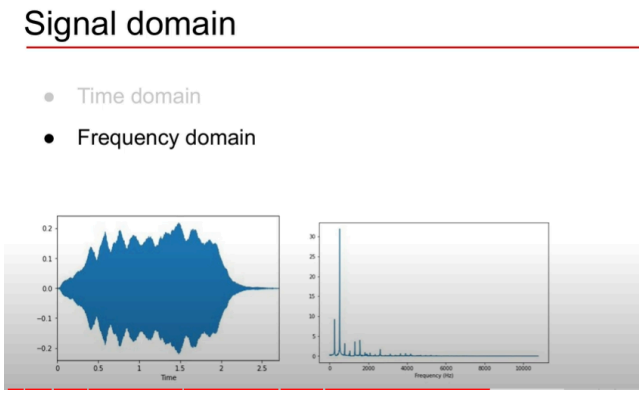
Music humming → detect which song
Knees P & Schedl M → Music similarity and retrieval ; intro to audio and web based strategies

Fourier transform is to change the time domain graph to spectral lines



https://www.youtube.com/@ValerioVelardoTheSoundofAI

LECTURE 6
Audio Processing for ML
Frequency domain features and Time domain features

Time domain feature
ADC conversion→ Framing (bundling samples) → Feature Computation → Aggregation (Mean, median, GMM) → Feature values/vector/matrix

[Sample 1-128 = frame 1 ==OVERLAPPING FRAMES
Sample 64-192 =frame2]

Why framing before removing its features?
Frames
Perceivable audio chunk
Since samples are very short. And ear resolution is like 10 ms but a sample is like 0.025ms so enough samples to hear is a frame
Power of 2 ^ yada samples is a frame
Typical value=256-8792

Duration of a frame = (1/sample rate inHz) * K[frame size]

Frequency domain feature
ADC → Framing →Time to frequency (With Fourier Transform) → Windowing → fourier transform →Feature Computation (of freq) → aggregation of results → feature vector/val/matrix

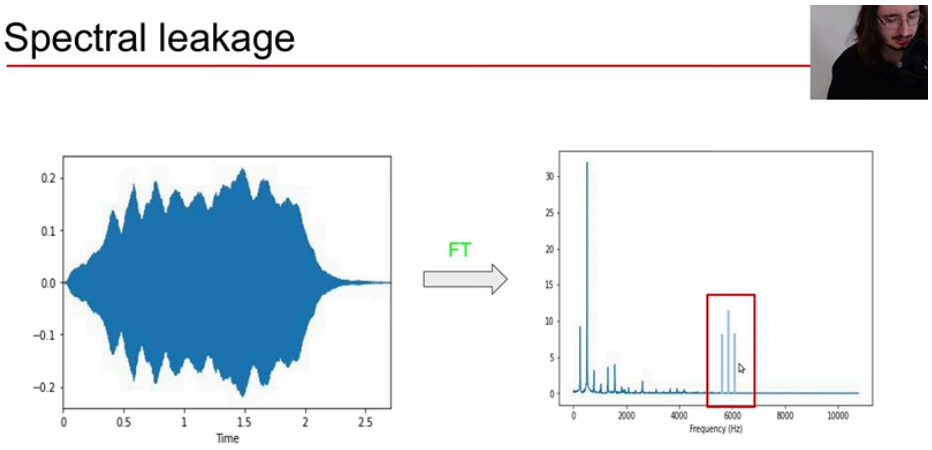
Prob: Spectral Leakage→ Endpoints of the Signal is Discontinuous because they are not int number of periods (if they are integers there wouldn't be any incompleteness)--> high frequency other components come because non integer periods

ChatGPT

Result in Frequency Domain

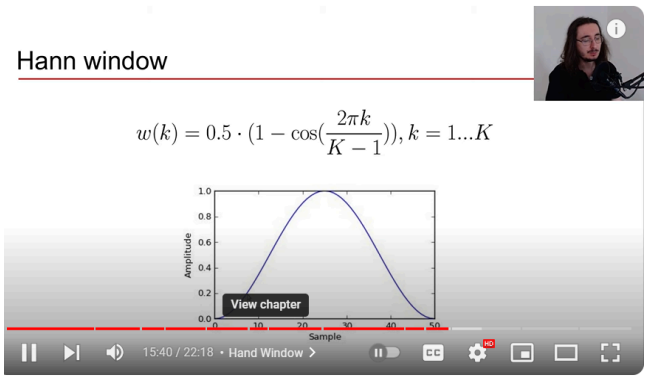
Integer Periods (1 second): In the frequency domain, we see a sharp peak at 5 Hz with little energy elsewhere.

