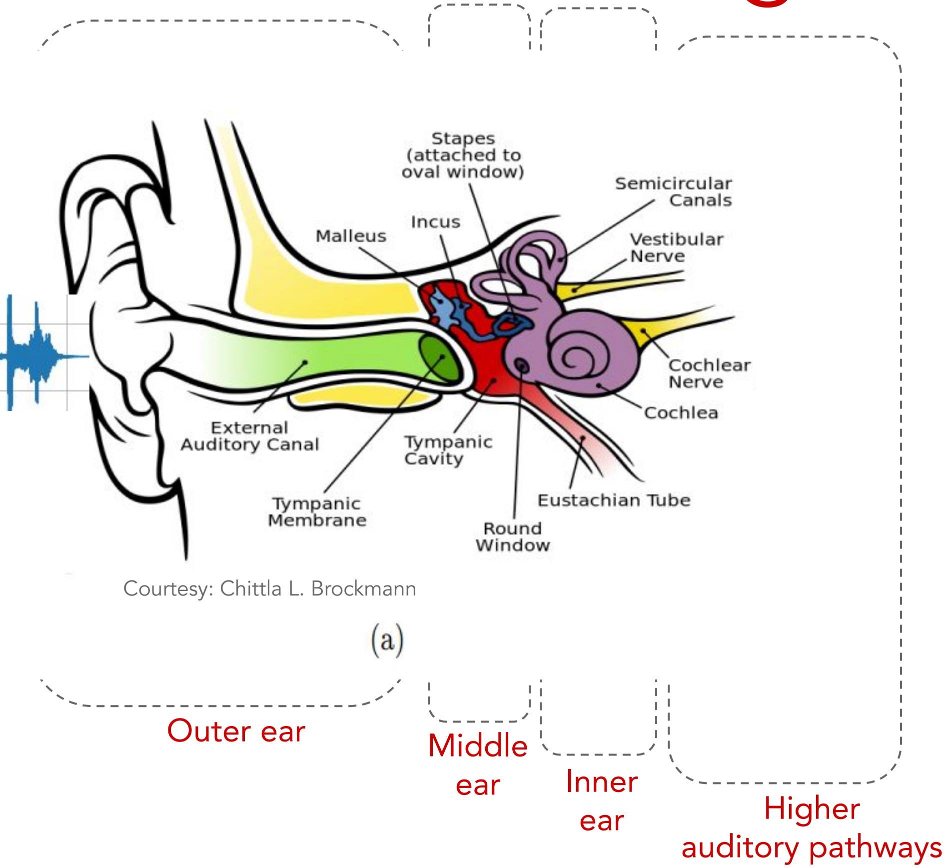
Outer ear

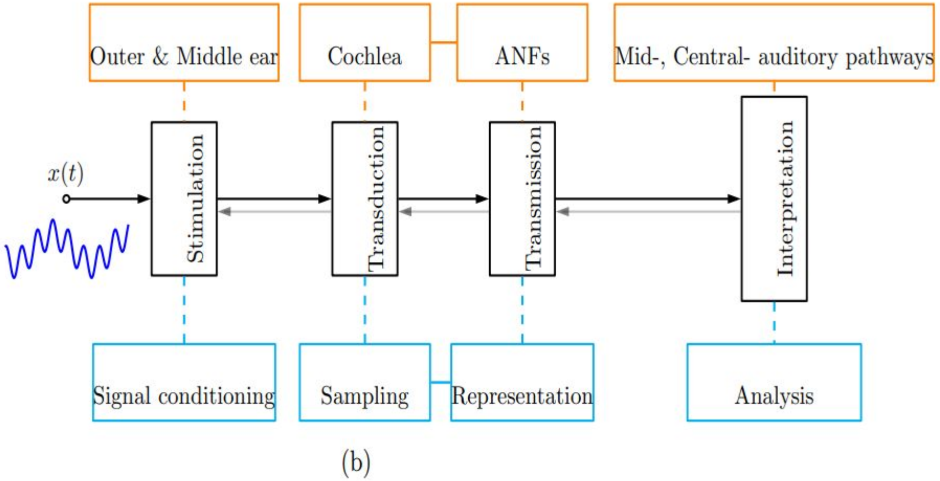
Middle
ear

Inner
ear

Higher
auditory pathways

# Human hearing mechanism

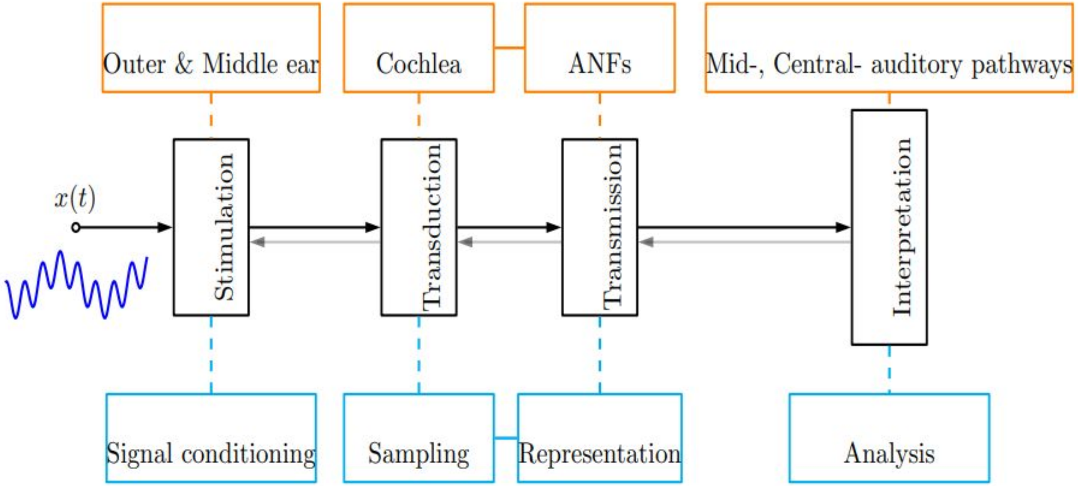



Data processing pipeline modeling hearing

Ref: N.K. Sharma, Information-rich sampling of time-varying signals, PhD Thesis, Indian Institute of Science, 2018

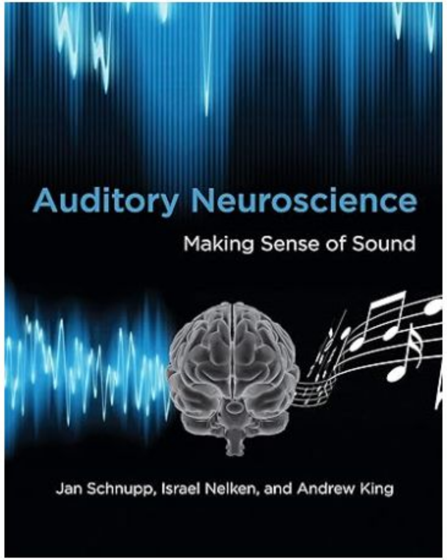
# Human hearing mechanism




(b)

Data processing pipeline modeling hearing

# Perceiving sound




Chap. 1

Why do things sound the way they do

https://auditoryneuroscience.com/sites/default/files/Schnupp_FM_Ch1.pdf

# First commercial **speech recognizer**

- The first machine that recognized speech was a toy from the 1920s
- Radio Rex was a celluloid dog that moved when a spring was released by 500 Hz acoustic energy
- Since 500 Hz is roughly the first formant of the vowel [eh] in "Rex", Rex seemed to come when he was called (David, Jr. and Selfridge, 1962)

# Sound signal as a file

SoXI – Sound eXchange Information, display sound file metadata

```
$ soxi file_name.wav
```

```
Input File     : 'nMIOAh7qRFf3pqbchclOLKbPDOm1_heavy_cough.wav'
Channels       : 1
Sample Rate    : 16000
Precision      : 16-bit
Duration       : 00:00:03.65 = 58378 samples ~ 273.647 CDDA sectors
File Size      : 117k
Bit Rate       : 256k
Sample Encoding: 16-bit Signed Integer PCM
```

