Notes I took from this amazing tutorial:

<https://youtube.com/playlist?list=PLwATfeyAMNqIee7cH3q1bh4QJFAaeNv0&si=m_d9Rh3AtDf6dmsl>

**Features**

**Waveform**

* **A diagram of different types of waves

  Description automatically generated**Carries information about: frequency, intensity, timbre

**Pitch (color)**

* Related to frequency (higher frequency = higher note)
* Hertz (Hz)
* Midi notes
  + Each note (semitone) with in an octave (different hues)
  + Each octave (progressing brightness)
    - each octave has 12 semitones/ notes
    - **A keyboard with black letters

      Description automatically generated with medium confidence**each octave has 1200 cents (each semitone has 100 cents)
  + A graph of a function

    Description automatically generatedthe same note sounds same across octaves, frequency progresses by 2^x
  + A black and white math equation

    Description automatically generatedMap pitch (p, midi number) to frequency (F)
  + For each subsequent note:

**A black text on a white background

Description automatically generated**

**Loudness (shape size)**

* Related to amplitude (higher amplitude = louder)
* Decibels (dB)
* Subjective measure of sound intensity
  + Sound power per unit area (W/m^2)
* Depends on frequency

**Timbre (shape type)**

* tone quality
* difference between two sounds with the same frequency, intensity, duration
* multidimensional
* 1) sound envelope
  + **A diagram of a curve

    Description automatically generated**Attack decay sustain release model
  + **A diagram of a waveform

    Description automatically generated with medium confidence**Eg.
* 2) harmonic content
  + Complex sound
    - Superposition of sinusoids (combination of sine waves)
    - Partial: a sinusoid used to describe a sound
    - Fundamental frequency: lowest partial (what normally give note names)
  + Harmonic partial: a frequency that is a multiple of a fundamental frequency (pitched instruments)
  + Inharmonicity: deviation form a harmonic partial (percussive instruments, noise)
  + Mel-spectrogram
    - Each line is a partial (very well defined) = very harmonic
    - **A screen shot of a computer screen

      Description automatically generated**A blue and orange lines

      Description automatically generated with medium confidenceThe line with the lowest Hz is the fundamental frequency
    - Left (violin) and right (piano) are differentiated by how energy is distributed throughout the partials
* 3) frequency modulation (vibrato)
  + **A diagram of a message signal

    Description automatically generated**Periodic variation in frequency
  + Common in violins
* 4) amplitude modulation (tremolo)
  + A diagram of different waves

    Description automatically generatedPeriodic variation in amplitude

**Digitization**

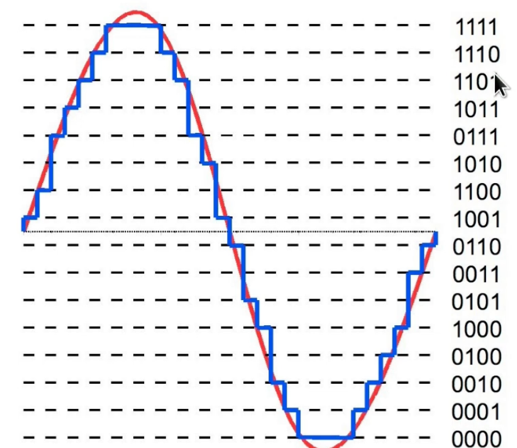
**Audio Signal**

* Representation of sound
* Encodes all the information we need to replicate a sound
* Analog signal
  + **A red line graph on a white background

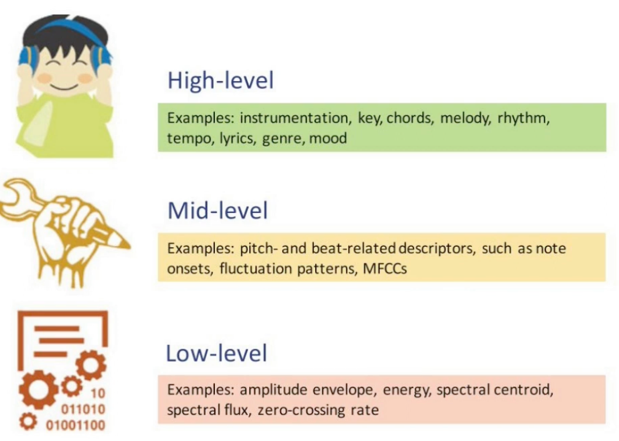
    Description automatically generated**Continuous for time and amplitude
* Digital signal
  + Sequence of discrete values
  + Data points can only take on a finite number of values
* Analog to digital conversion (ADC)
  + **A diagram of a sound wave

    Description automatically generated**Sampling (frequency)
    - **A black symbol with a white background

      Description automatically generated**Location samples
    - **A mathematical equation with numbers and symbols

      Description automatically generated**Sampling rate (Hz)
      * Higher sample rate = low sample error
  + Quantization (amplitude)
    - stored in binary
    - resolution = the number it bits (binary digit)