Non-Integer Periods (0.9 seconds): In the frequency domain, the peak at 5 Hz is spread out, and we see energy at frequencies other than 5 Hz. This spread of energy is spectral leakage.



The red box is the random high frequency that is present due to spectral leakage

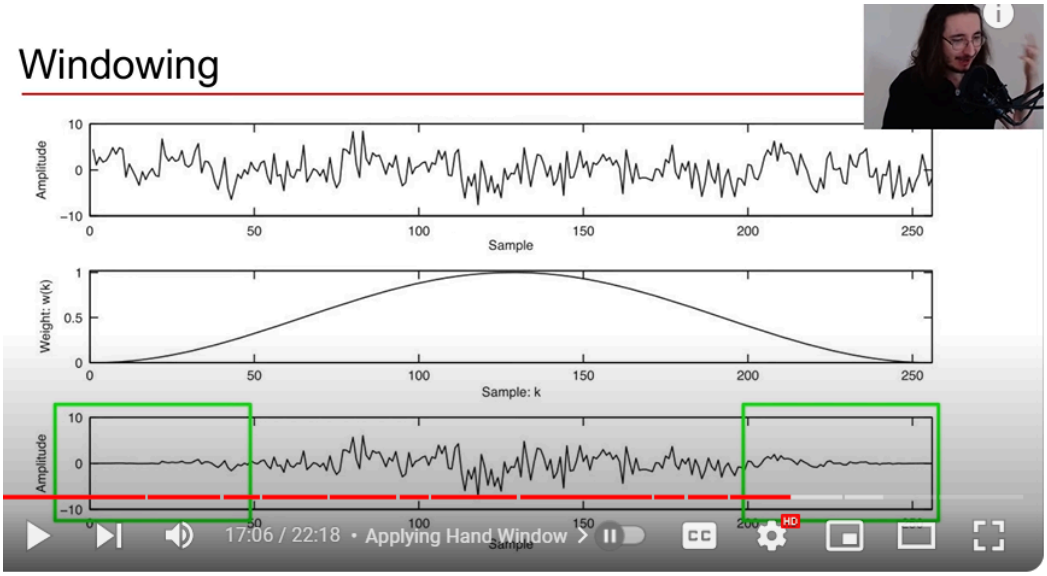
How to fix this?



Windowing function to each frame before we feed into FT. Remove the info from the endpoints and generates a periodic function

Windowing Function == HAAN WINDOW

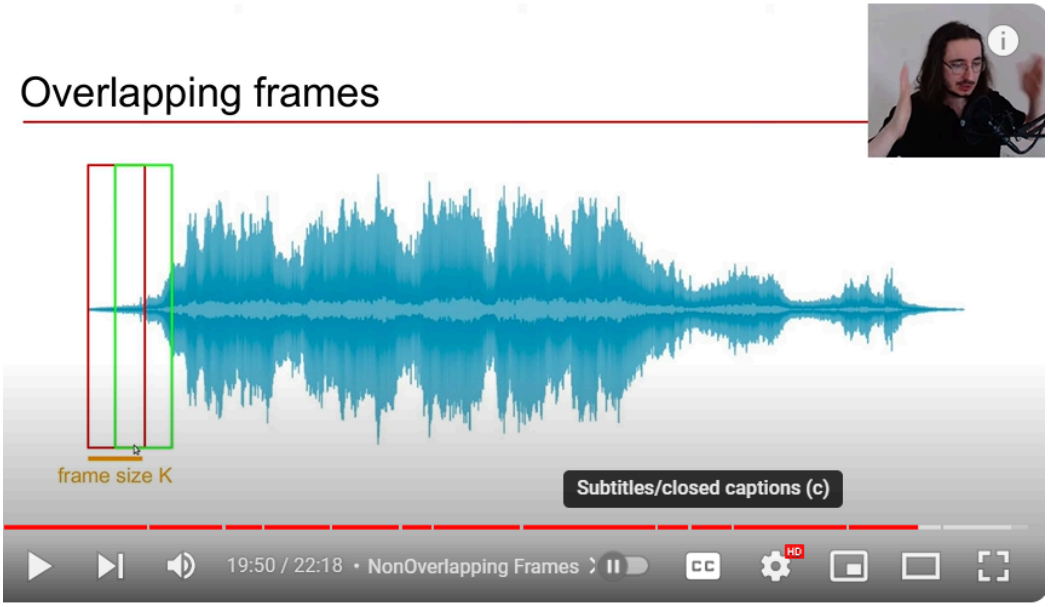
The graph is of the Haan window function



