# Audio Signal Processing

## Librosa Tutorial & General DSP concepts

5 September 2020

In this document will provide examples of the important used functions and methods mostly used in audio processing and MIR tasks. It is considered that no relevant experience with these modules before. Let the fun start😁.

## Some facts about sound waves:

- Sound is a mechanical wave that propagates through air causing the air particles to have areas of compression and contraction as it moves along the way.

- The sound wave carries multifactorial information such as **frequency, intensity** and **timbre**, and since all frequencies carry information. These are the information we need to extract to perform several audio signal processing tasks whether in music information retrieval, speech recognition or speaker verification.

- Sound can be periodic or aperiodic ( Fig.1)

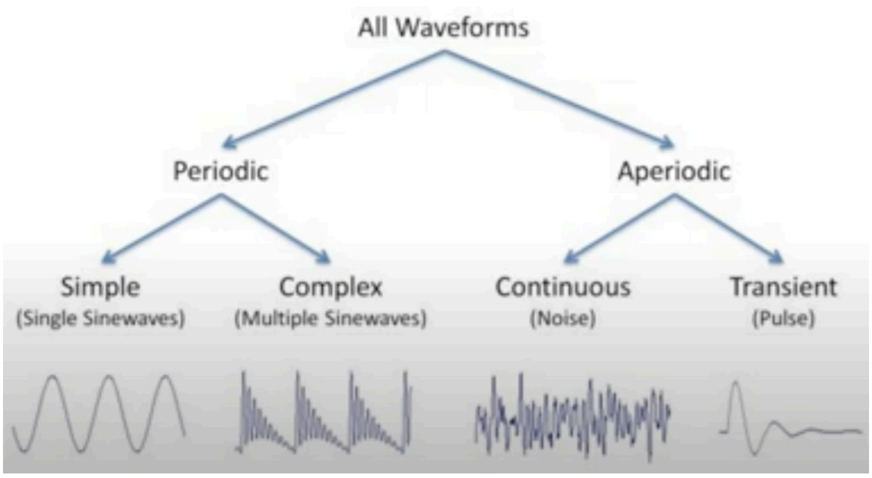


Figure 1 : Types of sound waveforms

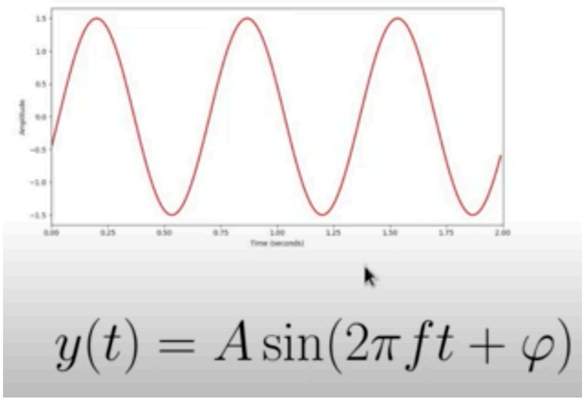


Figure 2: a simple sine wave

- Period is the distance between two peaks, Frequency= 1/ period.
- Period (T) is in Secs & Frequency in Hertz ( Expresses no. of cycles per second).
- Amplitude: is the quality that tels how high or low the waveform goes, starting from the midline from zero and then its length up or down.
- Phase: It enables us to shift the waveform to the right or to the left. It also tells us the position of the waveform at time zero.
- The higher the frequency the higher the sound, the lower or the shorter the amplitude.
- Larger amplitude is perceived as louder sound.
- The human hearing range is from 20 Hz - 20,000 Hz.
- Human hear or perceive sound logarithmically.
- Pitch is the concept we use as the perception of frequency, higher pitch notes means higher frequency sound waves.

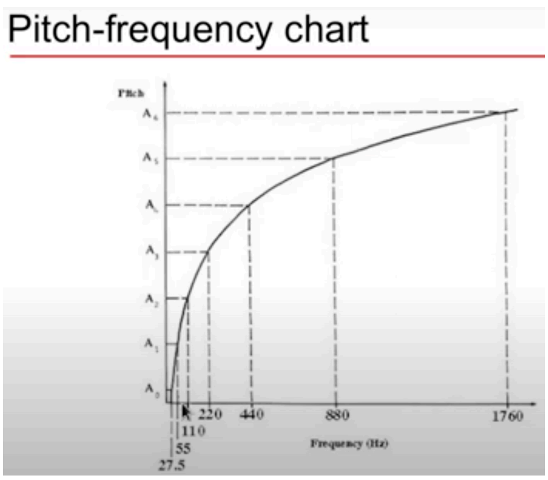


Figure 3: relationship between pitch and frequency.

