<https://youtube.com/playlist?list=PL-wATfeyAMNqIee7cH3q1bh4QJFAaeNv0>

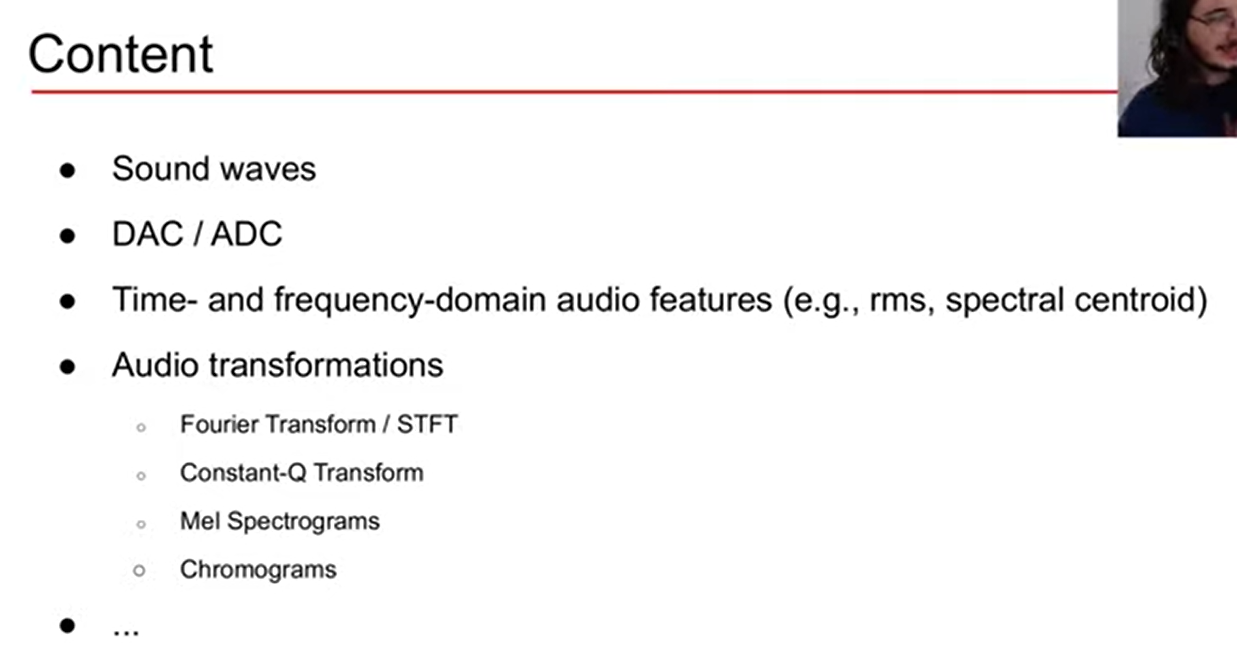
Video 1] Series Overview

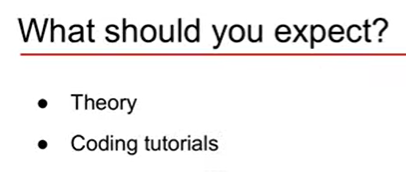
Audio signal processing applications -



There’s 2 parts for audio AI applications - preprocessing n DL model creation-testing. He has another playlist for latter [<https://youtu.be/iCwMQJnKk2c> ]

Content of the series -

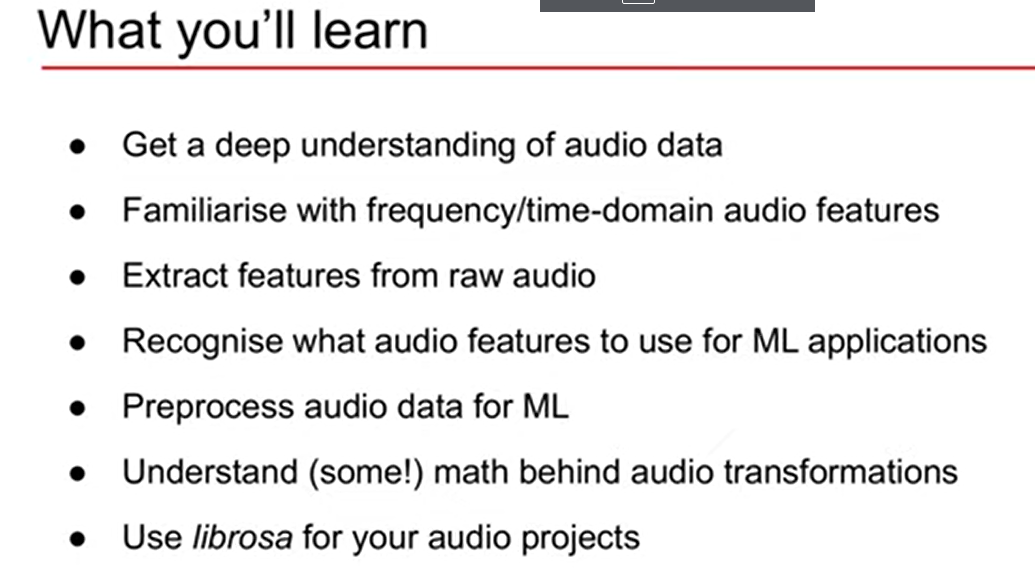


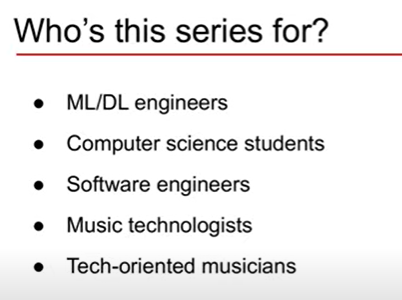




[Link to his Github acct](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbkRZVDB4ZGRGak9FcTN1bk5jMDBEOXd6MlNXd3xBQ3Jtc0tuNGE1ZWFnb1ZaTm92VUVUa2hsbHlCNkFKRzQ5YXU1TTYwQW9EdzZncW94UzV2NnBkR0pld1NFSk9qRndKX0l2VGl1cWxVNDJVVkM1QzBOYmtyMmNpUDJhM1ZrbzRZNER0c0xJcDhCWXM4bG9MVlprVQ&q=https%3A%2F%2Fgithub.com%2Fmusikalkemist%2FAudioSignalProcessingForML%2Ftree%2Fmaster%2F1-%2520Overview&v=iCwMQJnKk2c)

Tech stack - Python and Librosa.

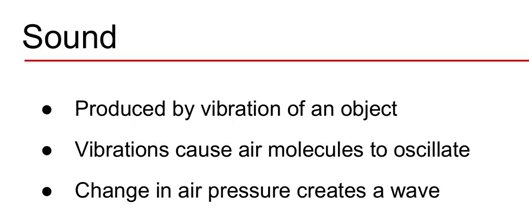




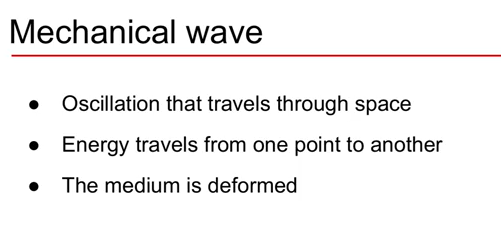
Intermediate Python skills needed.

Slack community - [Link to Slack community Sound of AI](https://www.youtube.com/redirect?event=video_description&redir_token=QUFFLUhqbHFObzR3cjVpVHcwNUVNaFF0dkdUMUZybDEtUXxBQ3Jtc0tsMFVtMS1YWVFqdnl6OFpfNWJaQlhfU1JodDQyOVR1RzhjS1VUS3paLTJoaHByWnZRTkJpbkxFTmtlV19qWElKVGlKcjlCME5VWV9XY3FkZ29NOEhVaTdaNFNrNWJ0OTkwWlpBYngzSDZKQjRna1FLQQ&q=https%3A%2F%2Fvaleriovelardo.com%2Fthe-sound-of-ai-community%2F&v=iCwMQJnKk2c)

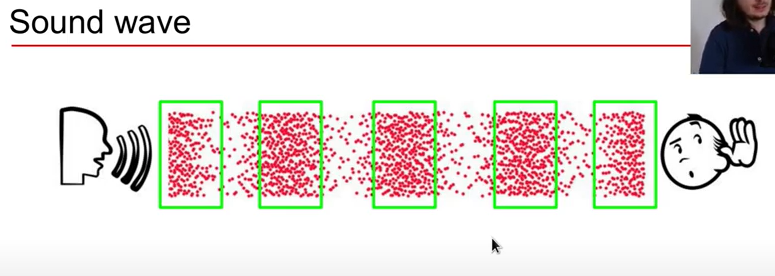
Video 2] Sound and Waveforms



Sound is a Mechanical wave.