# Sound signal as a file

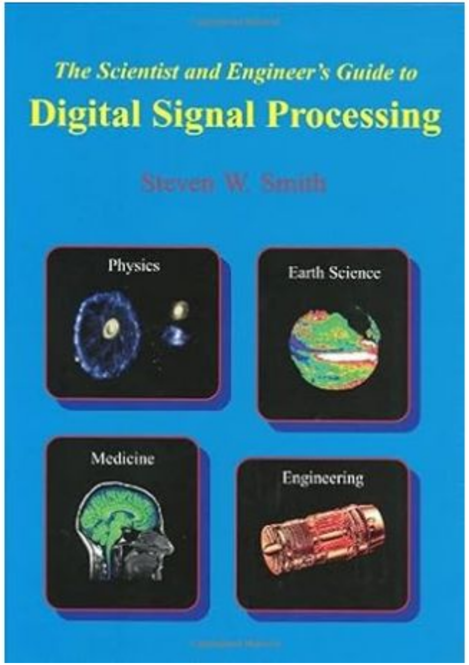
SoXI – Sound eXchange Information, display sound file metadata

```
$ soxi file_name.wav
```

note: a sound signal captured and stored is referred as **Audio**

```
Input File     : 'nMIOAh7qRFf3pqbchclOLKbPDOm1_heavy_cough.wav'
Channels       : 1
Sample Rate    : 16000
Precision      : 16-bit
Duration       : 00:00:03.65 = 58378 samples ~ 273.647 CDDA sectors
File Size      : 117k
Bit Rate       : 256k
Sample Encoding: 16-bit Signed Integer PCM
```

# Audio Processing




Chap. 22

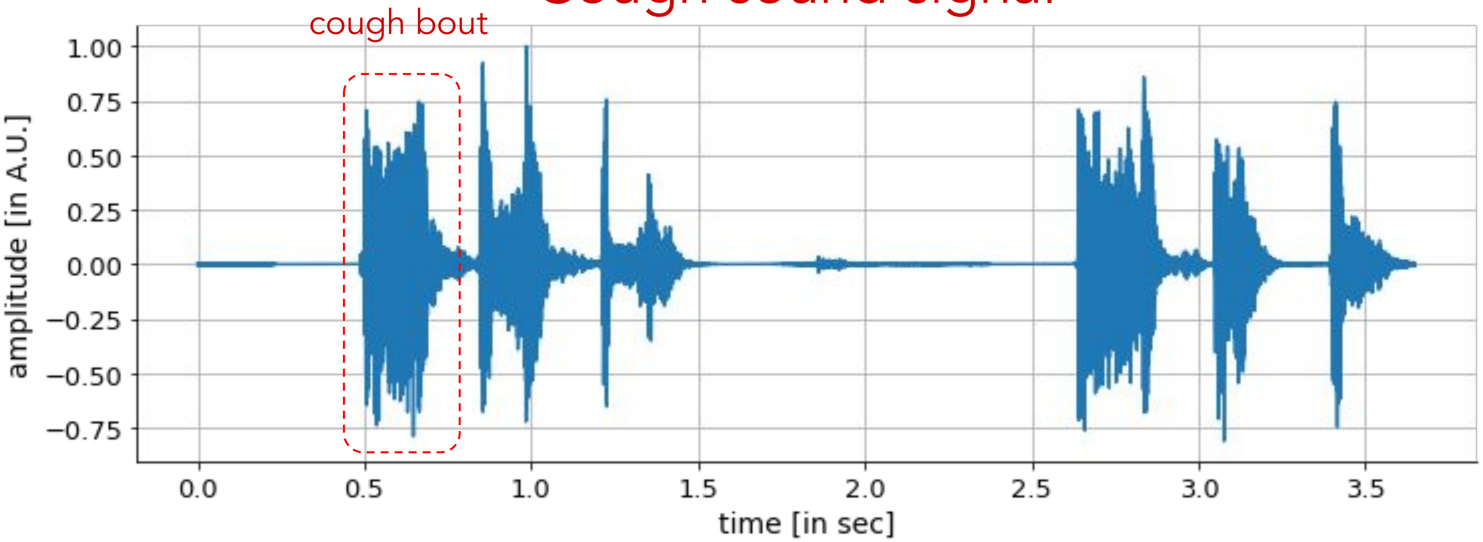
Audio Processing in The Scientist and Engineer's Guide to Digital Signal Processing By Steven W. Smith

http://www.dspguide.com/CH22.PDF

# Audio as a time-series

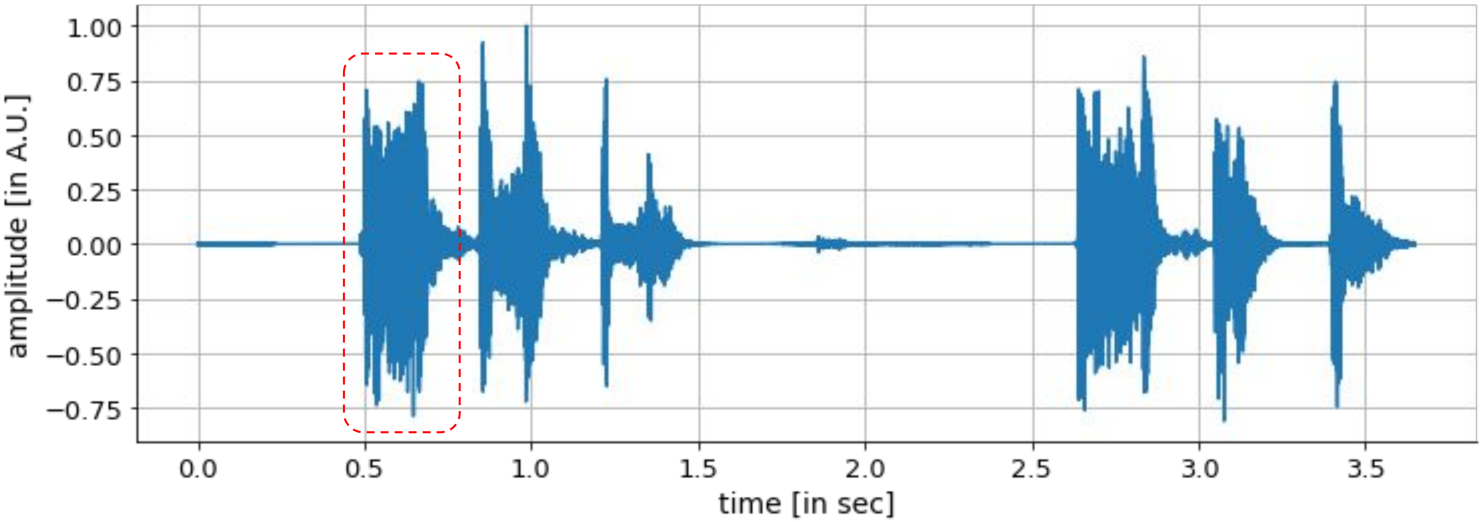
```
fs = 16000
fname = 'nMIOAh7qRFf3pqbchclOLKbPDOm1_heavy_cough.wav'
dname = './my_data/'
# load
x, sr = librosa.load(dname+fname,sr=16000)
x = x/max(np.abs(x))
times = np.arange(0,len(x))/fs
```

