

OUTLINE

1. Definitions
 - a. Sound Waveform and Frequencies
 - b. Harmonics
 - c. Overtones
 - d. Pitch
2. Frequency Math
 - a. Sinusoidal Graph
 - b. Fourier Series
 - c. Spektrogram
 - d. MFCC
 - e. FFT
3. Tools
 - a. Spectrum Analyzer
 - b. Overtone Analyzer
4. Idea
 - a. After Main Project
5. Initial Work

Definitions

Sound Waveform and Frequencies

Frequency (Hz)

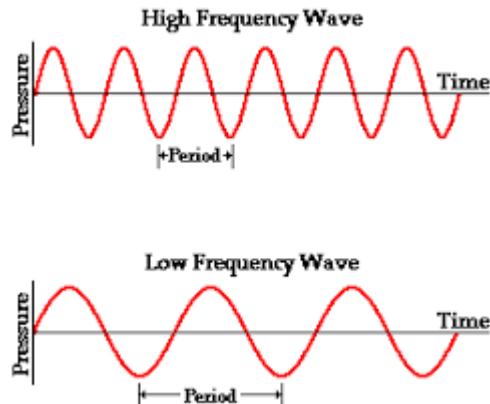
Each note has a frequency. For example, the frequency of the note "C" is generally accepted as 261.63 Hz. Frequency information is important for visualizing sound.

Waveform

Sound is represented as a waveform. This waveform usually has an amplitude value that changes over time, and you can manipulate this waveform.

Frequency Spectrum

You can use the Fourier transform (for example, FFT: Fast Fourier Transform) to visualize the different frequency components of sound.



Time-Domain Features

Amplitude Envelope: Captures the maximum amplitude within each frame, representing loudness and useful in onset detection (e.g., identifying when a note or event begins).

Root-Mean-Square Energy (RMS): Measures signal loudness by averaging the squared amplitudes within a frame, less sensitive to outliers than amplitude envelope, aiding in audio segmentation (e.g., speech versus silence).

Zero Crossing Rate (ZCR): Counts how often the signal crosses the horizontal axis within a frame, which helps distinguish percussive sounds from pitched ones, useful in monophonic pitch estimation and voice-unvoiced distinction.

Frequency-Domain Features

Band Energy Ratio (BER): Measures the energy ratio between lower and higher frequency bands, helpful in differentiating music from speech or classifying environmental sounds.

Spectral Centroid: Indicates the “brightness” of a sound by representing the frequency band where most energy is concentrated. Often used in music and sound classification, this feature is key in analyzing timbral properties.

Bandwidth: Shows the spread of energy around the spectral centroid, related to timbre. It's effective for music genre classification and pre-processing voice sounds for medical analysis.

These features help build machine learning models for sound classification, genre recognition, and speech analysis, each suited to different tasks in audio processing based on this article <https://sciendo.com/article/10.2478/ijasitels-2023-0006>

Not: Ses Dosyası Formatları

Ses verisini işleyebilmek için WAV, MP3, **FLAC** (matplotlib kullanıyor) gibi farklı formatlarda ses dosyalarını kullanılır. Ses dosyaları genellikle sıkıştırılmış formatta saklanır. Ancak FLAC ses formatı sıkıştırma yapmadan sesleri saklamaktadır.

Definitions

Harmonics

Harmonics are the frequencies of the harmonic series as they relate to a fundamental, which is the note you actually perceive.

When you play a C on the piano, while you only perceive that C, you actually hear several other notes above it being created by the vibrations of that same string.

- The string vibrates in full, creating the fundamental.
- It vibrates in halves, creating a C an octave up.
- It vibrates in thirds, creating a G above that.
- In fourths, creating a C above that.
- In fifths, creating an E above that.
- In Sixths, a G above.
- In Sevenths, a Bb above.
- The list goes on into infinity. It is these extra notes that give an instrument its timbre, the tone that makes it unique from other instruments.

It's additional tones that are vibrated when you play the fundamental. The fundamental is the first harmonic of the note you're playing. Every additional harmonic is an octave above it. Together, it's called the harmonic series. It's what creates timbre.