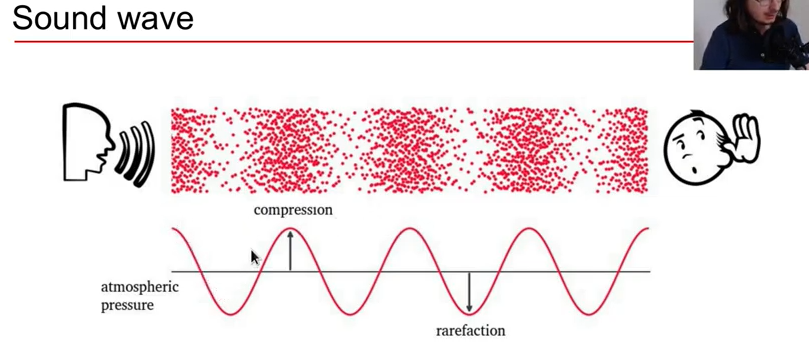


Mech waves need a medium, for propagation.

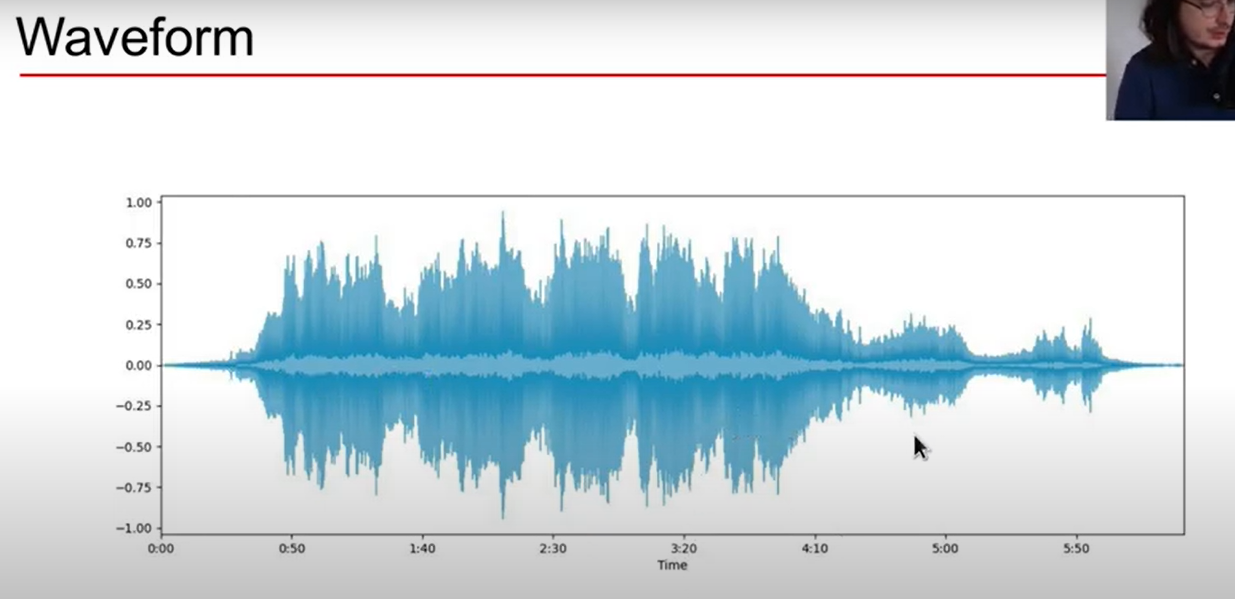


Points of high pressure vs low.

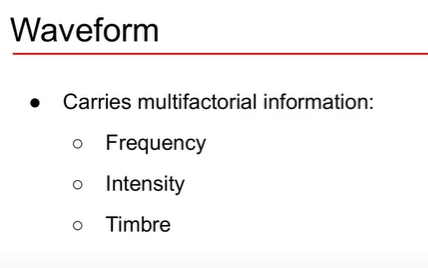
Can visualize like this

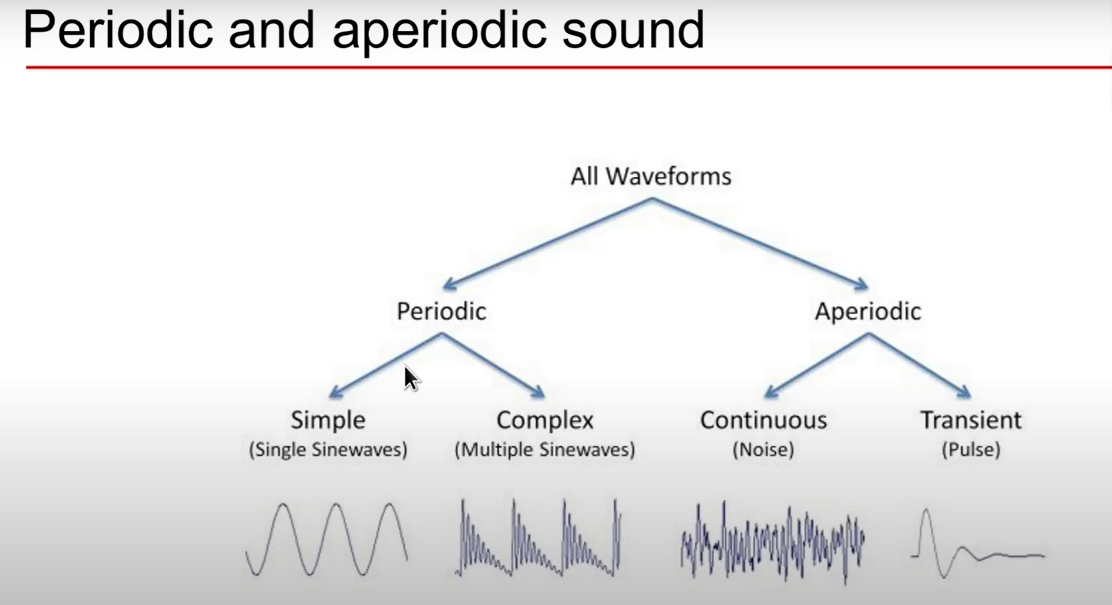


Can visualize with a waveform.



Plotting deviation from Atmospheric pressure [0.00].

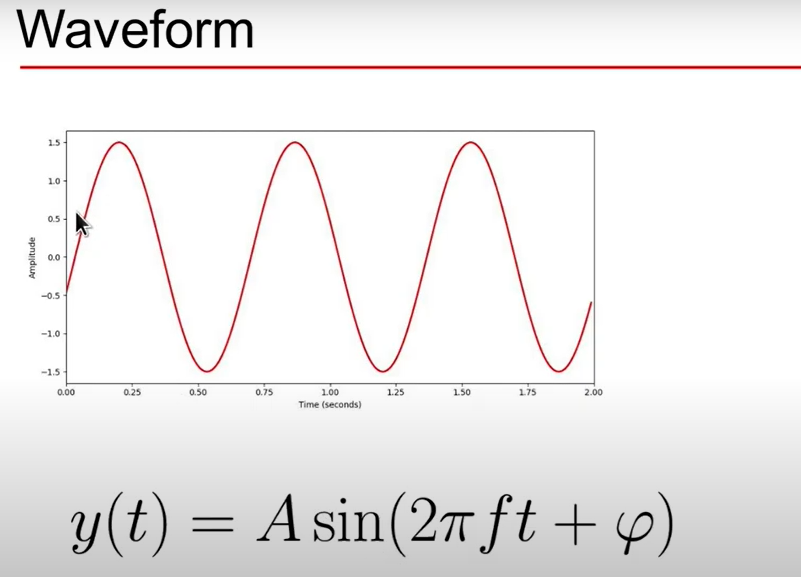




Transient is like a sudden burst of energy- like a pulse; clicking or popping sounds.

Complex periodic is like an orchestra.