**Audio Features of ML**

* ****Levels of abstraction
* Signal domain
  + Time domain:
    - Amplitude envelope
      * Max amplitude value of all samples in a frame
      * Gives rough idea of loudness
      * Sensitive for outliers
      * Use for onset detection, music genre classification
    - **Root-mean-square energy** 
      * Rms of all samples in frame
      * Indicator of loudness
      * Less sensitive to outliers than AE
      * Use for audio segmentation, music genre classification
    - Zero crossing rate
      * number of times a signal crosses the horizontal axis
      * recognize percussive vs pitched sounds
      * monophonic pitch estimation
      * voiced/unvoiced (vocal cord vibration) decision for speech recognition
  + Frequency domain:
    - band energy ratio
    - spectral centroid
    - spectral flux
  + Time-frequency representation
    - Spectrograms (x= time, y = frequency)
    - A graph with numbers and lines

      Description automatically generatedmagnitude displayed through color rather than the y axis
* Time-domain feature pipeline
  + Frames
    - Audio chunks that are perceivable
    - Because 1 sample is too short to hear
    - # of frames is the power of 2 (matter of speed)
    - Typical range : 256-8192
    - Duration of a frame (ms):
      * A black and white math equation

        Description automatically generatedDuration of a single sample \* K (total # of samples)
  + **A graph with numbers and a graph

    Description automatically generated with medium confidence**Pipeline:
* Frequency-domain feature pipeline
  + use Fourier Transform to transform from time-domain to frequency-domain
  + problem: spectral leakage
    - processed signal isn’t an integer number of periods (sin wave cycle)
    - endpoints are discontinuous, translated into high-frequency components not present in the original signal
    - solve by windowing
  + windowing
    - eliminates samples at both ends of a frame
    - generates a periodic signal
    - hann window
      * multiply with sample to compress ends
    - A diagram of a sample

      Description automatically generated with medium confidenceproblem: loss signals at beginning and end of a frame
      * Solve by overlapping frames
  + A blue sound wave

    Description automatically generatedOverlapping frames
    - Yellow = overlapped area
    - Frame size= width of red box
    - Hop length = frame size- overlap
      * Amount of samples that we shift every time we take a new frame
  + **A diagram of a graph

    Description automatically generated**Pipeline:
* Fourier Transform
  + Purpose: Time domain -> frequency domain
  + compare original signal with sinusoids of various frequencies
    - for each freq we get a magnitude (abs of ft) and a phase
    - high mag = high similarity between the signal and a sinusoid
* Mel spectrograms
  + Humans perceive sound logarithmically
  + Purpose: time-frequency representation, perceptually-relevant amplitude and freqency representation
  + Mel scale
    - Logarithmic scale
    - A blue line graph with numbers

      Description automatically generatedEqual distances on the scale have same “perceptual” distance
  + Extracting Mel spectrogram
    - Extract Short-Time Fourier transform
    - Convert amplitude to dBs
    - Convert frequencies to Mel scale
      * Choose number of Mel bands (depends on the problem, reference piano keys ~88)
      * Construct Mel filter bank
        + Convert lowest/highest frequency to Mel
        + Create equally spaced points in the range of Mel based on # of Mel bands
        + Convert points back to Hz
        + Round to nearest frequency bin (due to the discrete nature of signals)
        + A diagram of a frequency

          Description automatically generatedCreate triangular filters (weight <1 = the dumbing down of signals)
      * A black text on a white background

        Description automatically generatedApply Mel filter banks to spectrogram
        + A close-up of a graph

          Description automatically generatedMel spectrogram= M\*Y= (# of bands, # of frames)
    - Used for:
      * Audio classification
      * Automatic mood recognition
      * Music genre classification
      * Music instrument classification
* Mel-frequency cepstral coefficients
  + Cepstrum
    - Developed to study echoes
    - A graph of a diagram

      Description automatically generated with medium confidenceUsed for speech and music processing over time
  + Speech = convolution of vocal tract frequency response with glottal pulse
    - Glottal pulse = frequency
    - A close-up of a graph

      Description automatically generatedA diagram of a sound wave

      Description automatically generatedShaping of vocal tract = timbre
    - Use Mel frequency because the nature of log allows as to study E and H individually
    - We are interested in spectral envelope
  + Pipeline
    - Use Discrete Cosine Transform (simplified version of FT)
    - Enable us to decorrelate energy in different mel bands
    - Reduce # of dimensions to represent spectrum
    - Traditionally 13 coefficients
      * Each MFCC describes a progressively finer detail of the sound’s spectral shape (the “spectral envelope”).
      * A blue and yellow squares

        Description automatically generatedA diagram of a diagram

        Description automatically generateda hierarchy of timbral information
    - Get the first derivative (Rate of change of MFCCs over time) or second derivative(Acceleration of MFCC change) of MSCCs
      * If I want the shapes to react to timbre change more efficiently (within a frame)