Harmonics are series of notes that play along with the notes you play on an instrument, but gradually rise and fall with each note. They are what give instruments their sound or "Thumb".

Trivia What are harmonics?

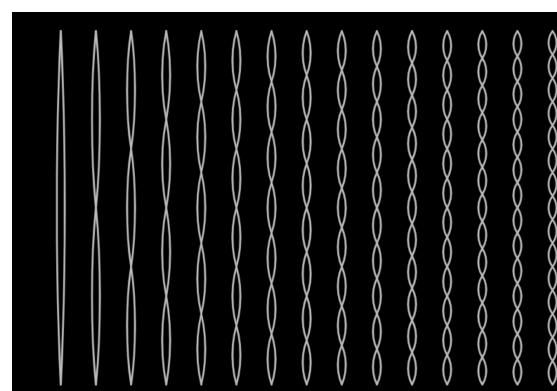
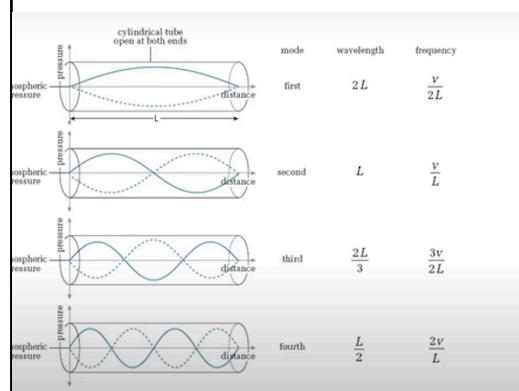
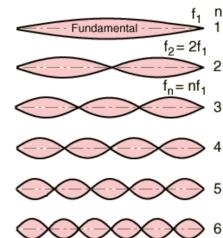
onic is a sound wave that has a frequency that is an integer multiple of a fundamental tone. The lowest icy sound that can be produced on the tube is the fundamental tone frequency. The frequency twice that of damental tone is the second harmonic, and the frequency three times the frequency of the fundamental tone rd harmonic.

ample, if the fundamental tone is C, then its second harmonic is C in the next octave up, and the third will be G ve and a perfect fifth higher, and the fourth will be C two octaves up. From low to high, the harmonics will be , C, E, C, A.

struments can produce harmonics with all pitches, but the volume at which various harmonics are produced based on the instrument, the player, and the pitch. This combination of harmonics is the very thing that creates pre of the instrument.

representative example of something that creates a pure sound, or only a fundamental tone with no harmonics, ing fork. For more about harmonics, we recommend Yoshinori Ando's *Instrument Acoustics* and Atsushi oto's *The Science of Instruments* (Blue Backs series).

The difference is HARMONICS! Different instruments have different harmonics on the frequencies (notes):



Appearance of harmonics in wind instruments and stringed instruments Available at:
<https://alexanderchen.github.io/harmonics/> for stringed instruments

Definitions

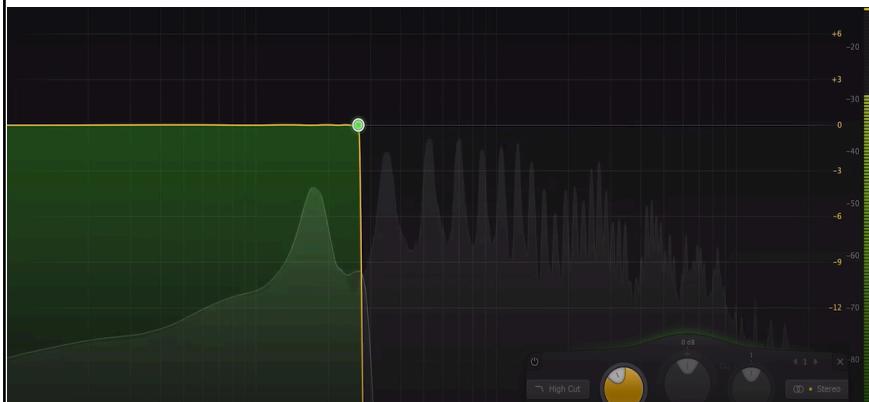
Overtones

Integer multiples of a note's fundamental frequency. Overtones don't have their own overtones.