Before, haan window function and after applying the function SINGLE FRAMES

We end up losing data(signal) when we have multiple frames

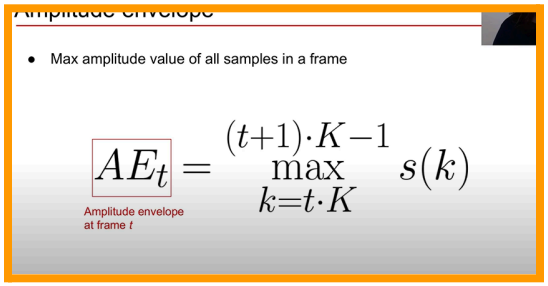
Overlapping frame



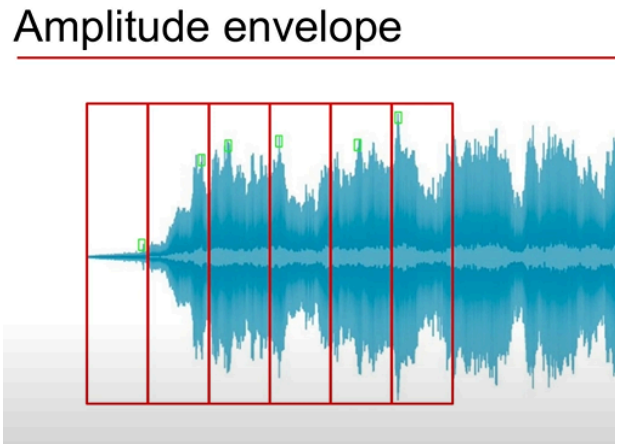
Hop length
Is the the length from the first red to first green
==how much samples we have to shift right to reach a new frame

low to Extract Audio Features

LECTURE 7
Understanding Time Domain Audio Features



Amplitude Envelope == Max amplitude value of all samples in a frame
K= frame size or number of samples in a single frame
s(k)= amplitude of kth sample
k= t.K where t is the number of sample (like which sample is it 0th , 1st etcc) ==GIVES THE FIRST SAMPLE FRAME t
[(t+1).K]-1 ==LAST SAMPLE OF FRAME t → How? It is the last because with t+1 you go to the next frame then you multiply with K and then you -1 cause you need to go back to the LAST frame



USE?
We can figure when the acoustic event starts (spike of amplitude)

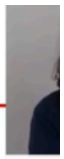
Root Mean Square, Zero Crossing Rate(Recognition of Precursive VS pitched Sound)
Monotonic pitch est,
Voice/unvoiced

CODE UNDERSTANDING
RMS is to find the loudness, energy of the signal
Zero Cross Rating is to find the
ZCR indicates the rate at which the audio signal changes its sign (from positive to negative or vice versa).
For speech processing, ZCR can be useful in detecting silence or pauses between words.

FOURIER TRANSFORM

`figsize=(18,0)` means that the figure has a width of 18 inches and a height of 0 inches.

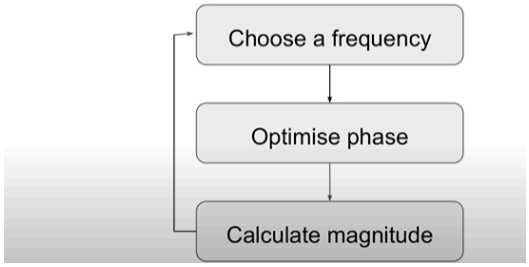
Fourier transform



$$\varphi_f = argmax_{\varphi \in [0,1)} \left(\int s(t) \cdot sin(2\pi \cdot (ft - \varphi)) \cdot dt \right)$$