Starting with simple sine wave-

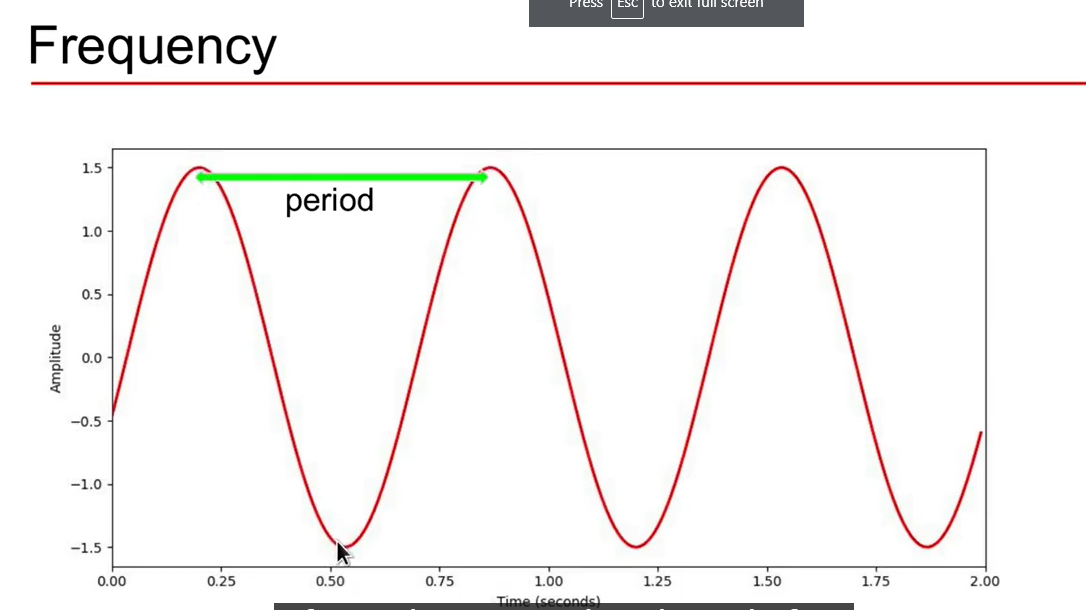


A is amplitude.

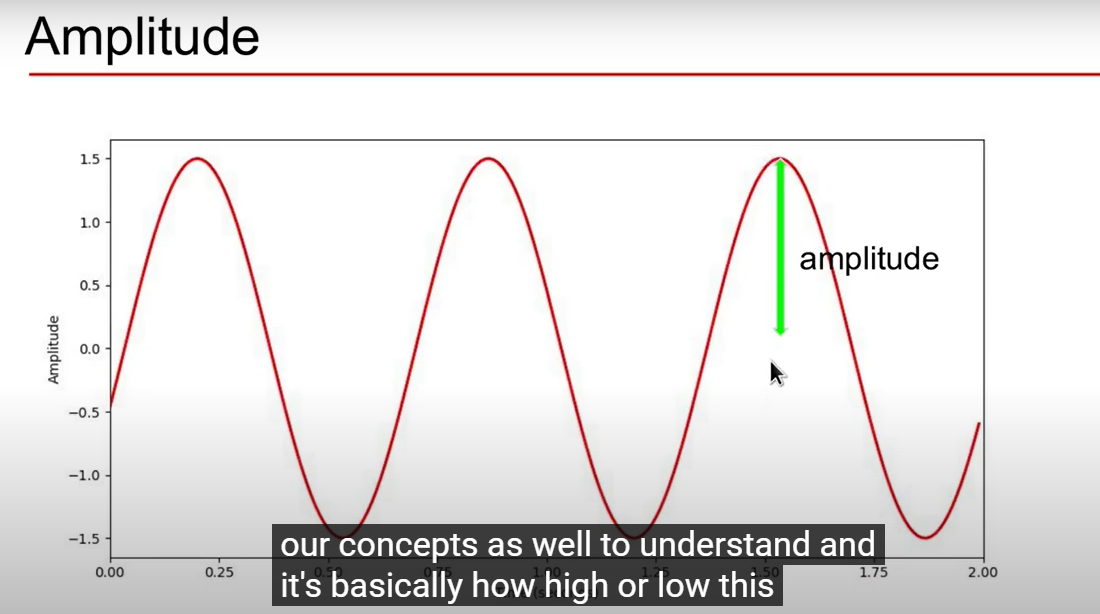
F is freq. T is time.

That symbol is phase.

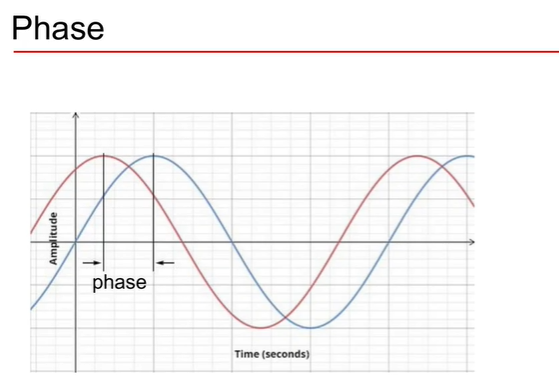
Terms:



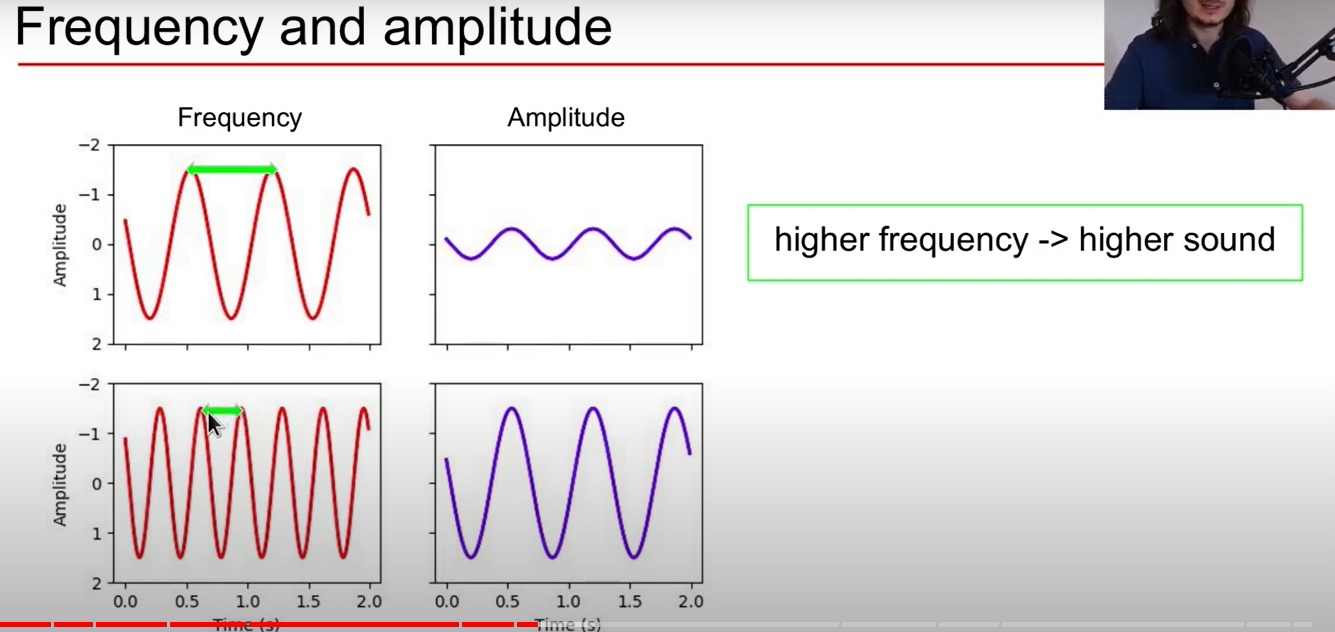
Frequency is inverse of period, expressed in Hz [Number of cycles per second]

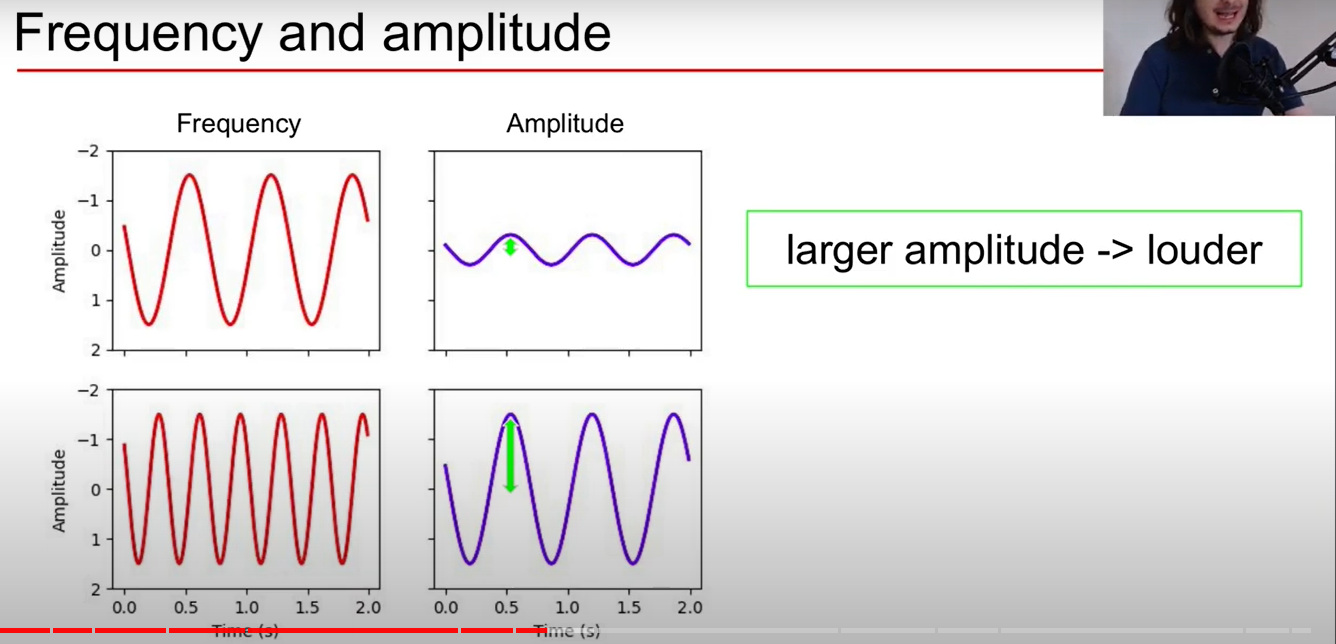


How high or low the perturbation goes.