Overtones are any resonant frequencies above the fundamental frequency of a sound. (Overtones may or may not be harmonics.) In other words, overtones are all the pitches higher than the lowest pitch within a single sound, the fundamental being the lowest pitch.



Guitar's First Overtone (Technically second harmonic) is stronger than the fundamental note



EQ Curve: EQ curves are standardized recording characteristics that ensure records are mastered and manufactured with equalization that ensures a smooth listening experience at the beginning of the note. In the video

Adding the brick wall EQ curve you can clearly hear that, both guitar's and clarinet's sounds are identical when the adding EQ curve.

Source: https://www.youtube.com/watch?v=Wx_kugSemfY

Result Of This Analysis : When we leave the overtones behind and extract only the fundamental note from the sound, the resulting wave is a sine wave.

Definitions

Pitch

A tone that has no harmonics is a sine wave therefore all harmonics are sine waves. Sine wave is the purest tone possible it trully is just a single pitch.

Pitch, sesin **algılanan yüksekliği veya alçaklığıdır** ve temel olarak sesin frekansına dayanır. Ancak, harmonikler, timbre, ses yoğunluğu ve insan kulağının frekansa duyarlılığı gibi etkenler, pitch algısını etkileyebilir.



What is pitch?

- In music the **pitch** of a sound means how **high** or **low** a note is.
- Musical pitches are relative, any note is **higher** or **lower** than another.
- It is the degree of relative **highness** or **lowness** of a sound

Pitch

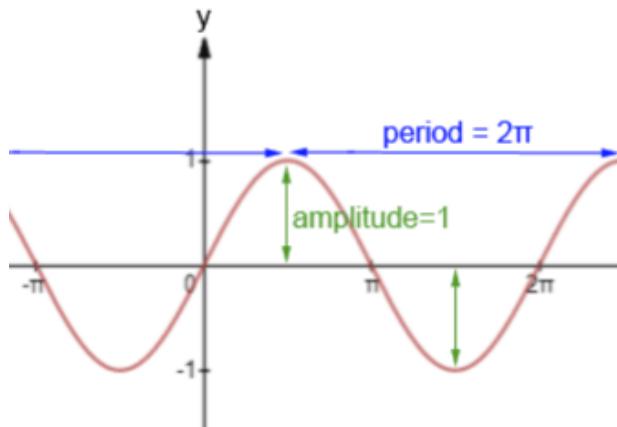
In music, pitch is the perceived frequency of a note or sound, which determines how high or low we hear a sound. In Western music, there are 12 pitches, all of which correspond to a letter, such as A or C. Pitch is one of the core elements of music and is fundamental to constructing melodies and harmonies. It's also one of the first things we notice when we hear a sound and helps us to distinguish one sound from another. In simple terms, pitch refers to how high or low a sound is, determined by the frequency of its vibrations. It's the attribute that allows us to classify sounds as being high, like the chirp of a bird, or low, like the growl of an engine. To bring our example to the world of musical instruments, you can think of the sounds played with the left hand on a piano as low pitch, whereas the keys on the right produce a high-pitched sound. A high frequency (e.g., 880 hertz [Hz; cycles per second]) is perceived as a high pitch and a low frequency (e.g., 55 Hz) as a low pitch.

Sesin tonu, frekansı ile ilgilidir. Bir ses notasını görselleştirirken, frekans değerleri ile notaların ilişkisini kurmanız gereklidir.

Frequency Math

Sinusoidal Graph

The term sinusoidal is used to describe a curve, referred to as a sine wave or a sinusoid, that exhibits smooth, periodic oscillation. Sinusoids occur often in math, physics, engineering, signal processing and many other areas.