The whole integral is used to calculate the area signal and the sin wave of the frequency and the arg max is to basically finding the phase which has the highest similarity

Fourier transform: Step by step



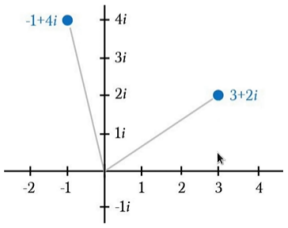
https://teropa.info/harmonics-explorer/

LECTURE 11

Why get involved with complex numbers?
→ To showcase magnitude and phase
i : c = a +ib a and b are real units ib is the imaginary part

Plotting complex numbers on the complex plane
→ X axis is real axis and Y is the imaginary axis

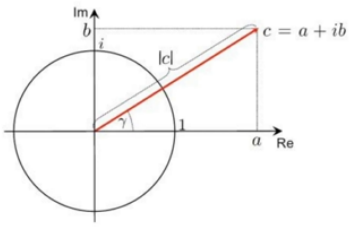
Plotting complex numbers



==CARTESIAN

POLAR REPRESENTATION

Polar coordinate representation



Find c and gamma
Find using pytho theorem
And find the gamma
Tan gamma = b/a
Gamma = arctan(b/a)

$$a = |c| \cdot \cos(\gamma) \quad b = |c| \cdot \sin(\gamma)$$

$$c = a + ib$$

$$c = |c| \cdot (\cos(\gamma) + i \sin(\gamma))$$

Euler formula → $e^{i * gamma}$ = cos gamma +i sin gamma
Euler identity → $e^{i * pi}$ +1 = 0

Pi as gamma and it works

—----------

C = |c| . $e^{i*gamma}$

|c| IS THE DISTANCE FROM THE CENTRE origin
And the other part is the direction of a number in the complex plane→ theta tan inverse tan
Magnitute and direction

LECTURE 12

$$\varphi_f = argmax_{\varphi \in [0,1)} \left(\int s(t) \cdot sin(2\pi \cdot (ft - \varphi)) \cdot dt \right)$$

Phase

$$d_f = \max_{\varphi \in [0,1)} \left(\int s(t) \cdot sin(2\pi \cdot (ft - \varphi)) \cdot dt \right)$$

Magnitude

C = |c| . $e^{i*gamma}$

$$c_f = \frac{d_f}{\sqrt{2}} \cdot e^{-i2\pi\varphi_f}$$

Mangnitude and phase in an complex number
FT coefficient
Mapping this onto the polar form ie.

C = |c| . e ^{i*gamma}

Due to the - sign present, as phase increases the polar coordinate goes clockwise which is opposite

Continous audio SIGNAL

Complex Fourier Transform
g cap (f) = c_f
G cap : Real number → Complex number

$$d_f = \max_{\varphi \in [0,1)} \left(\int g(t) \cdot sin(2\pi \cdot (ft - \varphi)) \cdot dt \right)$$

$$\varphi_f = argmax_{\varphi \in [0,1)} \left(\int g(t) \cdot sin(2\pi \cdot (ft - \varphi)) \cdot dt \right)$$

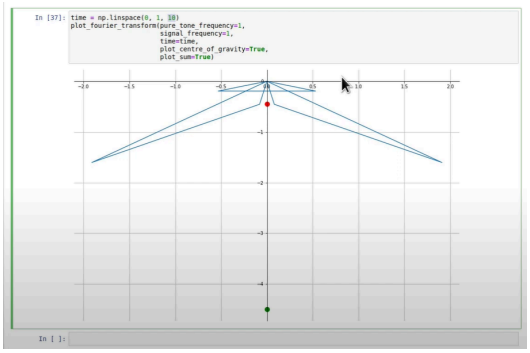
$$\hat{g}(f) = \int g(t) \cdot e^{-i2\pi ft} dt$$

Whatever is present with e →shows a unit circle in a complex plane that is a pure tone freq
→clockwise since -
→ Speed is the frequency f ; if f=2 it will take 0.5 secs to go through the entire unit circle

The main idea behind the Fourier transform is that any function, especially periodic ones, can be represented as a sum of simpler sinusoidal functions (sines and cosines). The Fourier transform generalizes this concept to non-periodic functions as well, breaking them down into their frequency components.