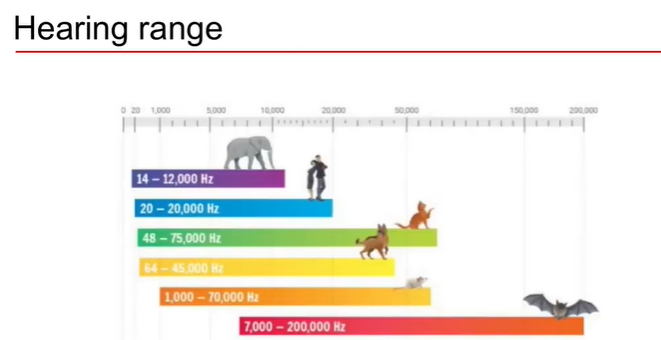


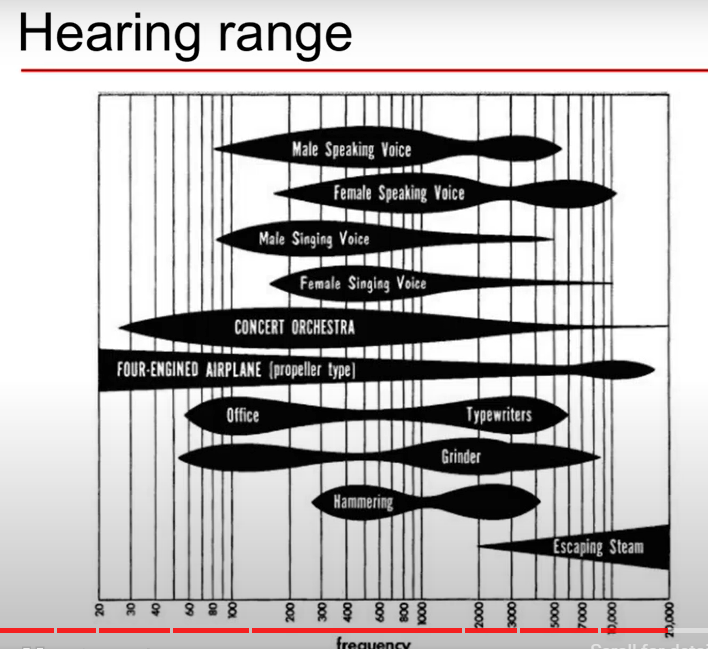
Enables us to shift waveform to right or left. Tells us position of the waveform at time 0.

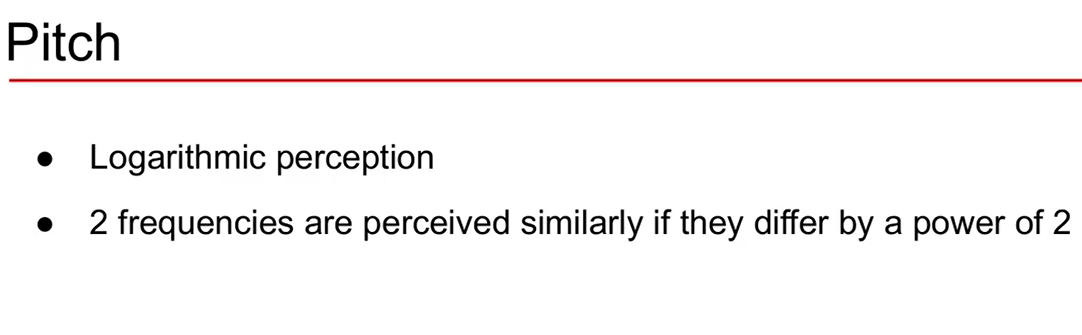




Coz amplitude measures amt of perturbation in air, so higher the perturbation louder would be the sound transmitted.

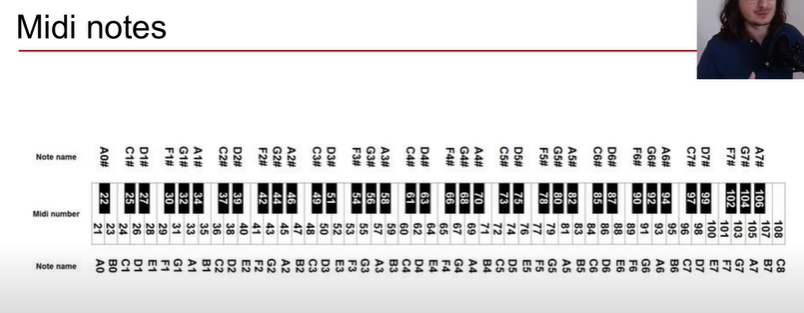






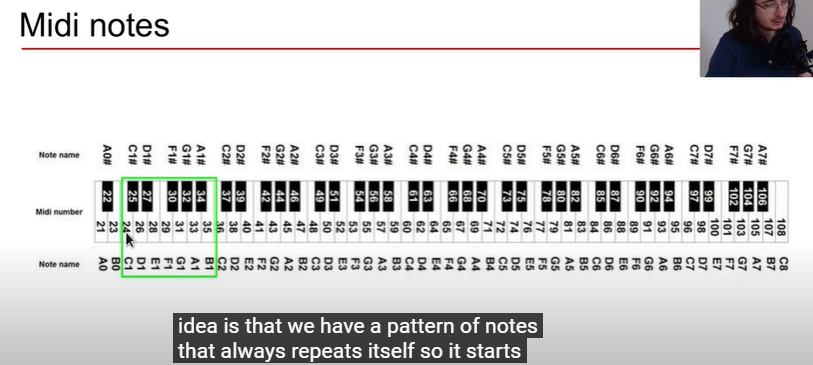
Pitch is the concept used for perception of frequency. We hear freq in logarithmic way. [so it’s v different from the og freq]

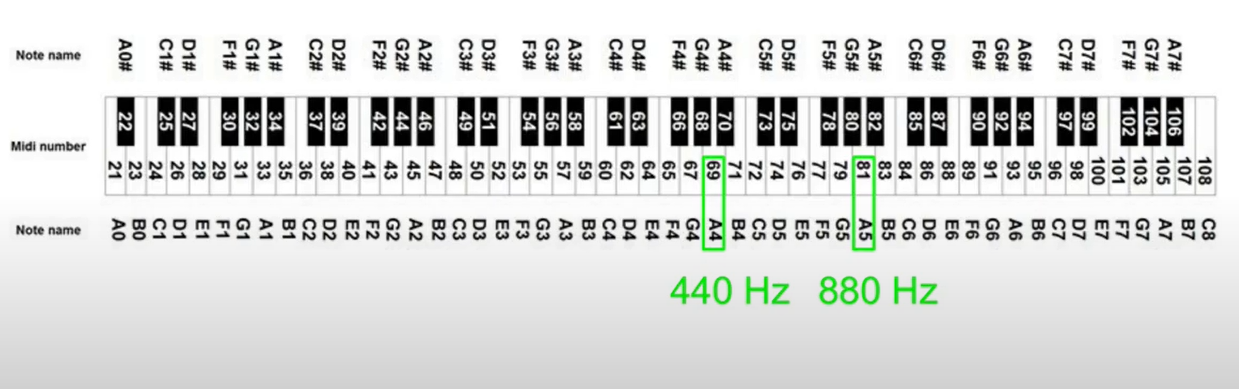
Power of two wala point is related to Octave concept, covered later. .



