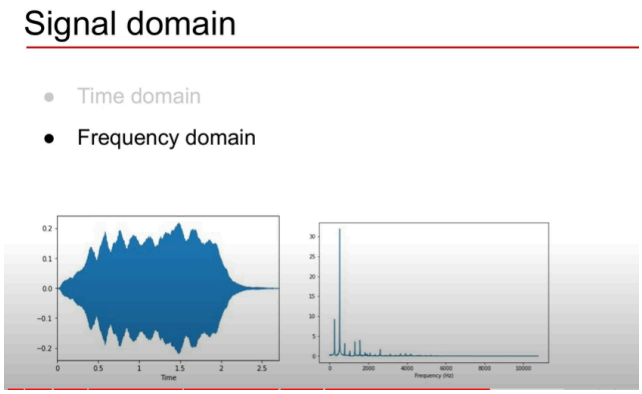
Music humming → detect which song
Knees P & Schedl M → Music similarity and retrieval ; intro to audio and web based strategies

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

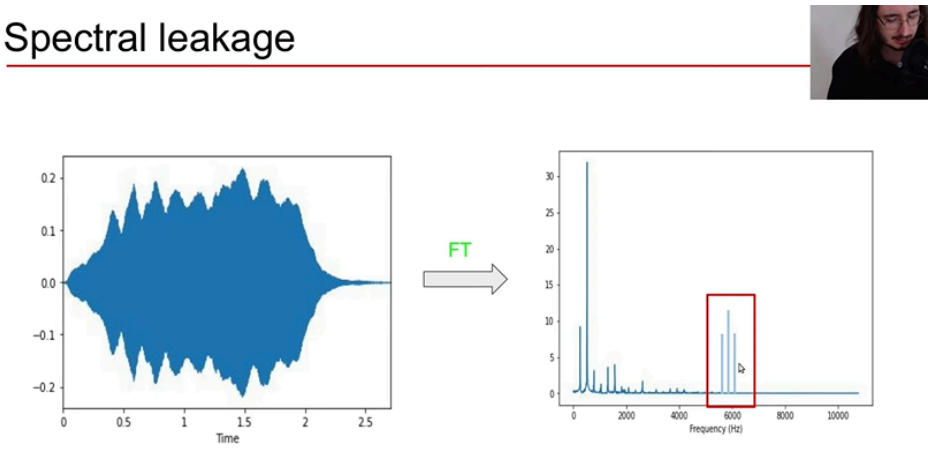
Prob: Spectral Leakage→ Endpoints of the Signal is Discountinous because they are not int number of periods (if they are integers there wouldnt be any incompleteness)--> high frequency other components come because non integer periods

ChatGPT

Result in Frequency Domain

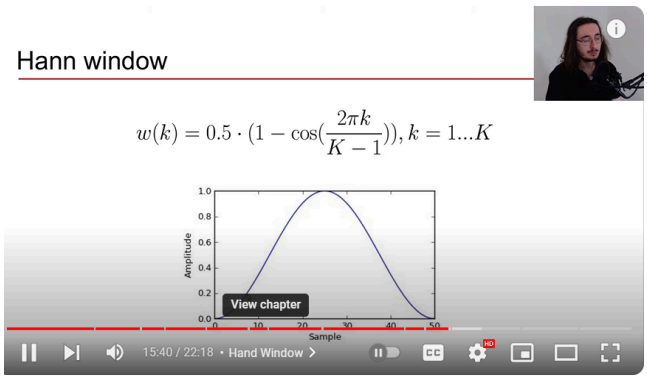
Integer Periods (1 second): In the frequency domain, we see a sharp peak at 5 Hz with little energy elsewhere.

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

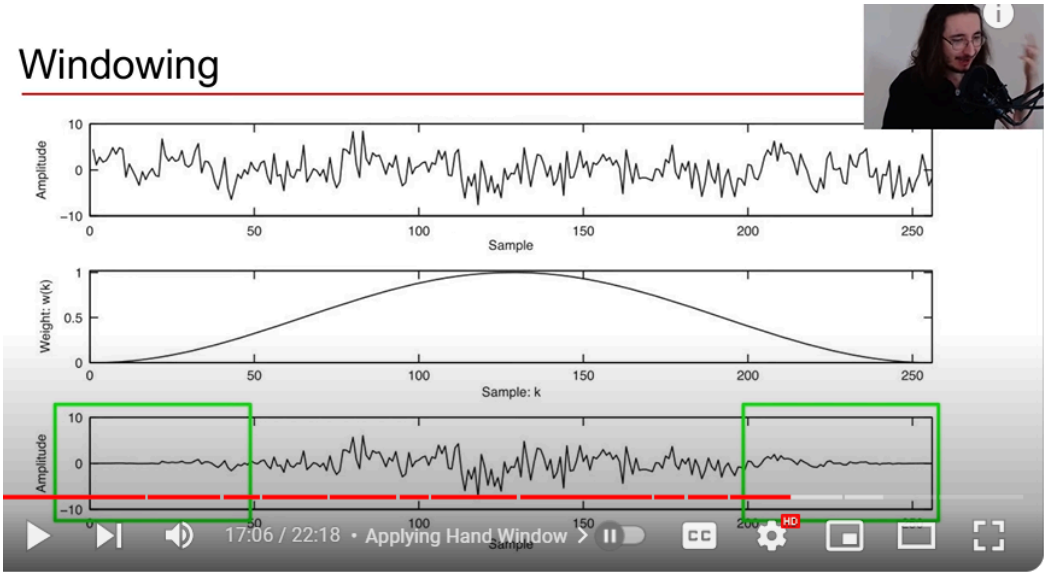
How to fix this?