## Cough sound signal

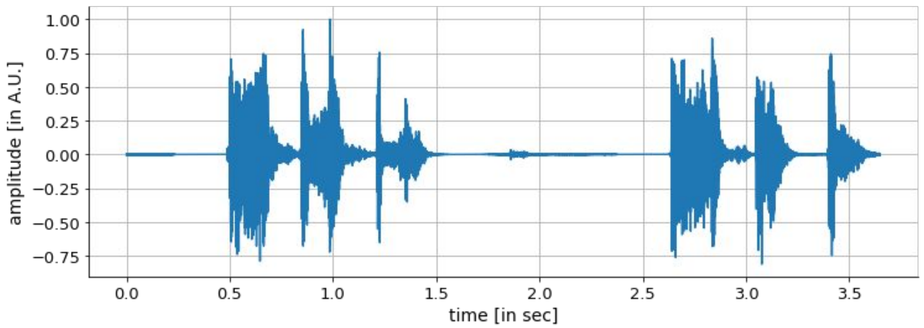


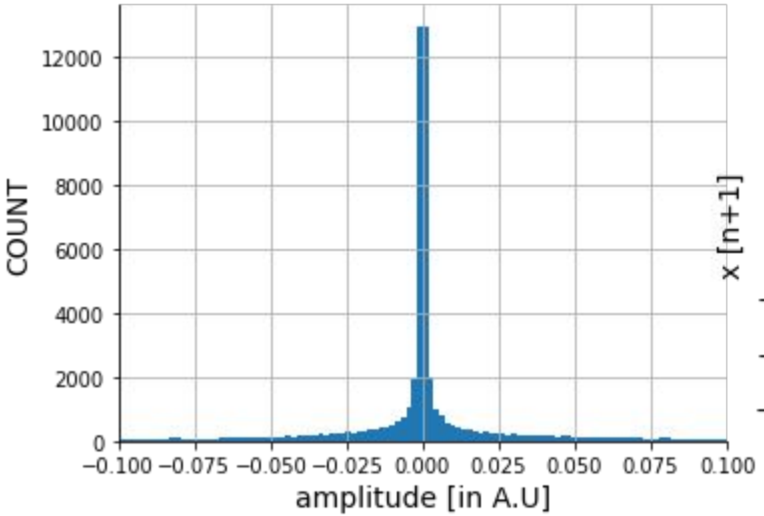
# Audio as a time-series

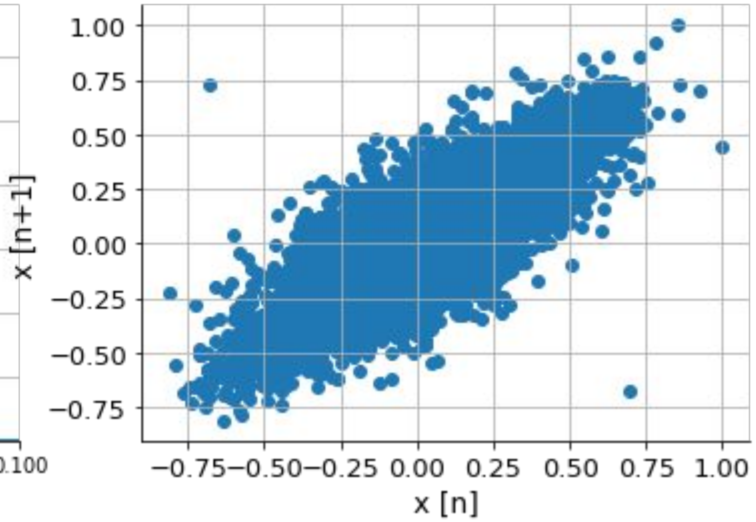
```
fs = 16000
fname = 'nMIOAh7qRFf3pqbchclOLKbPDOm1_heavy_cough.wav'
dname = './my_data/'
# load
x, sr = librosa.load(dname+fname,sr=16000)
x = x/max(np.abs(x))
times = np.arange(0,len(x))/fs
```