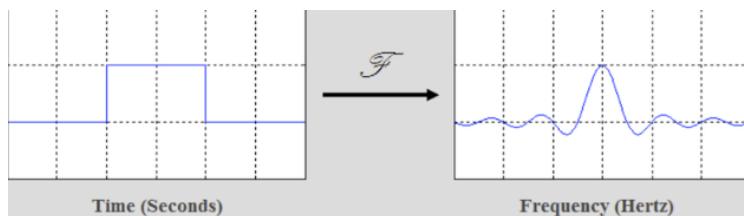


Sinusoidal function formula: $y = A \cdot \sin(B(x-C)) + D$

Frequency Math

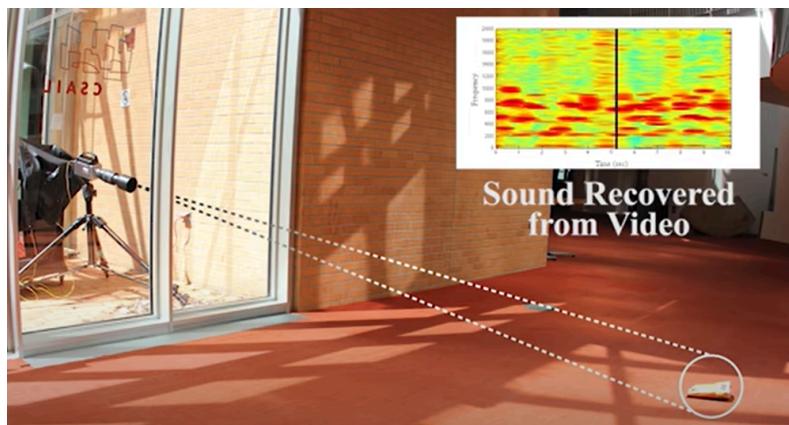
Fourier Series

Fourier Dönüşümü, bir dalga formunu (bir fonksiyon veya sinyal) değişen frekansların sinüs ve kosinüs fonksiyonları ile karakterize edilen alternatif bir gösterime bölen bir araçtır. Fourier Dönüşümü, herhangi bir dalga formunun sinüzoidallerin toplamı olarak yeniden yazılabilceğini gösterir.

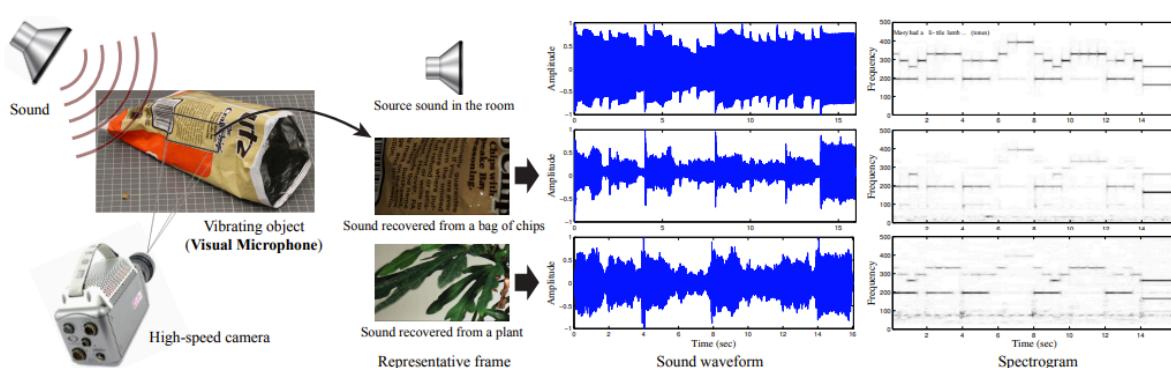


Evrende ne çizerseniz çizin veya gözlemlerseniz gözlemleyin, bütün dalga formları aslında farklı frekanslardaki basit sinüzoidlerin toplamından ibarettir.

Aşağıdaki videoda Fourier Series açıklanmış ayrıca Georgia Tech'de öğrenci olan Doğa'nın (Postdoctoral Scientist) <https://bilimneguzellan.net/fuyye-serisi/> adlı blogundan açıklama yapılmış Kaynak: <https://www.youtube.com/watch?v=ds0cmAV-Yek>



Doğa'nın blogunda yer alan Fourier Series ile ilgili bir çalışma: Videodan ses elde edimi (breathtaking) Fourier analizi kullanılarak bir videodan ses çıkarılabiliriyor. Yani, komşunuzun salonundaki çiçeği kamera ile sessiz olarak görüntüleyorsunuz, geliyorsunuz ve komşunuzun sizin arkanızdan çevirdiği dedikoduları dinleyebiliyorsunuz (Doğa).