In figure 3 we see that at every octave ( eg: A3 to A4) we double the frequency in hertz. it is very interesting because the way we receive sound and not only frequency but also amplitude and intensity is logarithmic and based on power of 2.

## Midi key note:

There is a common method ( equation) to find the midi key note and its corresponding frequency " **http://www.sengpielaudio.com/calculator-notenames.htm**".

## Sound features:

- **Sound Power** : Rate at which energy is transferred/ Energy per unit time emitted by a sound source in all directions ( Measured in watt"W").

- **Sound Intensity:** Sound power per unit area. Measured in W/m^2.

- Humans can perceive sound with very small intensities, called threshold of hearing "TOH". , TOH = 10^-12 W/m^2.
- Threshold of pain, TOP = 10 W/m^2.
- Intensity level is also on a logarithmic scale, because as we said before, human perceive sound and its features logarithmically.
- Intensity of sound is measured in decibel scale which is a logarithmic unit expressing the ratio between two values.
- dB(I) = 10. log10( I/I TOH) ——> Eq. that describes the intensity.
- Every time we go up by 3 dBs, intensity doubles.
- Check page 25 in Muller's Book " Fundamentals of music processing".

- **Loudness**: The same sound can be perceived to have different loudness depending on the individual. Age can affect the perception of sound. Also the duration of the sound affects perception of its loudness. For example a sound of duration 200ms is perceived to be louder than a similar sound of 50 ms.

- **Timbre**: It is the color of sound, it is the quality that tells which musical instruments is playing. It can be described with words like bright, dark, harsh, dull, warm.

**For example,** two sounds with same intensity, frequency and duration but with played different musical instruments.

**Timbre is multidimensional and have some features:**

- a- **Sound envelope:** Also called amplitude envelope. It passes by four stages known as ADSR, and as shown in below fig.4.

- It is obvious that the initiation of a sound have an attack, or spike in amplitude like hitting a piano key. Then there is decay where the sound stabilises.
- Then the sustain where sound becomes constant.
- Then the release where sound starts to fade

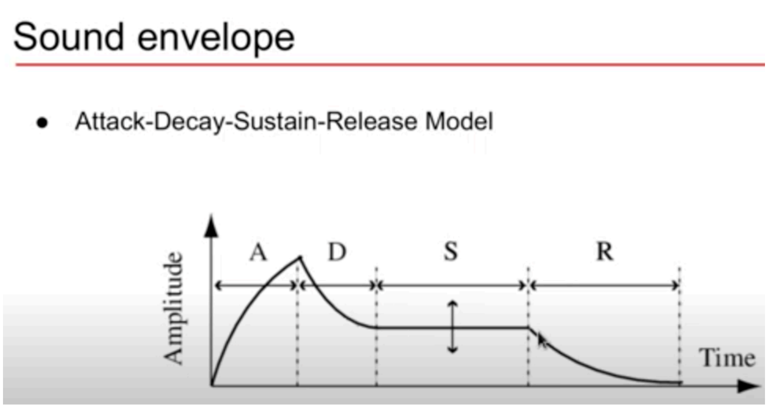


Figure 4: ADSR Envelope

ADSR envelope is a main component in many synthesisers, it helps to modulate aspects and features of an instrument's sound, often it's loudness over time.

The main idea is that when a mechanical musical instrument produces sound, the loudness of the sound produced changes over time and differs from instrument to instrument. for example when playing pipe organ, it plays a note at constant volume and the sound fades away quickly when releasing the key. On the contrary, a guitar's sound is loudest immediately after a string is plucked and quickly fades.

**The synthesizer's ADSR envelope** is a way to tailor and modulate the timbre for the synth. Modulation is like a quick attack with little decay makes it sound like an organ. A longer decay and zero sustain makes it sound more like a guitar.

It is known that envelopes are applied often for volume change but are also commonly used to control filter frequencies or oscillator pitches.