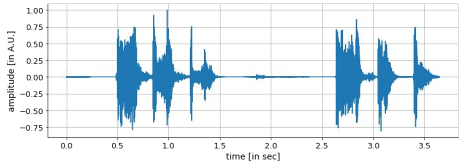
# Audio as a time-series



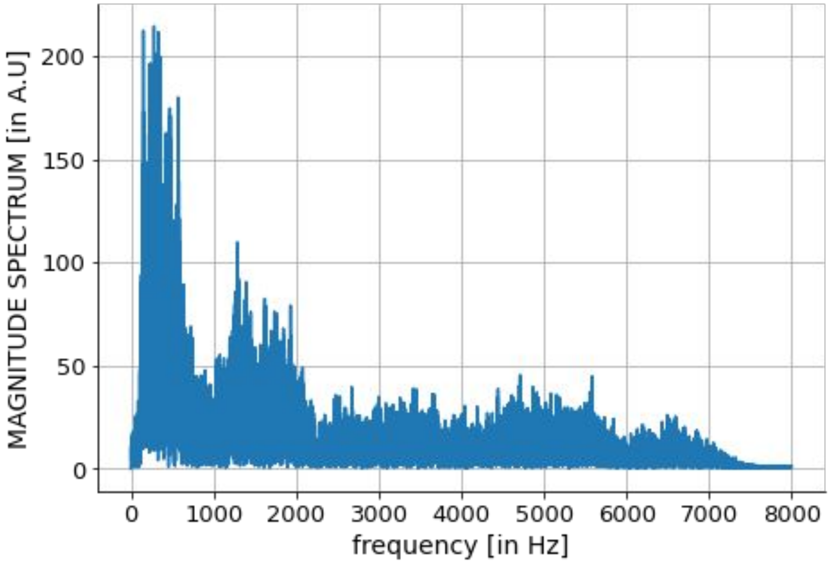



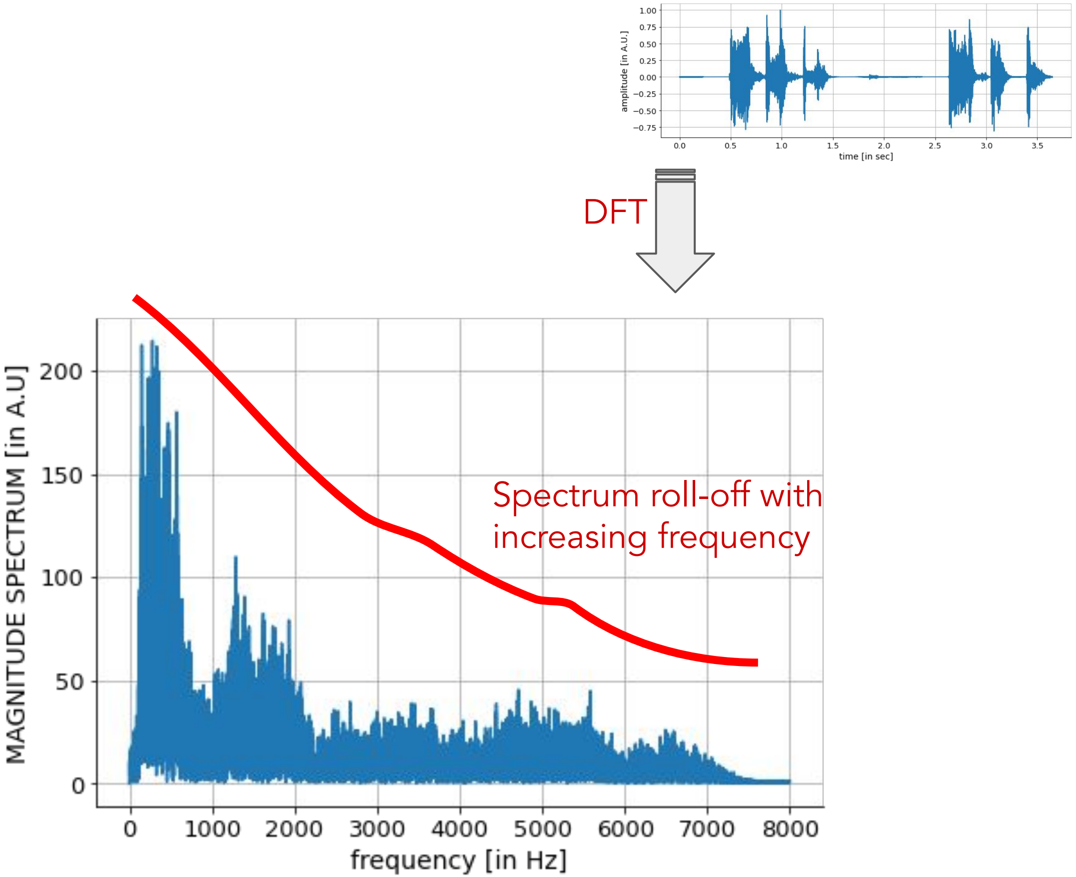
# Audio as a frequency spectrum



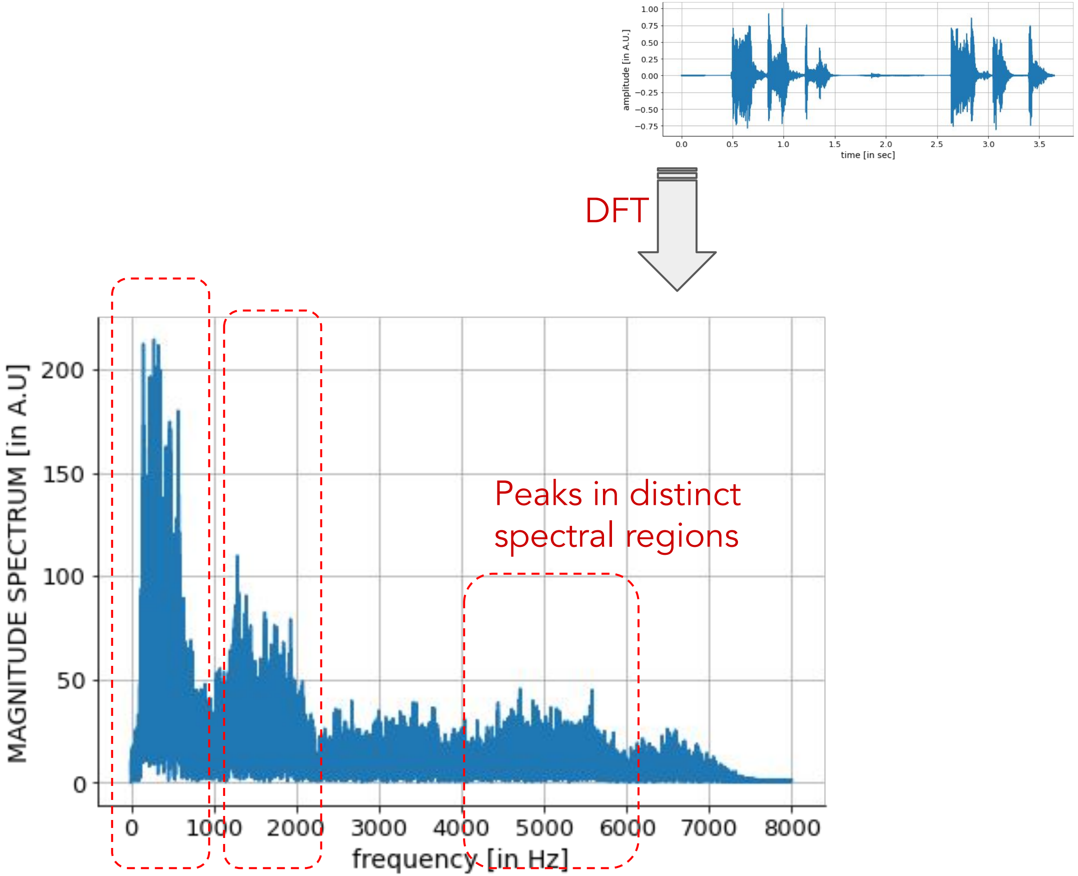

Magnitude spectrum
obtained using DFT



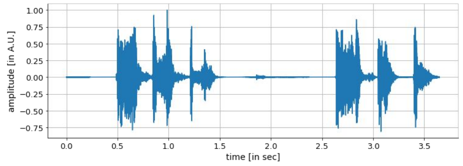
# Audio as a frequency spectrum



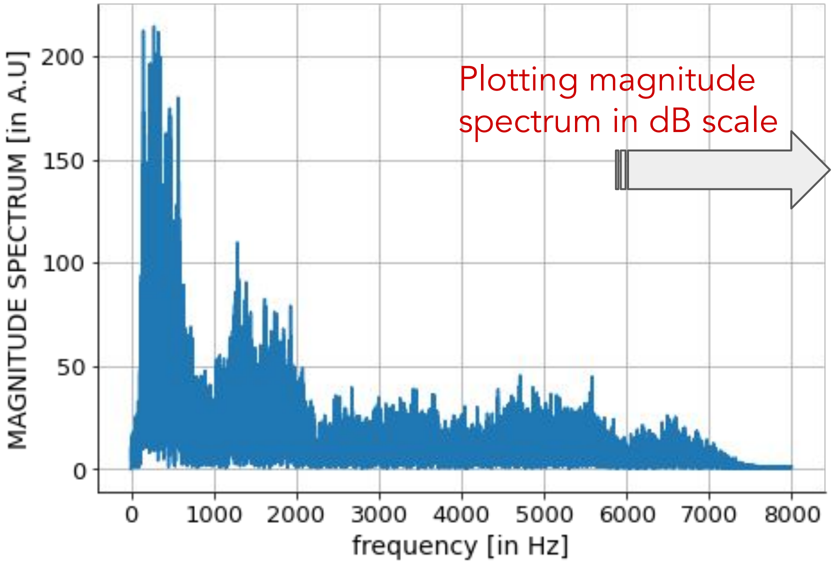
# Audio as a frequency spectrum



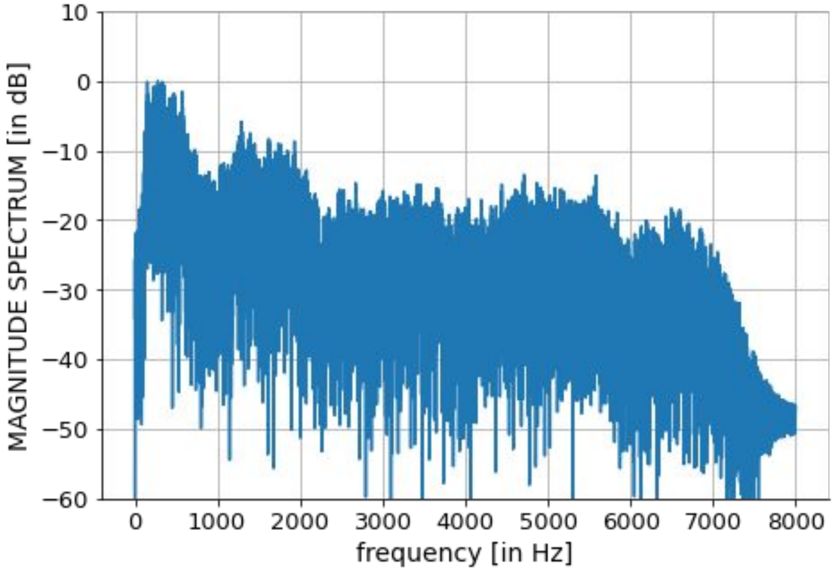