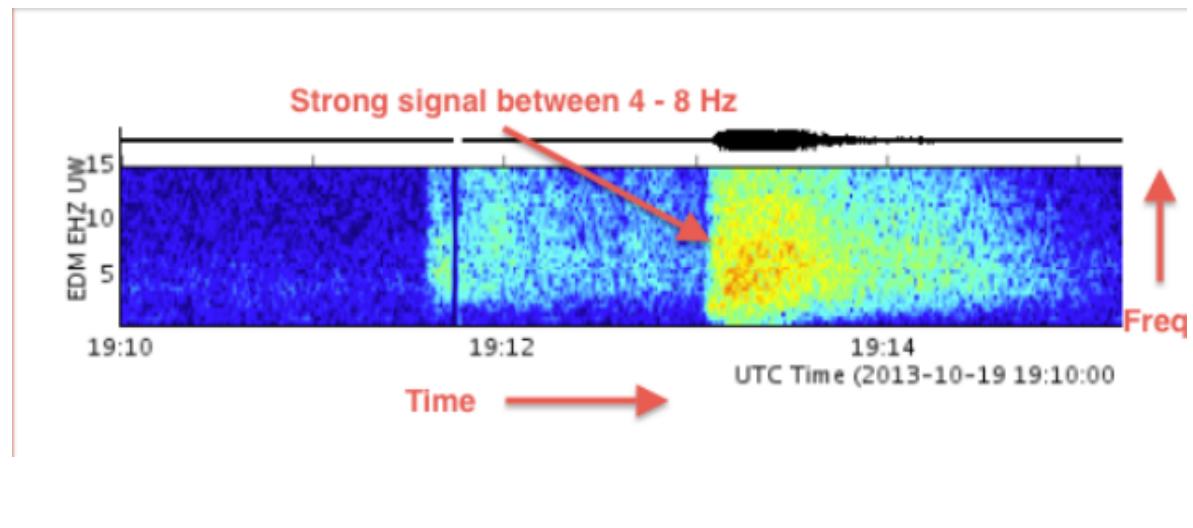
Project Webpage: <https://people.csail.mit.edu/mrub/VisualMic/>

Frequency Math

Spektrogram

Bir spektrogram, belirli bir dalga formunda bulunan çeşitli frekanslarda bir sinyalin sinyal gücünü veya "ses yüksekliğini" zaman içinde temsil etmenin görsel bir yoludur.

Örneğin 2 Hz ile 10 Hz'de daha fazla veya daha az enerji olup olmadığını görmekle kalmaz, aynı zamanda enerji seviyelerinin zaman içinde nasıl değiştiğini de görebilirsiniz.



Frequency Math

MFCC

In sound processing, the mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency.

Mel Frequency Cepstral Coefficients (MFCCs) refer to a set of features developed to analyze seismic sound echoes and model human voice characteristics. They are basic sound features used in various applications, obtained by taking the discrete Fourier transform of a signal, applying a logarithm, and then taking a Fourier inverse.

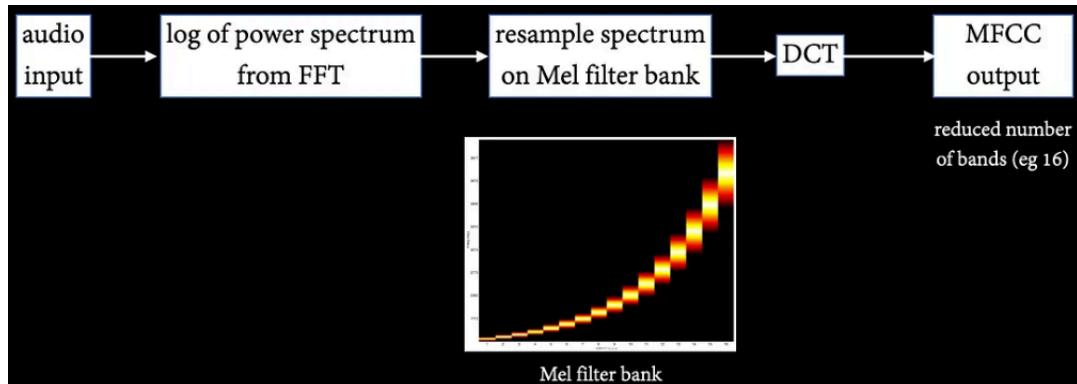
In Python Programming Language have a library that named "Librosa"