There is also ADSHR, which they added hold parameter after sustain level for a specific length of time. please read this ———> **https://www.wikiaudio.org/adsr-envelope/**

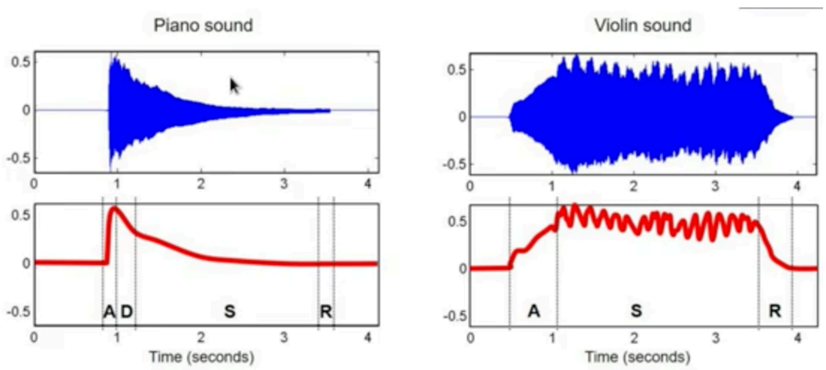


Figure 5: Sound envelopes for piano and violin showing the differences between ADSR in both

The attack in piano happens suddenly and short, while in violin its longer, and less sharp.

- Complex sounds are composed of many sinusoids that makes an audio signal.

- A partial is a sinusoid that used to describe sound.

- The lowest partial is called the **Fundamental Frequency**.

- **The fundamental frequency** is the lowest frequency sinusoidal in the sum.

- **A partial**: Is any of the sine waves of which a complex tone is composed, not necessarily with an integer multiple of the lowest harmonic " The fundamental". It also contains the fundamental frequency.

- **A harmonic** is any member of the harmonic series, an ideal set of frequencies that are positive integer multiples of the fundamental.

- **An inharmonic partial :** Is any partial that doesn't match an ideal harmonic. **Inharmonicity** is a measure of the deviation of a partial from the closest ideal harmonic, typically measured in cents for each partial. ——-> **https://en.wikipedia.org/wiki/Harmonic_series_(music)**.

- **An overtone:** Is any partial above the lowest partial, any frequency except the fundamental but multiples of the fundamental.

- A harmonic partial is a frequency that's a multiple of the fundamental frequency, f1=440 Hz. , second partial f2= 2*440=880 Hz, third partial f3=3*440= 1320 Hz, and so on, ….

- Inharmonicity indicates a deviation from a harmonic partial.

- Music instruments like pitched instruments are harmonic while percussive instruments are known to be inharmonic.

- **b- harmonic Content:**
- **c- Amplitude/Frequency modulation:**

---

**Facts about sine waves:**

- The sound used in music production is not a single sine wave, but rather is a combination of sine waves with varying amplitude, frequency and phases.
- Sinusoidal Functions posses an explicit physical meaning in terms of frequency. The composition of the sinusoids that make up the sound unfolds the frequency spectrum of the signal just like a prism that break up light into its spectral colors.

- The Fourier transform converts an audio signal that depends on time into a representation that depends in frequency or can be said converts a signal from the time domain to the frequency domain.
- There is two concepts known as **Vibrato** which means " **frequency modulation**" and **Tremolo** which is " **amplitude modulation**". Together can be called " Signal Modulation".
- Both Vibrato and Tremolo depend on two parameters, first is the extent of the variation and the rate at which the frequency or amplitude is varied. They both are expressed as local changes in intensity and frequency, it is not necessarily that they evoke the perception in loudness or pitch of the overall musical tone, but can be considered as features that influence the timbre of a musical tone.

**<u>- Some python implementations and functions that will be frequently needed for some tasks.</u>**

1- Checking if a file exists or not.

os.path module is a sub module of Os module used for common path name manipulation.

os.path.isfile('path') : It checks whether the specified path is an existing file or not ( " It returns True or False").

os.listdir('path') : It returns a list of all directories and files in the given path, if no specified path it return all dir. and files in the current working directory.

- To get the current working directory ——> os.getcwd()

---

# <u>Librosa</u>

- Loading an audio signal and plotting its waveform.

"Firstly, what is the amplitude envelope of a signal? " : The **amplitude** of a wave refers to the maximum amount of displacement of a particle on the medium from its rest position. In a sense, the **amplitude** is the distance from rest to crest. Similarly, the **amplitude** can be measured from the rest position to the trough position.

mono_signal, sr= librosa.load ('path of the audio signal' , sr=22050, duration=10)

NB: The default sampling rate is 22050, you can change and specify a sample rate for the signal ( Take in consideration sampling theorem to avoid aliasing).