# Audio as a frequency spectrum

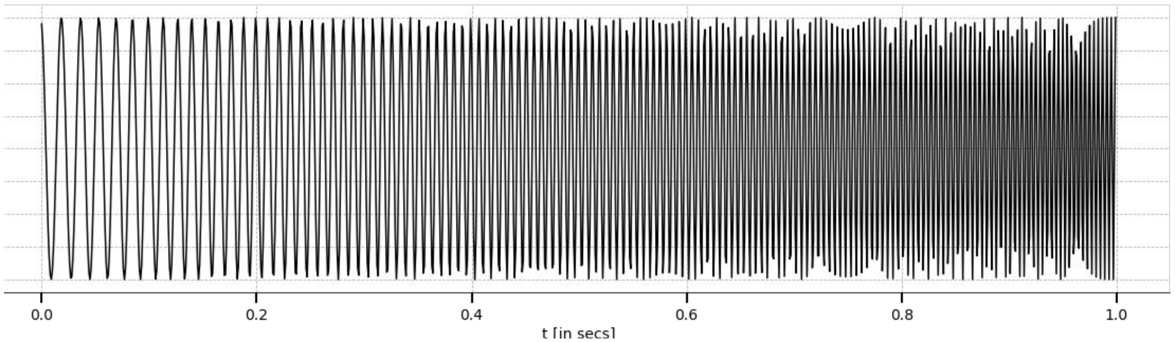



DFT

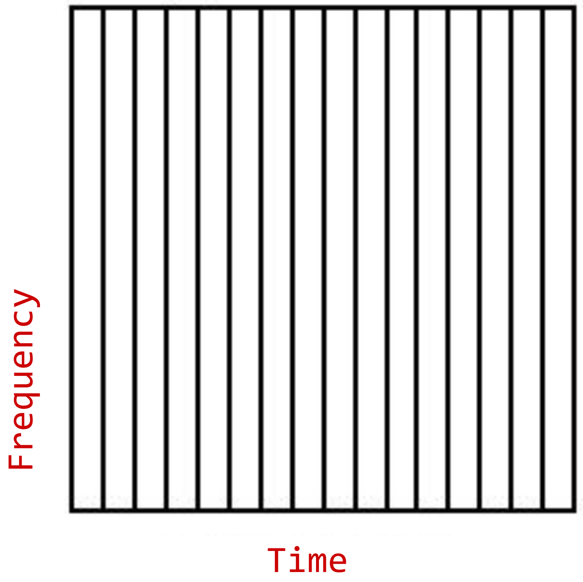



Plotting magnitude spectrum in dB scale

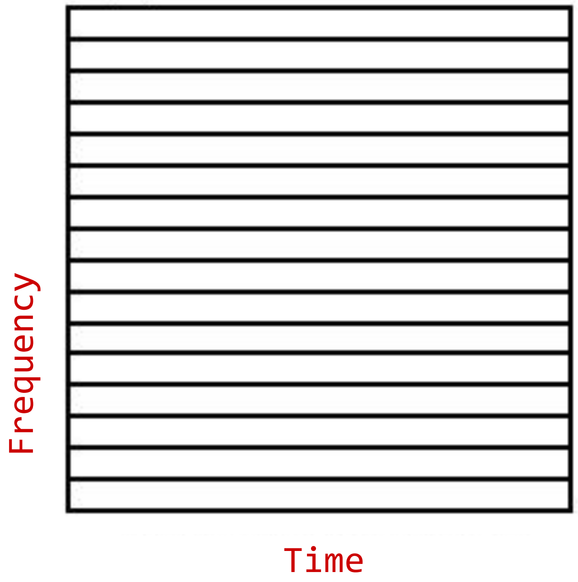


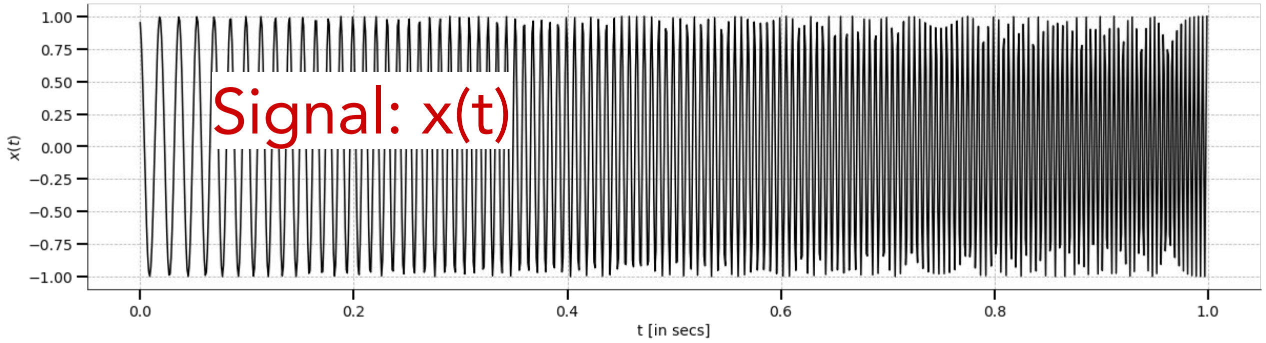


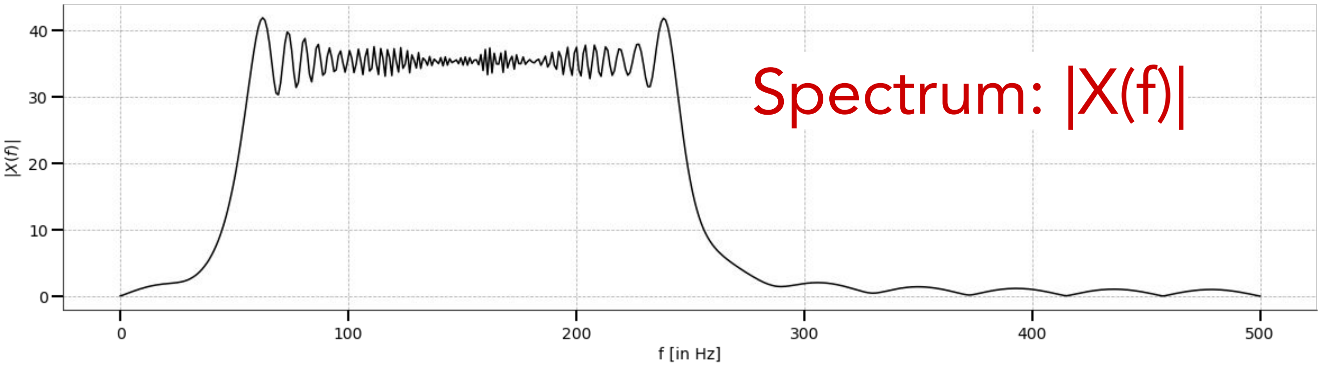

(a) Time series



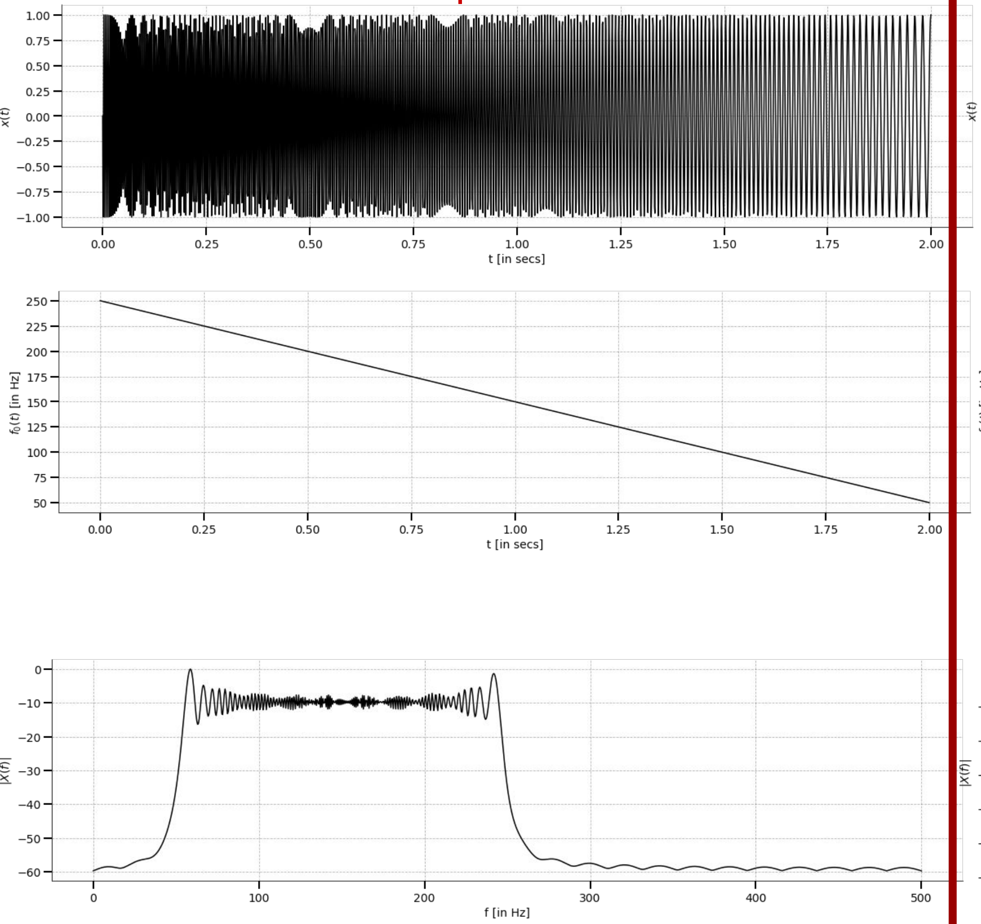

(b) Fourier transform



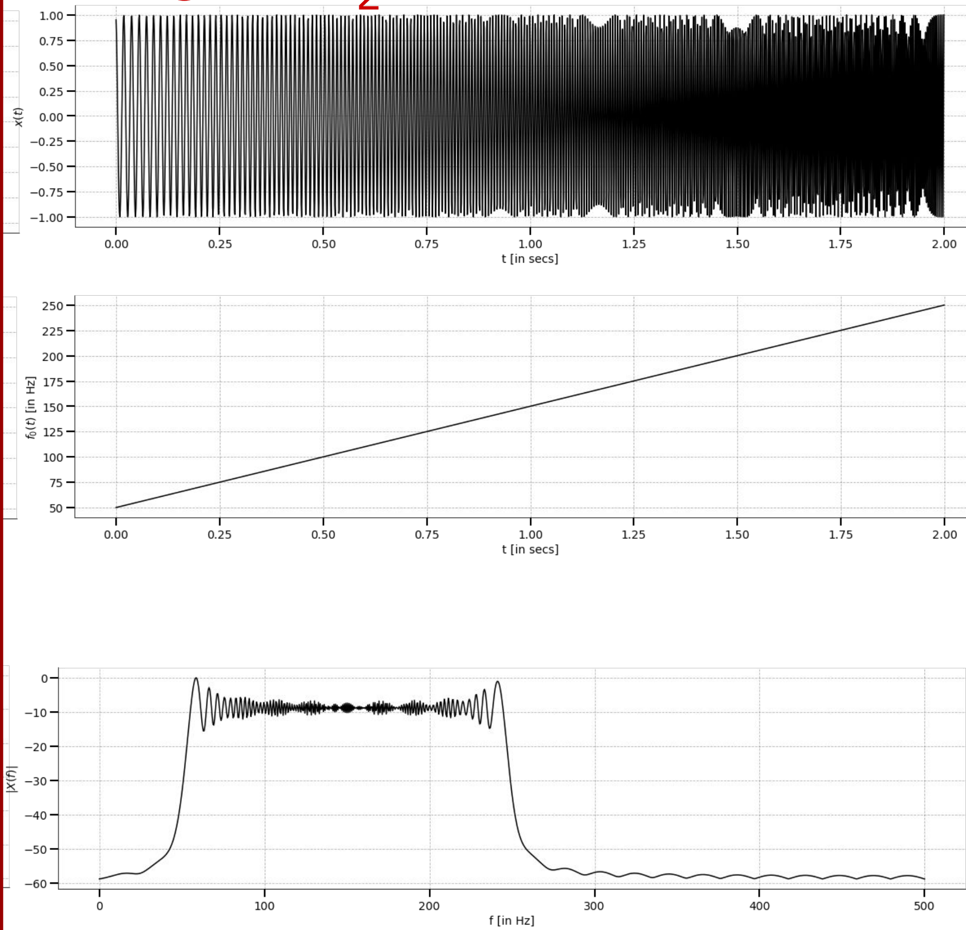

Signal: x(t)
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
x(t)
0.0
0.2
0.4
0.6
0.8
1.0
t [in secs]