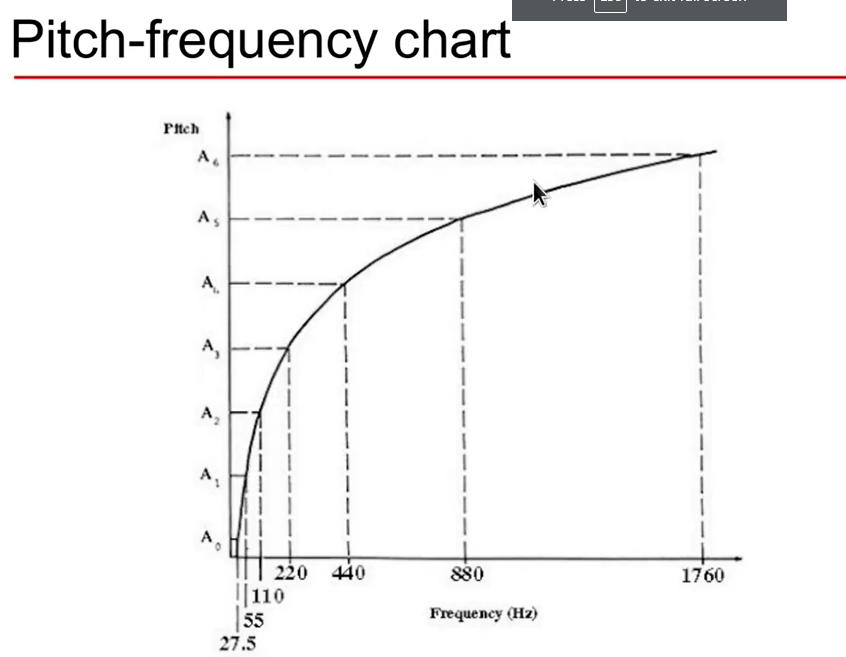
Piano keyboard. Eg: Midi Note 60 = C4. [we’re assigning numbers to each note]

Note = Letter [name] + number [Octave we’re at]

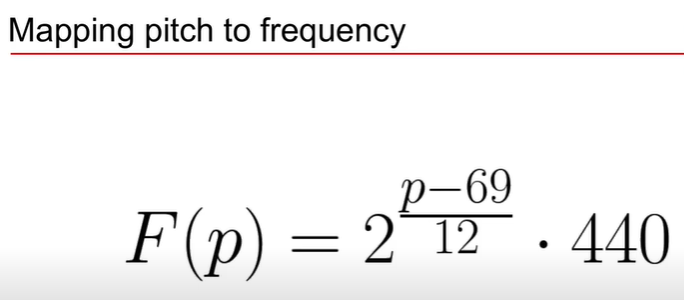
Pattern of 12 notes repeat and then we go up an octave. The same notes are called semitones [the ones that repeat]. Basically the notes with same letters sound similar but if the number is greater, the note sounds higher. 



Notes an octave above have double the frequency.

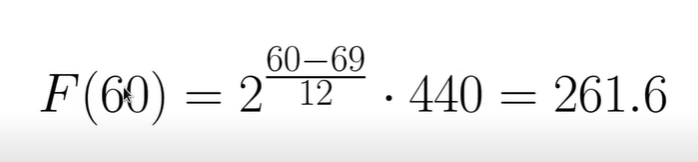


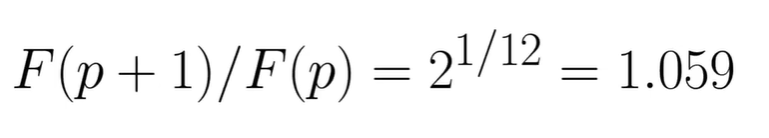
Logarithmic.



Freq F as function of p which is a Midi note. [Plug in p=69, the A4 Midi note we saw earlier, and we get back its freq as 440]

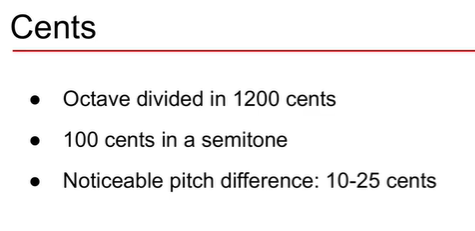
Example-





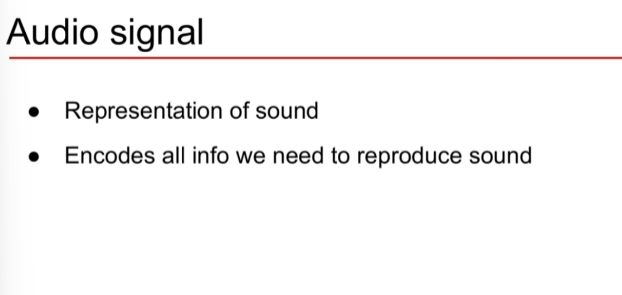
Relation between two subsequent semitones [pitches]

[remmbr how an octave is divided into 12 semitones]

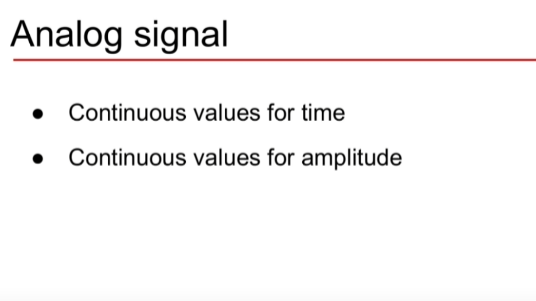


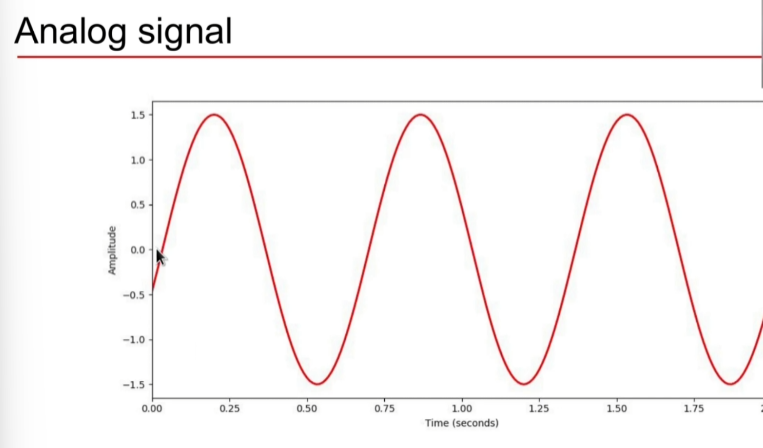
**4) Understanding Audio Signals for ML : -**

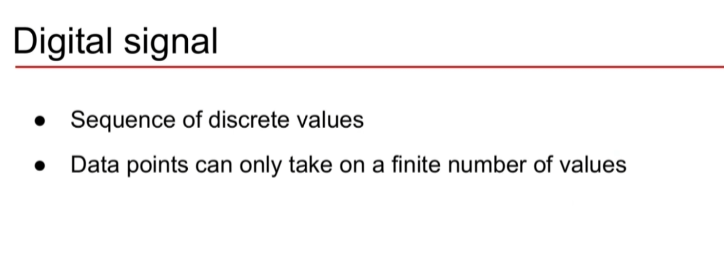
We can reconstruct a sound with the help of it’s audio signal.



The main focus - Convert Analog Signal to a Digital Signal.

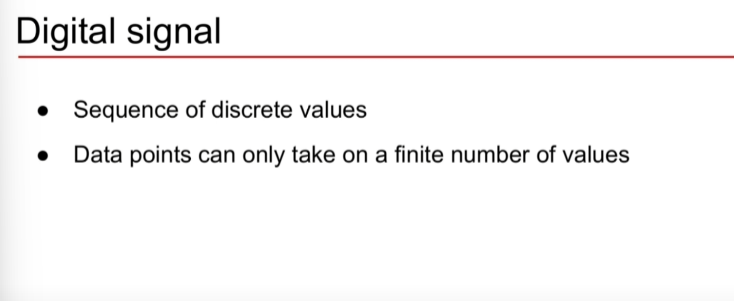


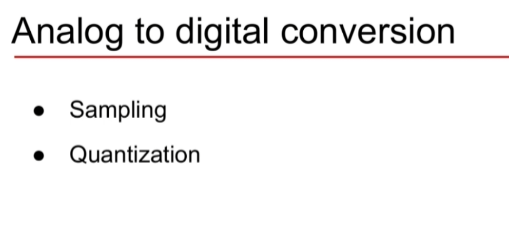


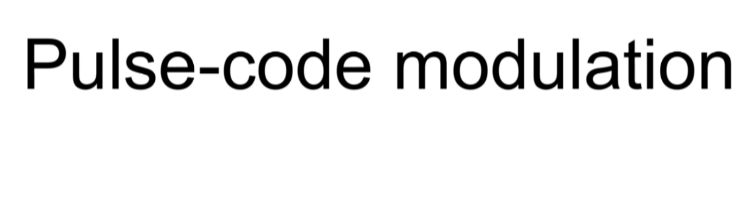


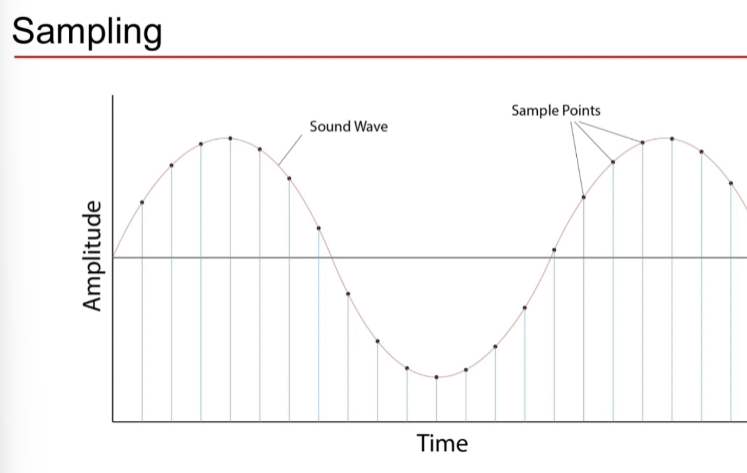
Infinite memory and time to store an analog signal.