Audio signal will be automatically resampled to 22050, if you want to preserve the original sampling rate of the audio file use sr=None.

Duration: Specify the duration of the signal, if you selected an audio file of 60 secs, you can frame it to 10 by selecting ( duration=10).

- Plot a mono waveform.

First, import matplotlib.pyplot as plt in order to plot the audio signals.

fig, ax = plt.subplots( nrows= ?, ncols= ?, sharex, sharey)

choose the number of rows and cols based on how many subplots you want. Based on your choice for nrows, ncols you will choose the axes.

For example: figure, axes= plt.subplot( nrows=2,ncols=2) , then when later plotting axes you have to specify its location because you have an array of 2 rows and 2 cols. Then your subplots can have the possible locations on the grid as ( axes[0,0], axes[0,1], axes[1,0], axes[1,1] ).

**- For more details on subplots ——->**

1- https://matplotlib.org/3.1.0/gallery/subplots_axes_and_figures/subplots_demo.html

2- https://jakevdp.github.io/PythonDataScienceHandbook/04.08-multiple-subplots.html

**- To control the shape and color , line style or marker of the graphs, here are the list of the string formats ———> https://matplotlib.org/2.1.1/api/_as_gen/matplotlib.pyplot.plot.html**

librosa.display.waveplot( mono_signal,  sr,  ax[0] )

<u>**Code for plotting mono and stereo signals:**</u>

```
import matplotlib.pyplot as plt
import librosa
import librosa.display
```

```python
mono_signal, sr= librosa.load('/Users/zaynkh/Desktop/
blues.00086.wav', duration=10)
fig, ax= plt.subplots(nrows=2)
librosa.display.waveplot(mono_signal, sr, ax=ax[0])
ax[0].set(title= 'mono')
```

```python
steroe_signal, sr2= librosa.load('/Users/zaynkh/Desktop/
blues.00086.wav', duration=10, mono=False )
librosa.display.waveplot(steroe_signal, sr2, ax=ax[1])
ax[1].set_title('stereo')
fig.tight_layout()
plt.show()
```

Note that the line " fig.tight_layout() " makes the subplot's title not to interfere with the one above it.

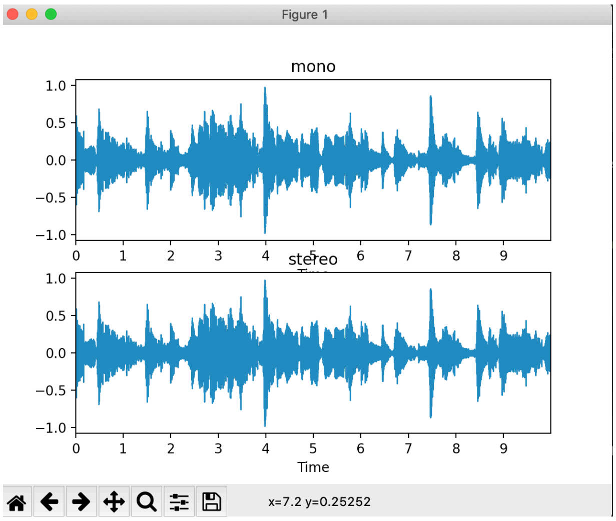


Figure 1 : Monophonic Signal

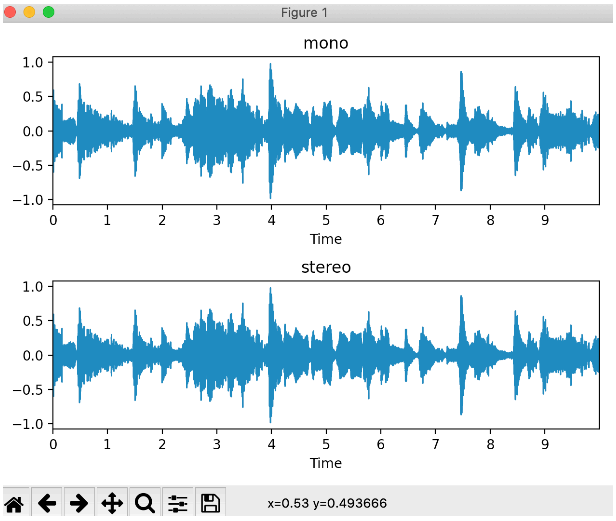


Figure 2: Stereo Signal