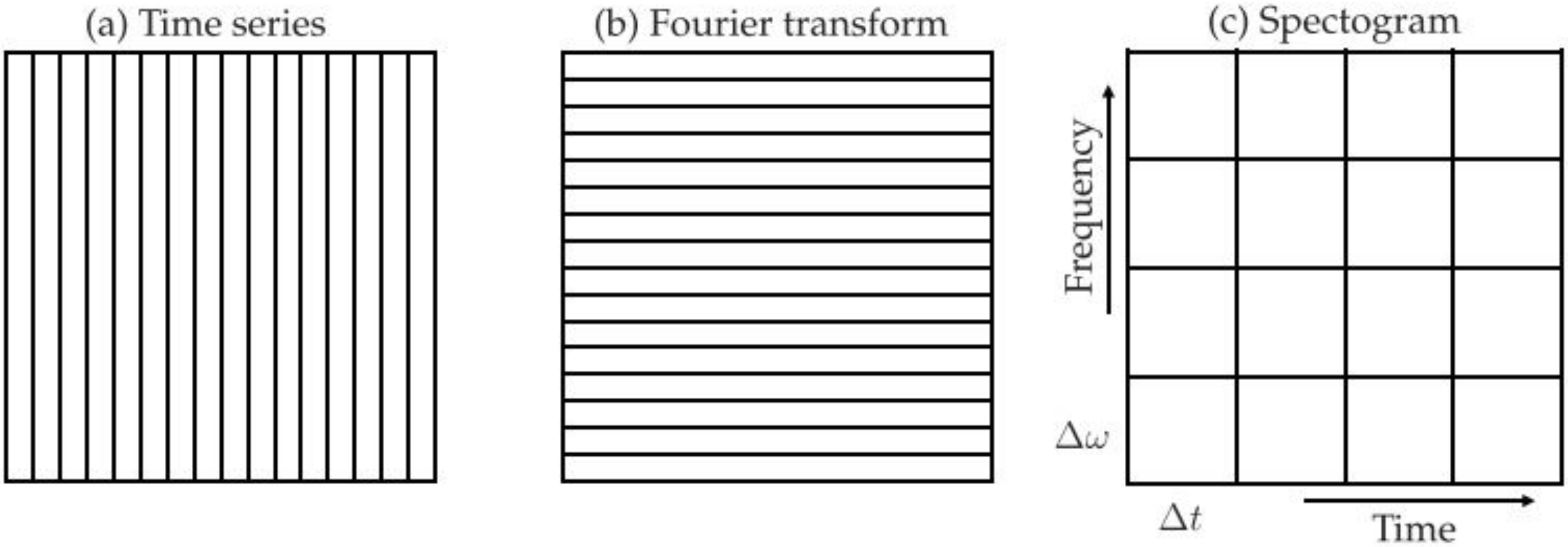

Spectrum: |X(f)|
40
30
20
10
0
|X(f)|
0
100
200
300
400
500
f [in Hz]

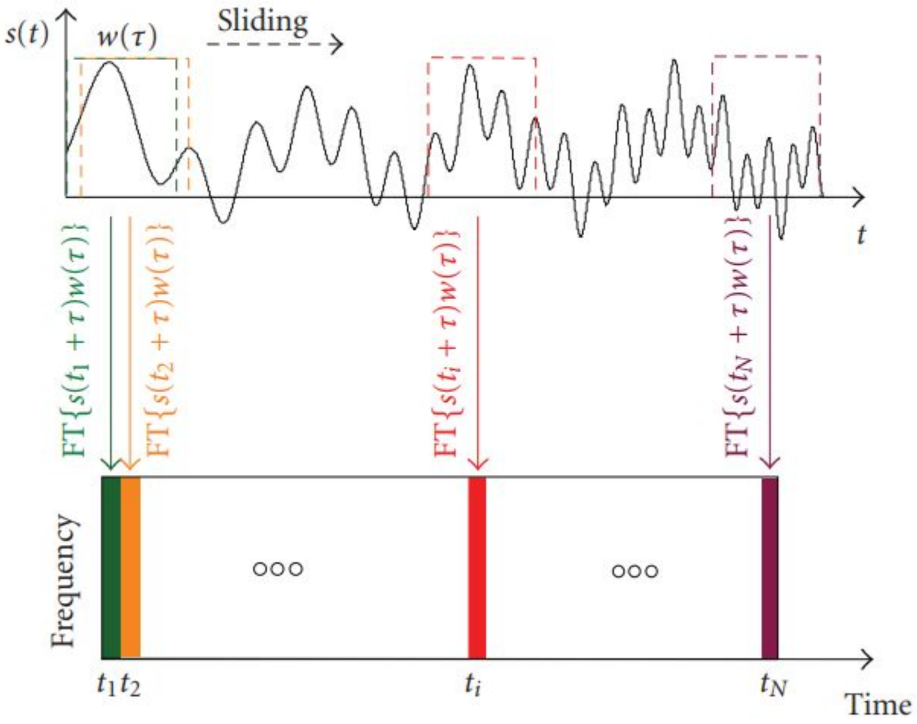
# Signal: $x_1(t)$




# Signal: $x_2(t)$



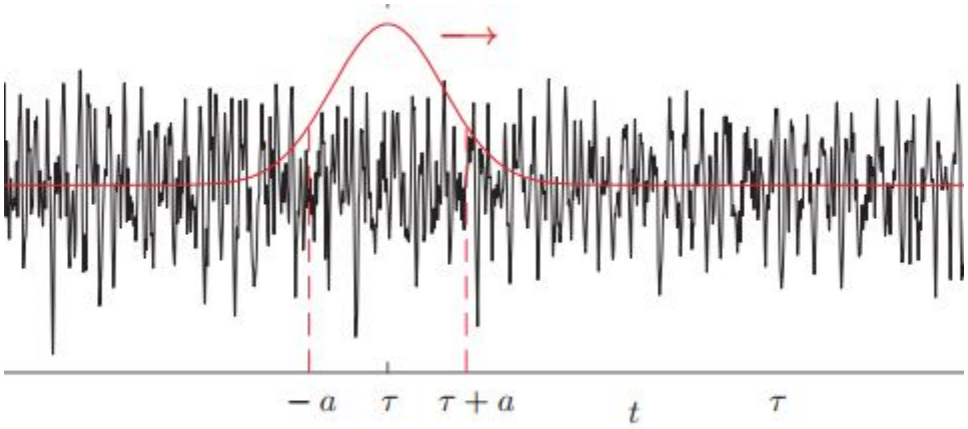

(a) Time series
(b) Fourier transform
(c) Spectogram
Frequency
$\Delta\omega$
$\Delta t$
Time

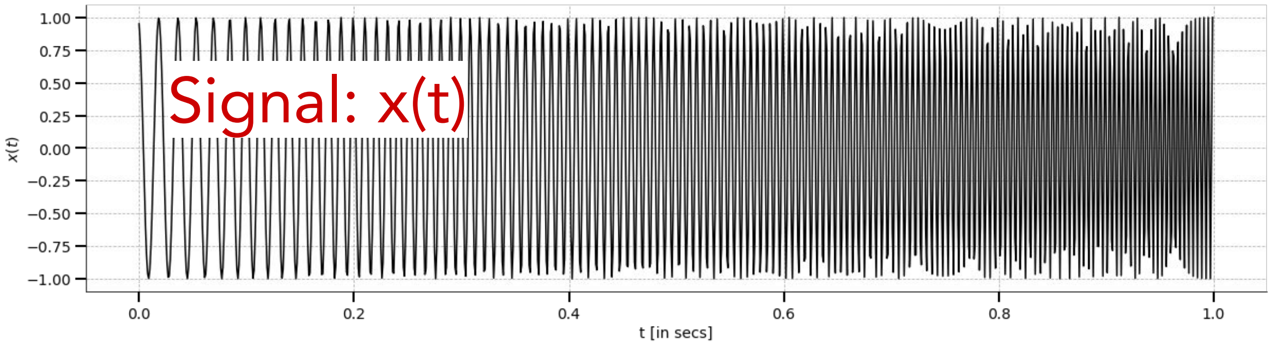


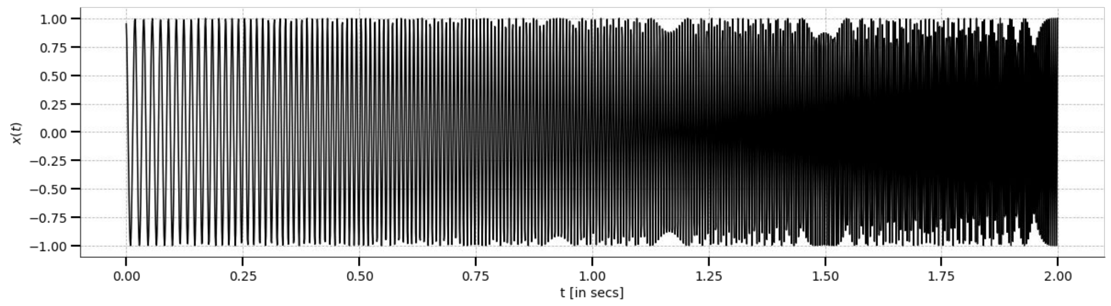

Srdjan Stankovic, Time-Frequency Analysis and Its Application in Digital Watermarking, Journal EURASIP, 2010