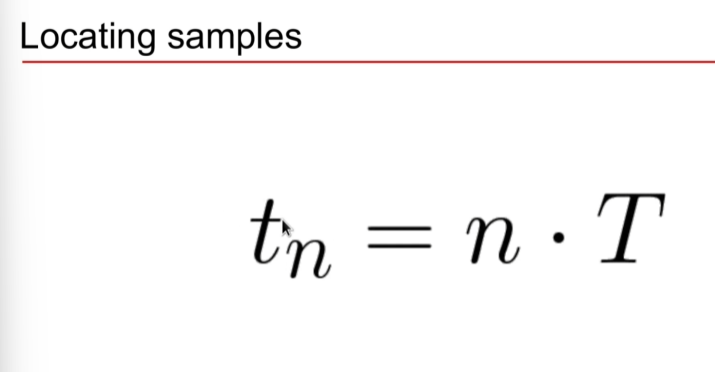
Digital Signal :

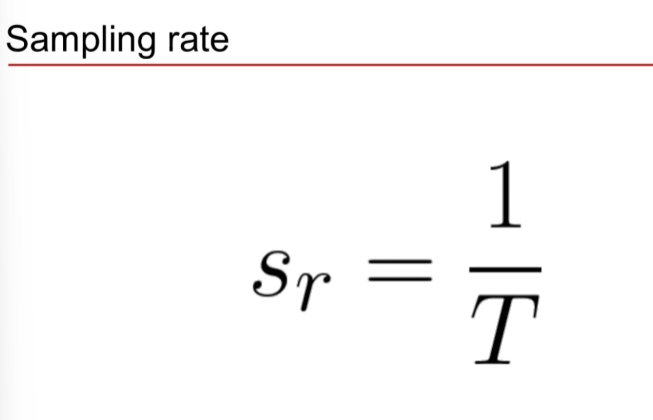


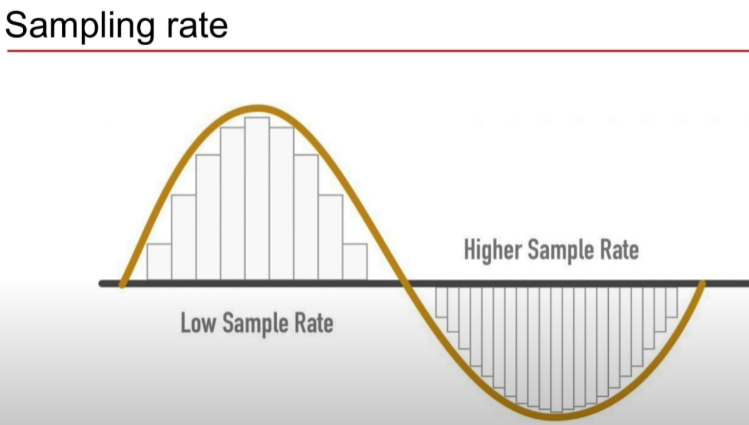




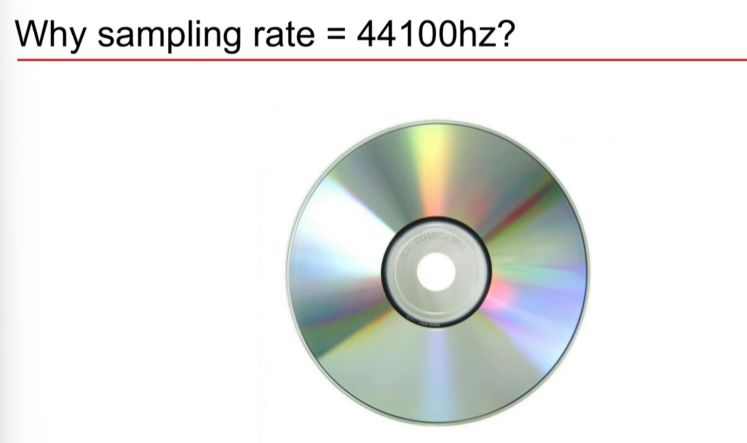


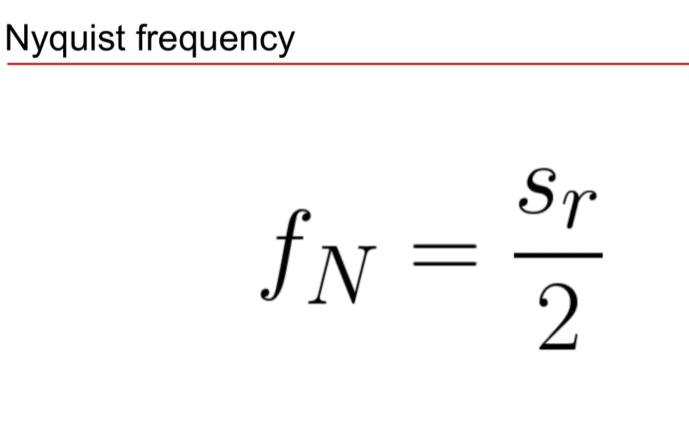


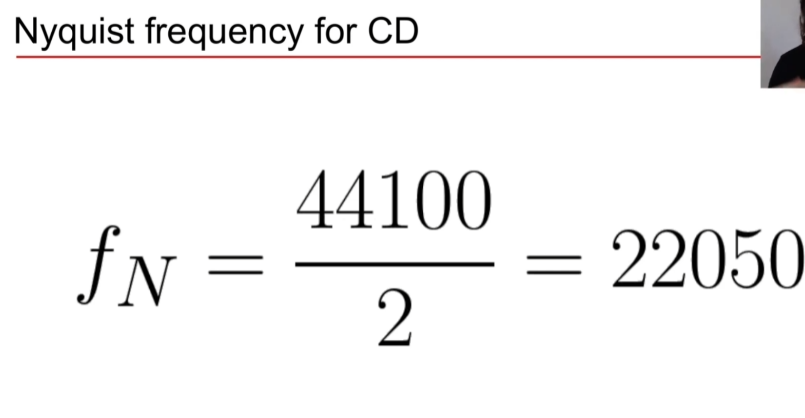




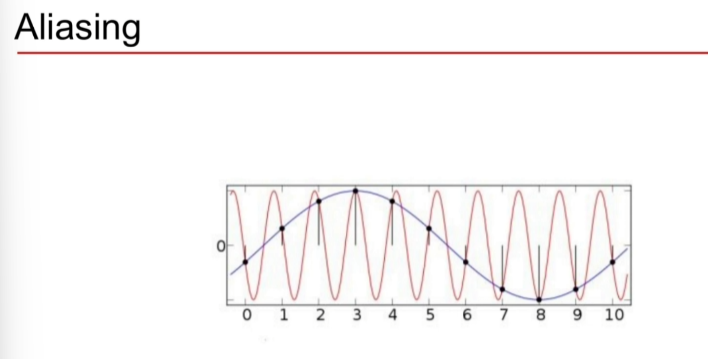
Why Sampling rate 44100hz ?