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

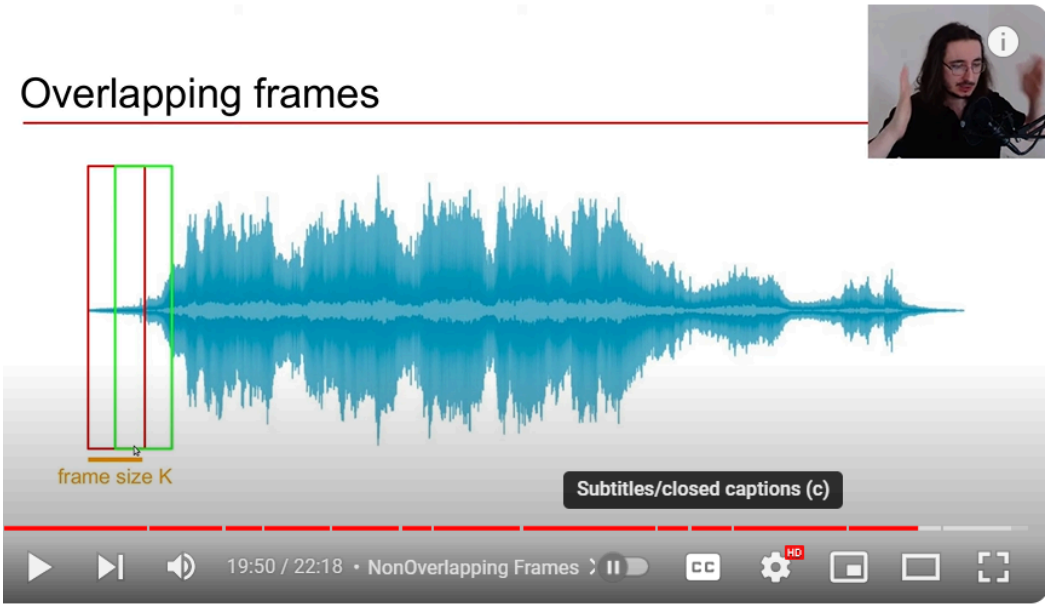
The graph is of the Haan window function



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

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

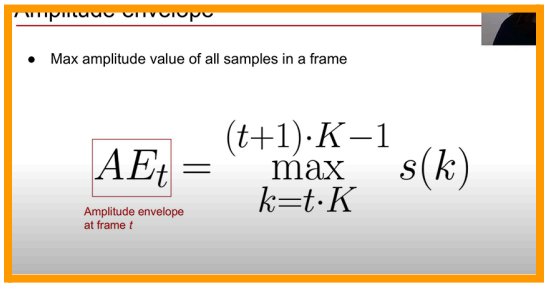
Overlapping frame



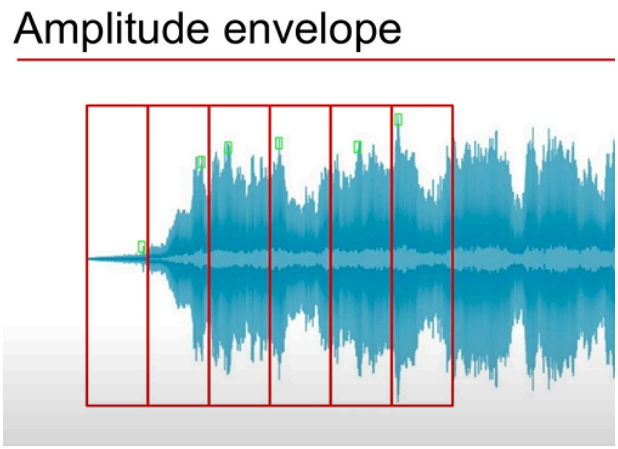
Hop length
Is the the length from the first red to first green ==how much samples we have to shift right to reach a new frame

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 Precurrsive VS pitched Sound)
Monotonic pitch est,
Voice/unvoiced