# Short-time Fourier Transform (STFT)

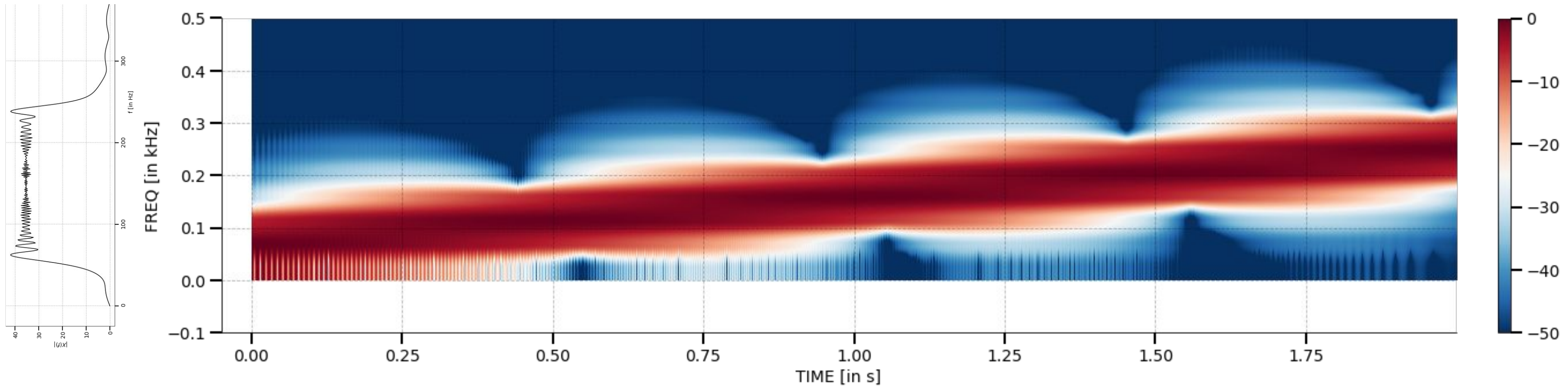
$$X(t, f) = \int_{-\infty}^{\infty} w(t - \tau) x(\tau) e^{-j2\pi f \tau} d\tau$$




Signal: x(t)
x(t)
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0.0
0.2
0.4
0.6
0.8
1.0
t [in secs]



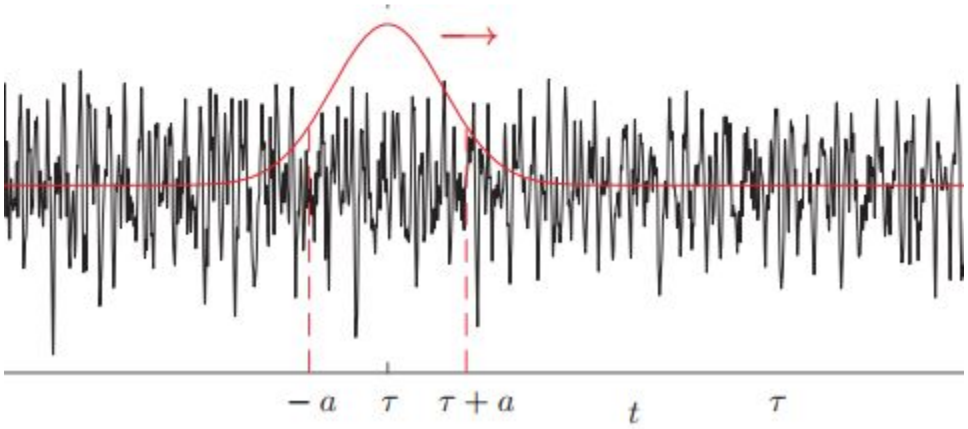
x(t)
t [in secs]

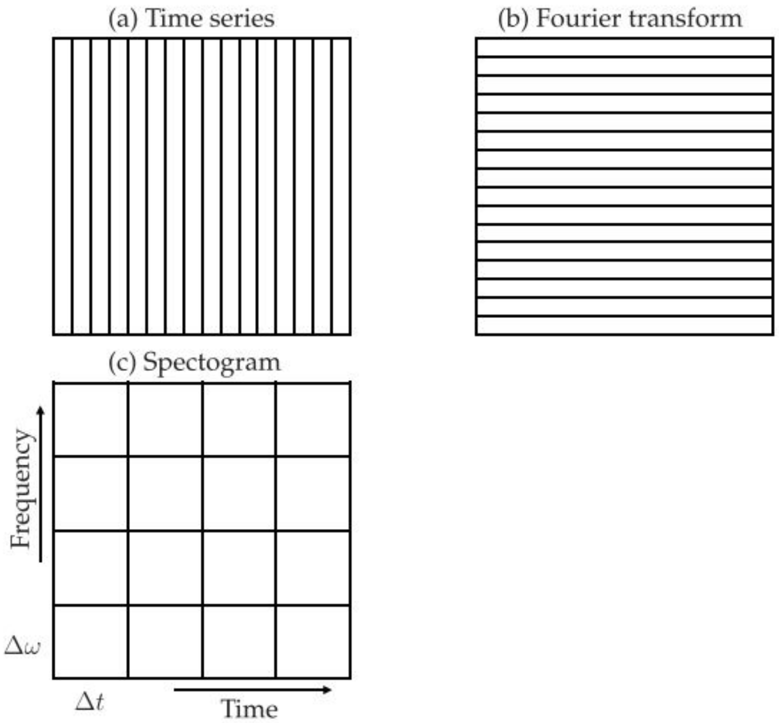
FREQ [in kHz]
TIME [in s]

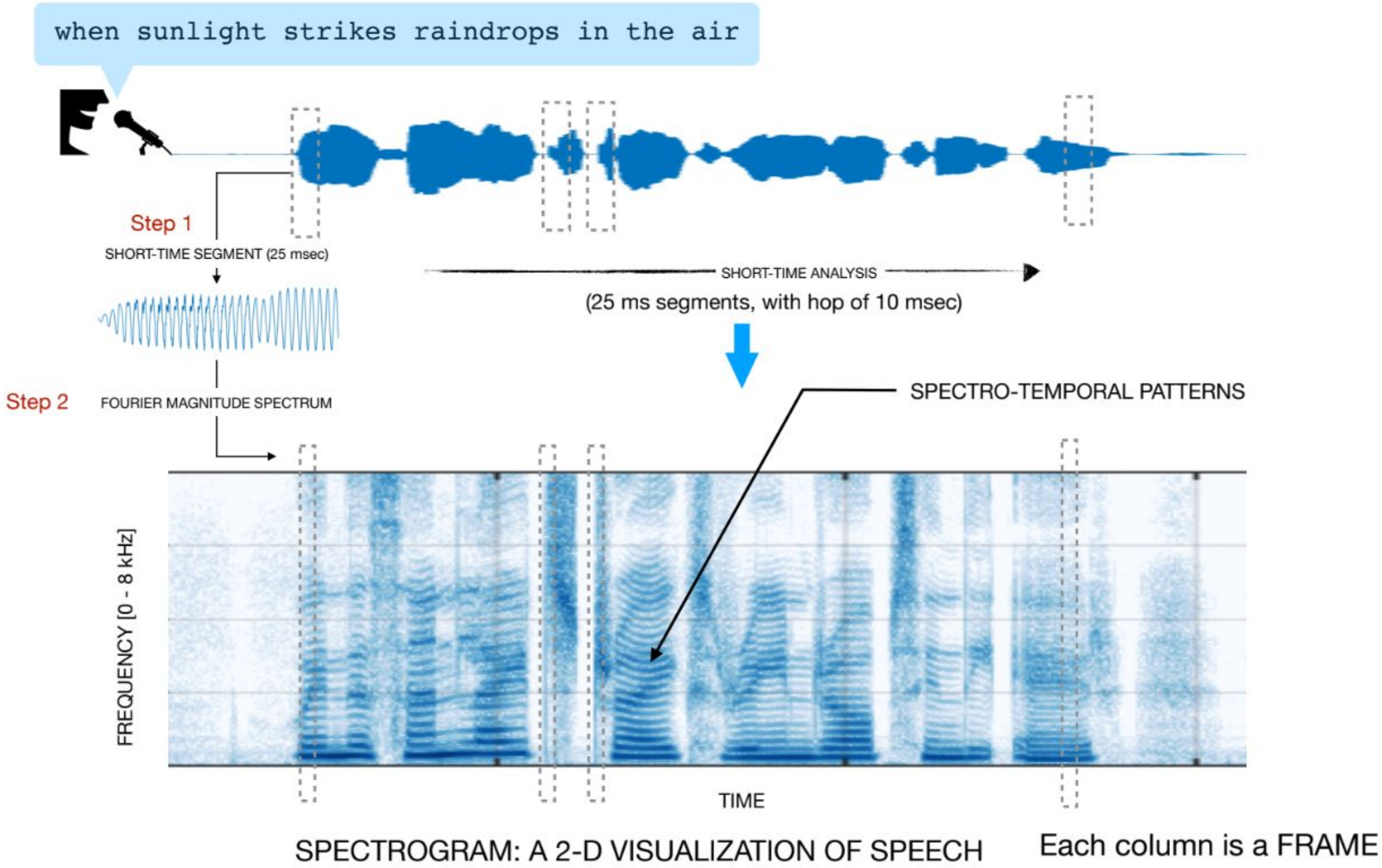
# Short-time Fourier Transform (STFT)