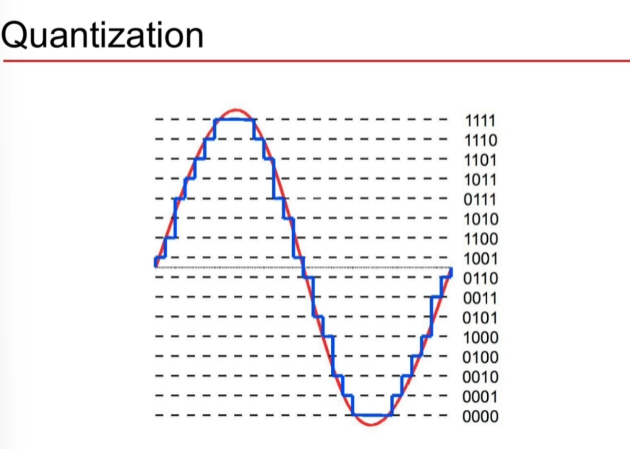


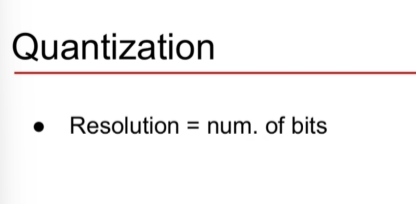


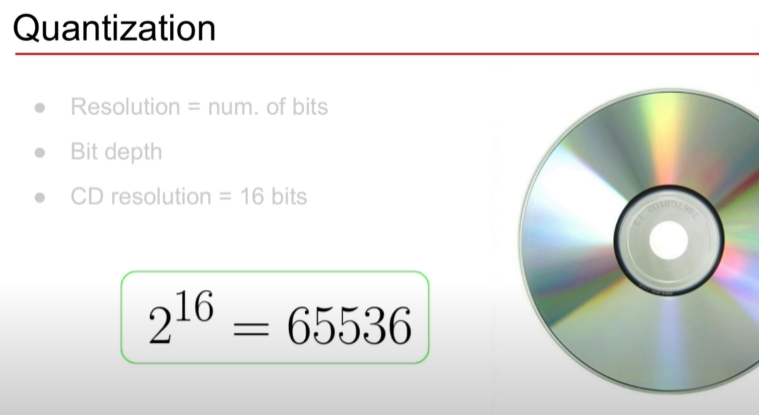


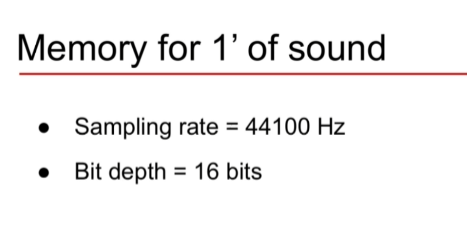
Undersampling leads to aliasing.