A **pure tone frequency** refers to a sound wave that's a sinusoidal waveform and is characterized by a single frequency component. This means it oscillates at a single, constant frequency without any harmonics or overtones. Pure tones are often used in various fields for testing and analysis due to their simple and predictable properties.

Average →plot _Centre_of_gravity is the point which shows the coeff of
But hte integration is summing
== what we are doing is multiplying the centre of gravity with the number of steps we are taking



$$\hat{g}(f) = \int g(t) \cdot e^{-i2\pi ft} dt$$

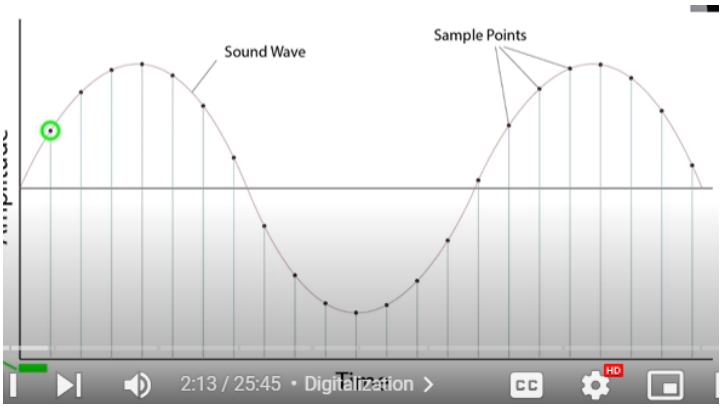
First is Complex FT

$$g(t) = \int c_f \cdot e^{i2\pi ft} df$$

Inverse Fourier Transform

LECTURE 13 discrete fourier transform

Discrete signal is what we use
Analog signals are continuous signal—> We have to digitize it for you to use by sampling



g(t) |→x(n) where n =0,1,2…

t=nT n=current sample
T = time period \

Its is a function of t in the continuous fourier transform → Its is broken at sample points of n

$$\hat{g}(f) = \int g(t) \cdot e^{-i2\pi ft} dt$$

→ Problems = we have continous freq and infinite time we are dealing with

$$\hat{x}(f) = \sum_{n} x(n) \cdot e^{-i2\pi fn}$$

→ how to fix == Focussing in finite time and fixing it like a song of 3 mins so no probs here
→ We can transform for the finite number of freq

Number of freq = Number of samples

Why?

As we can have the change of freq-> time and vise versa

Limit 0 ->N-1 training only N samples

\

$$\hat{x}(k/N) = \sum_{n=0}^{N-1} x(n) \cdot e^{-i2\pi n \frac{k}{N}}$$

And the freq has changed to k/N

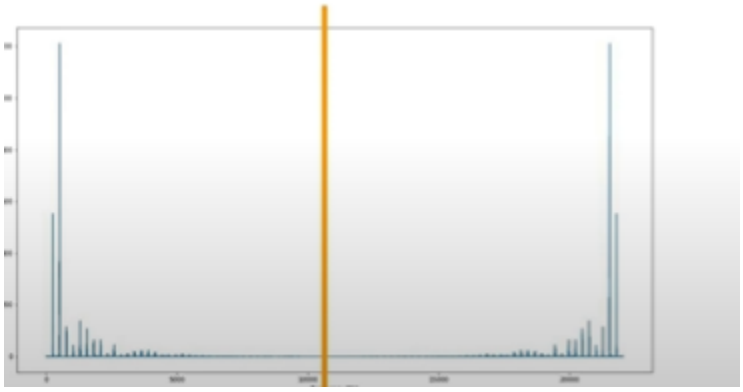
k=[0 , M - 1] = [0 , N - 1]

Sampling rate is the inverse of sampling period

$$F(k) = \frac{k}{NT} = \frac{ks_r}{N}$$

O/P of the FT is the coefficients that give us information at each freq Information ie Magnitude and Phase

Redundancy in DFT



Since we have a symmetry here we can say

$k=N/2 \rightarrow F(N/2) = s_r/2$

So we have to only care about the first half that is $s_r/2$

Nyquist Freq == The highest freq above which we cannot reconstruct a digital signal into its analog singal without aliasing

DFT is not as efficient compared to fast fourier transform (uses redundancies in the signal)