$$X(t,f)=\int_{-\infty}^{\infty} w(t-\tau)x(\tau)e^{-j2\pi f\tau}\,d\tau$$

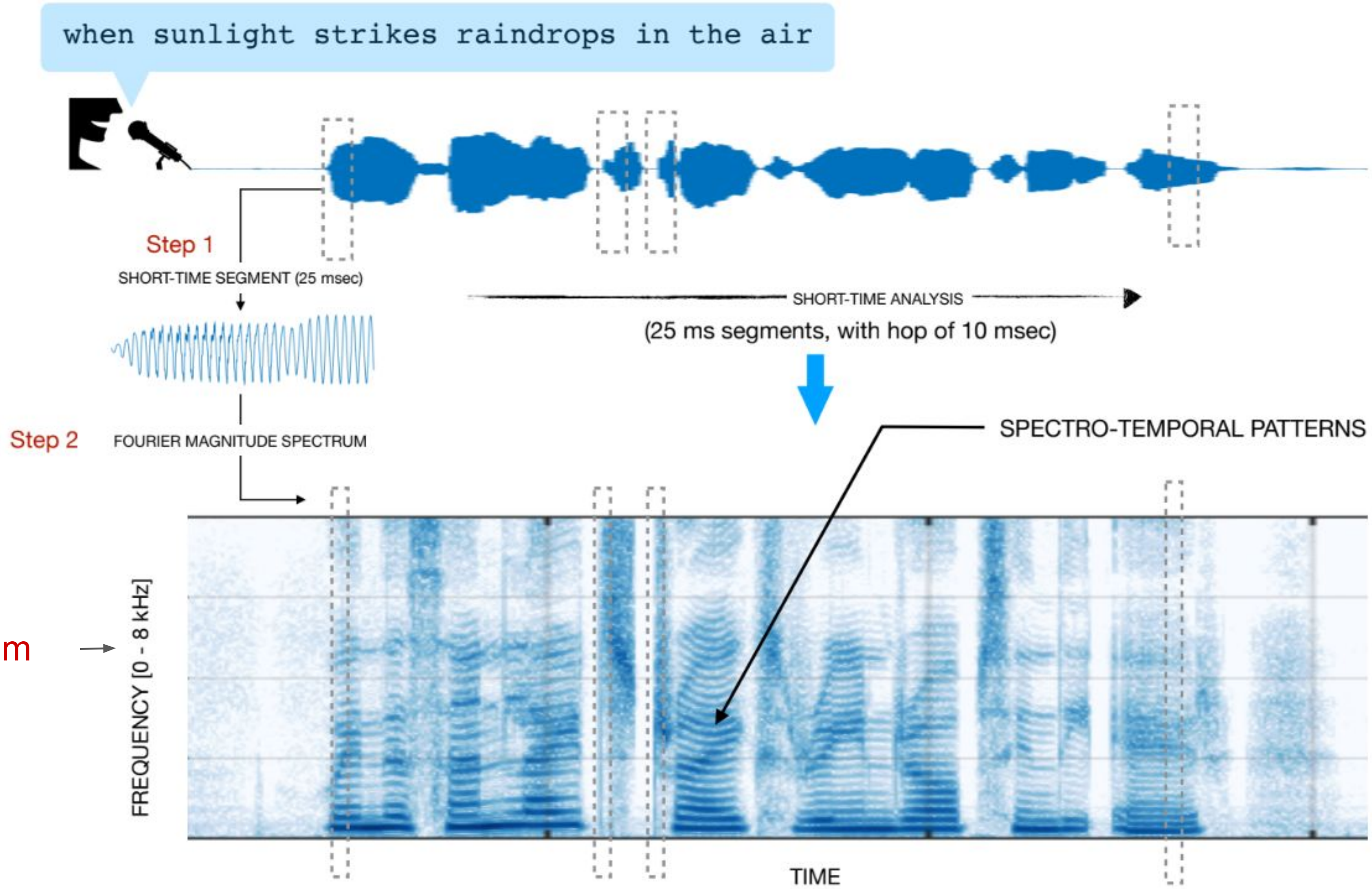



## Resolution




when sunlight strikes raindrops in the air
Step 1
SHORT-TIME SEGMENT (25 msec)
SHORT-TIME ANALYSIS
(25 ms segments, with hop of 10 msec)
Step 2
FOURIER MAGNITUDE SPECTRUM
SPECTRO-TEMPORAL PATTERNS
FREQUENCY [0 - 8 kHz]
TIME
SPECTROGRAM: A 2-D VISUALIZATION OF SPEECH
Each column is a FRAME

Spectrogram




SPECTROGRAM: A 2-D VISUALIZATION OF SPEECH

Each column is a FRAME

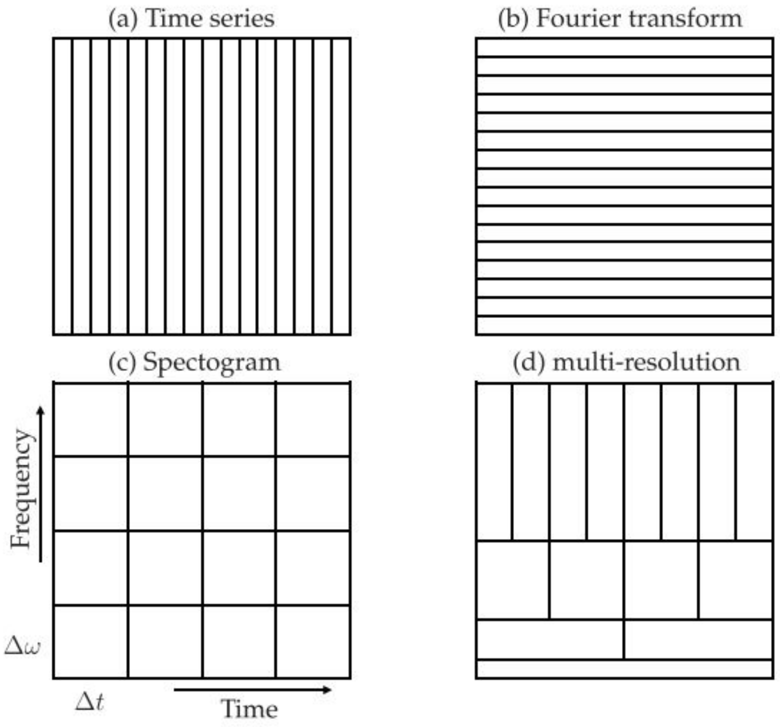
# Multi-resolution analysis

## Resolution

- Wavelets

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)$$

$$\mathcal{W}_\psi(f)(a,b) = \langle f, \psi_{a,b}\rangle = \int_{-\infty}^{\infty} f(t)\bar{\psi}_{a,b}(t)\,dt$$



# Reading Material

**Estimating and Interpreting The Instantaneous Frequency of a Signal—Part 1: Fundamentals**

BOUALEM BOASHASH, SENIOR MEMBER, IEEE

**THEORY OF COMMUNICATION***

By D. GABOR, Dr. Ing., Associate Member.†

# Thank you