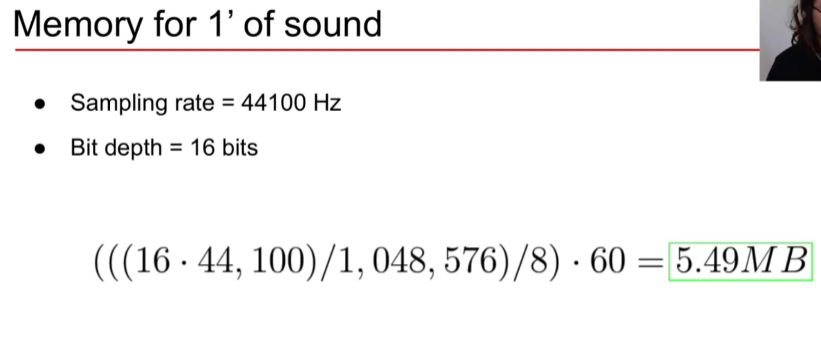


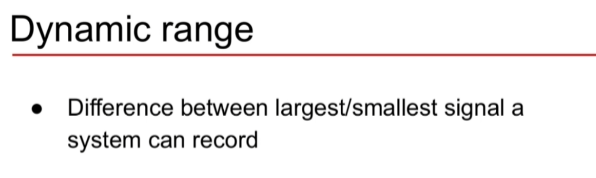


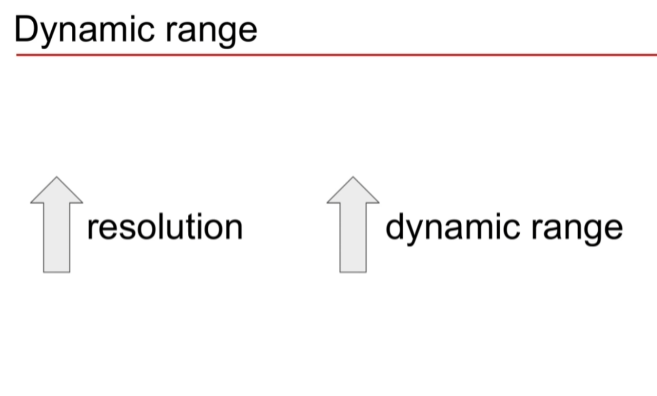




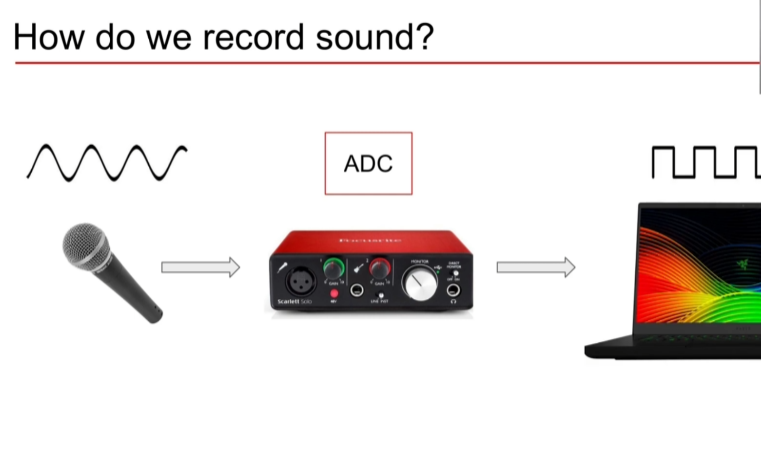


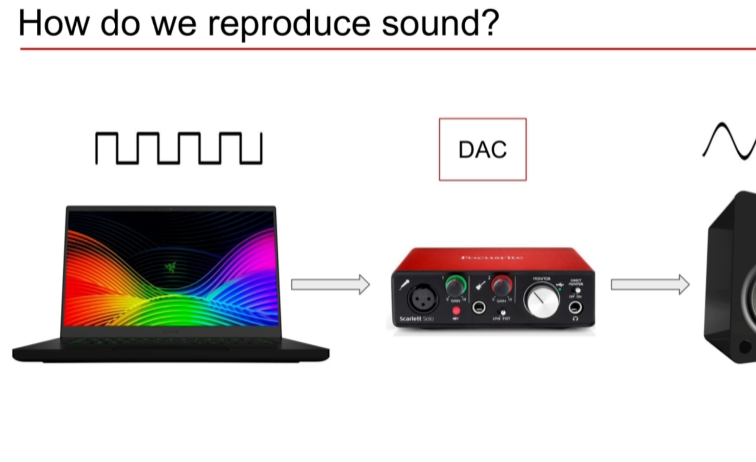




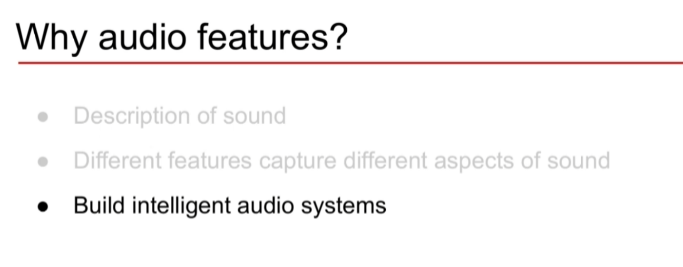


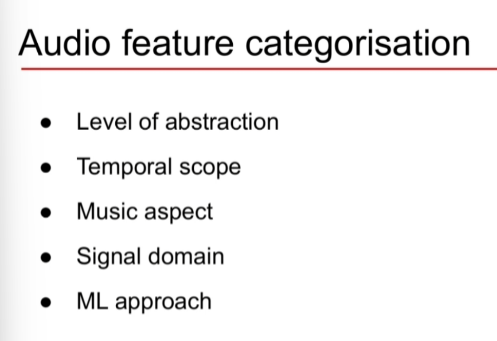


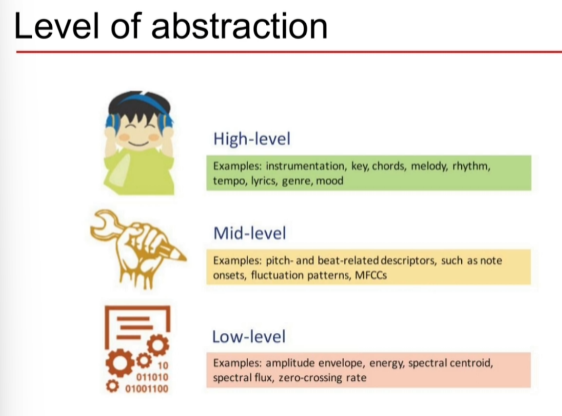


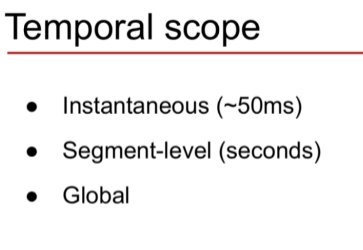


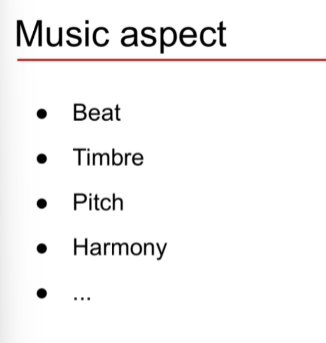
**Types of Audio Features for ML** : -



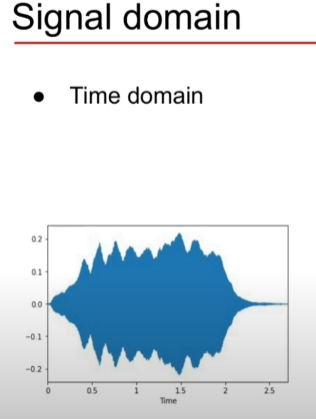




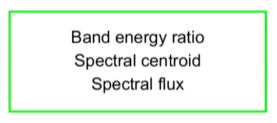
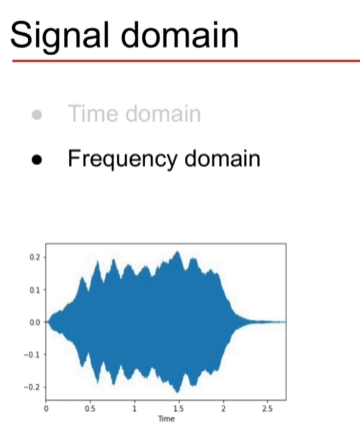








Time domain is capture through the waveform. Time domain audio features extract data from this representation.



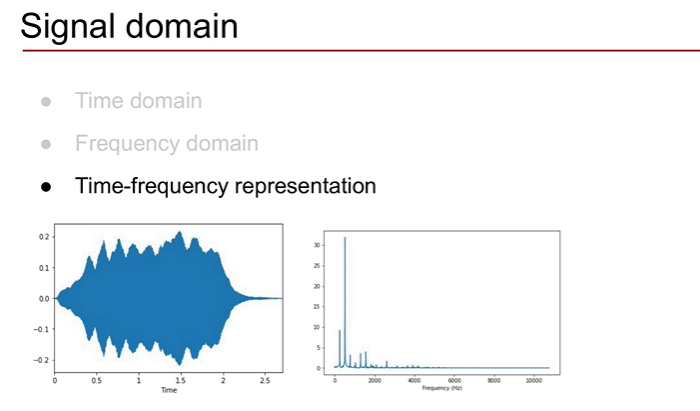


Time domain to frequency domain - Fourier transform.

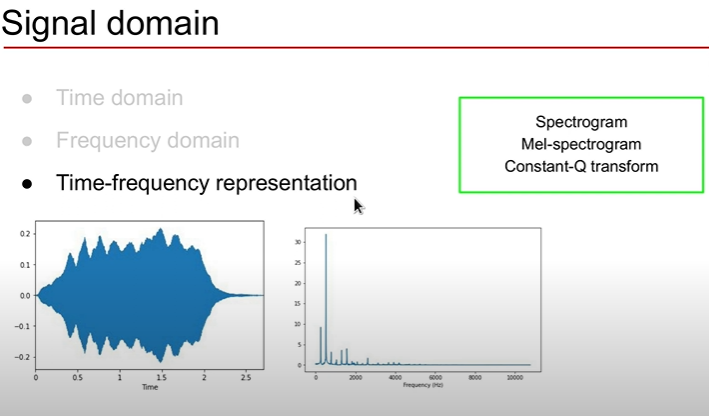
Here we are observing the description of all the frequencies that make up the sound.

The drawback of both of them being separated is we cannot of information about both of

them at the same time.



Time-frequency representation.

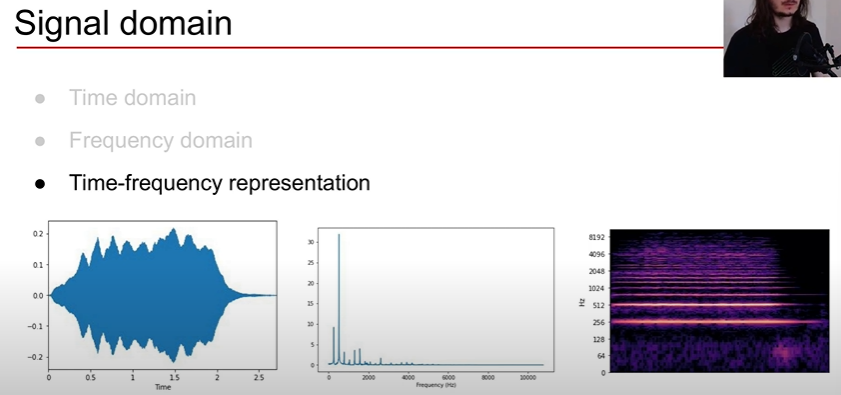


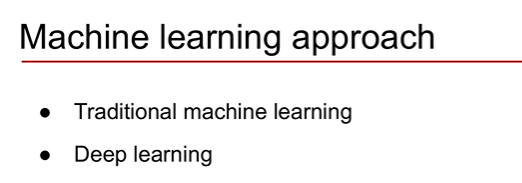
We apply stft to time frequency representation to get the spectrogram.

X -> time,

y-> frequency.

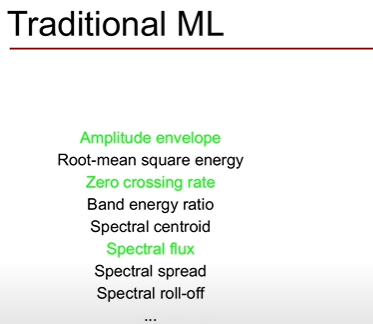
The more the color, the more the contribution of that frequency band at that time.

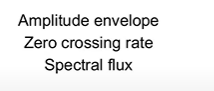


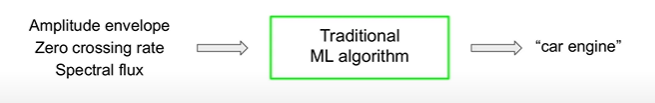




We consider all the features. After that we observe which ones are contributing most to the problem that we want to solve and choose them.

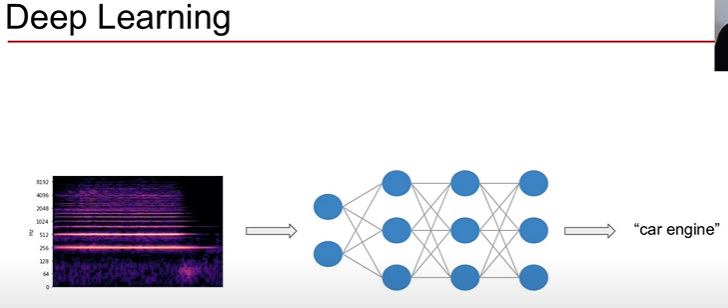




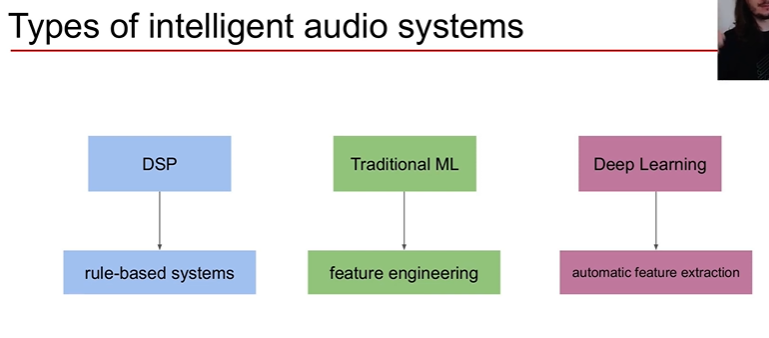


In Deep learning we tend to use un structured data.

In the image format we can pass time domain or a spectrogram or mfcc.



DSP -> digital signal processing systems were traditionally used, before the advent of Machine Learning.



We will be analysing time domain audio features and frequency domain audio features.

**How to extract audio features :**

Look at an analog signal, now to operate on it with our machine we need to have it in digital format. For that we need to sample and quantize it.

After we have this digitalized data,

1. The first step that we have to do is framing. Note that these frames are overlapped somehow.