MFCCs are commonly derived as follows:^{[2][3]}

1. Take the [Fourier transform](#) of (a windowed excerpt of) a signal.
2. Map the powers of the spectrum obtained above onto the [mel scale](#), using [triangular overlapping windows](#) or alternatively, [cosine overlapping windows](#).
3. Take the [logs](#) of the powers at each of the mel frequencies.
4. Take the [discrete cosine transform](#) of the list of mel log powers, as if it were a signal.
5. The MFCCs are the amplitudes of the resulting spectrum.

There can be variations on this process, for example: differences in the shape or spacing of the windows used to map the scale,^[4] or addition of dynamics features such as "delta" and "delta-delta" (first- and second-order frame-to-frame difference) coefficients.^[5]

The [European Telecommunications Standards Institute](#) in the early 2000s defined a standardised MFCC algorithm to be used in [mobile phones](#).^[6]



Source:

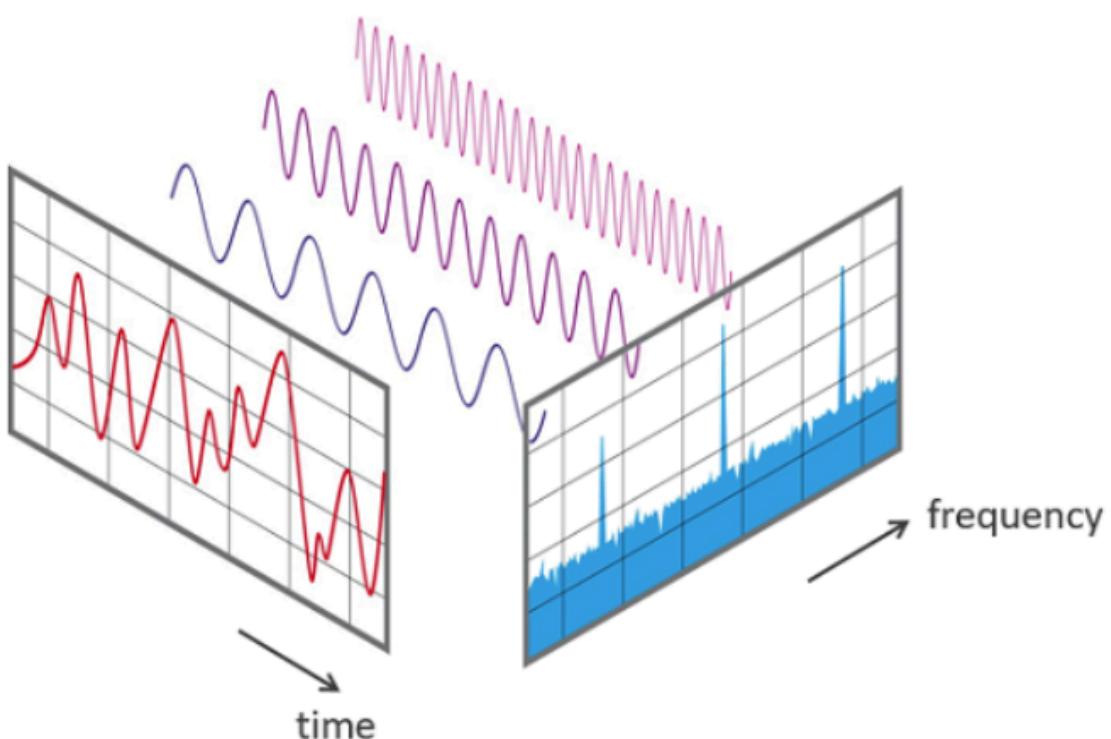
<https://medium.com/@derutycsl/intuitive-understanding-of-mfccs-836d36a1f779>

Frequency Math

FFT

The "Fast Fourier Transform" (FFT) is an important measurement method in the science of audio and acoustics measurement. It converts a signal into individual spectral components and thereby provides frequency information about the signal. FFTs are used for fault analysis, quality control, and condition monitoring of machines or systems. This article explains how an FFT works, the relevant parameters and their effects on the measurement result.

Sesin frekans bileşenlerini analiz etmek için kullanılır. Ses dosyasındaki frekansları görselleştirmeniz gereğinde bu dönüşüm çok faydalıdır.



View of a signal in the time and frequency domain

Source:

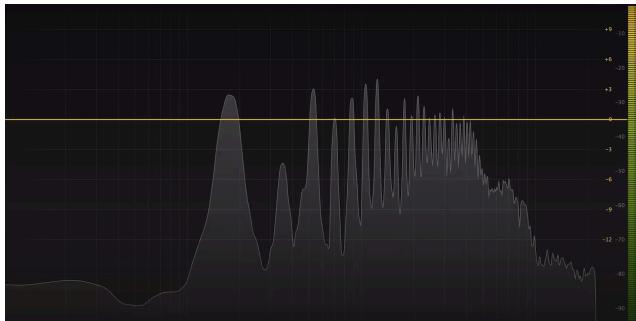
<https://www.nti-audio.com/en/support/know-how/fast-fourier-transform-fft#:~:text=The%20%22Fast%20Fourier%20Transform%22%20>

Tools And Use Case

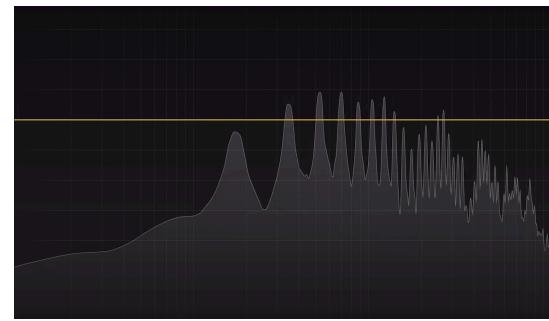
Spectrum Analyzer

A spectrum analyzer is a device that measures and displays signal amplitude (strength) as it varies by frequency within its frequency range (spectrum).

In the images below, the same note was recorded both Clarinet and Guitar on a computer and analyzed with Spectrum Analyzer.

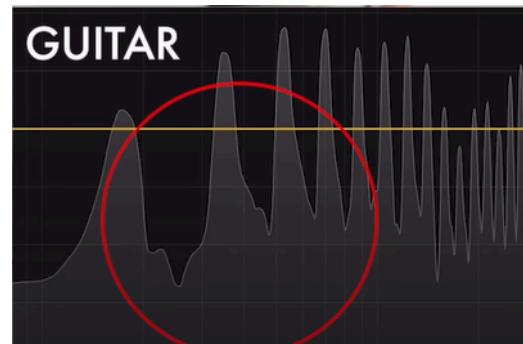
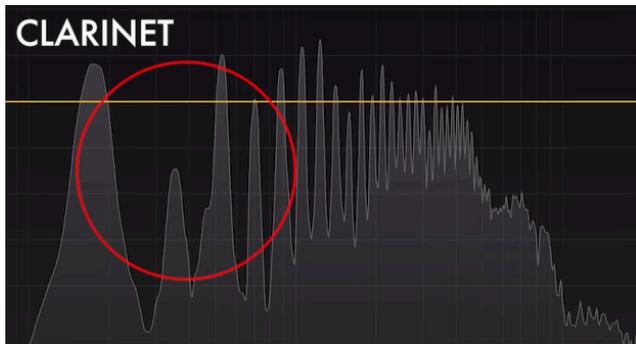


Spectrum Analyzer - Clarinet



Spectrum Analyzer - Guitar

Even though these two instruments are different, their fundamental tones are very similar.



Tools And Use Case

Overtone Analyzer

A screenshot of a Google search results page. The search query "sygt overtone analyzer" is entered in the search bar. The results show a link to "Sygyt Software – Software for Voice Analysis and Sound ...". The snippet below the link describes the software as making it easy to visually compare audio files and suited as a feedback tool for overtone singing.

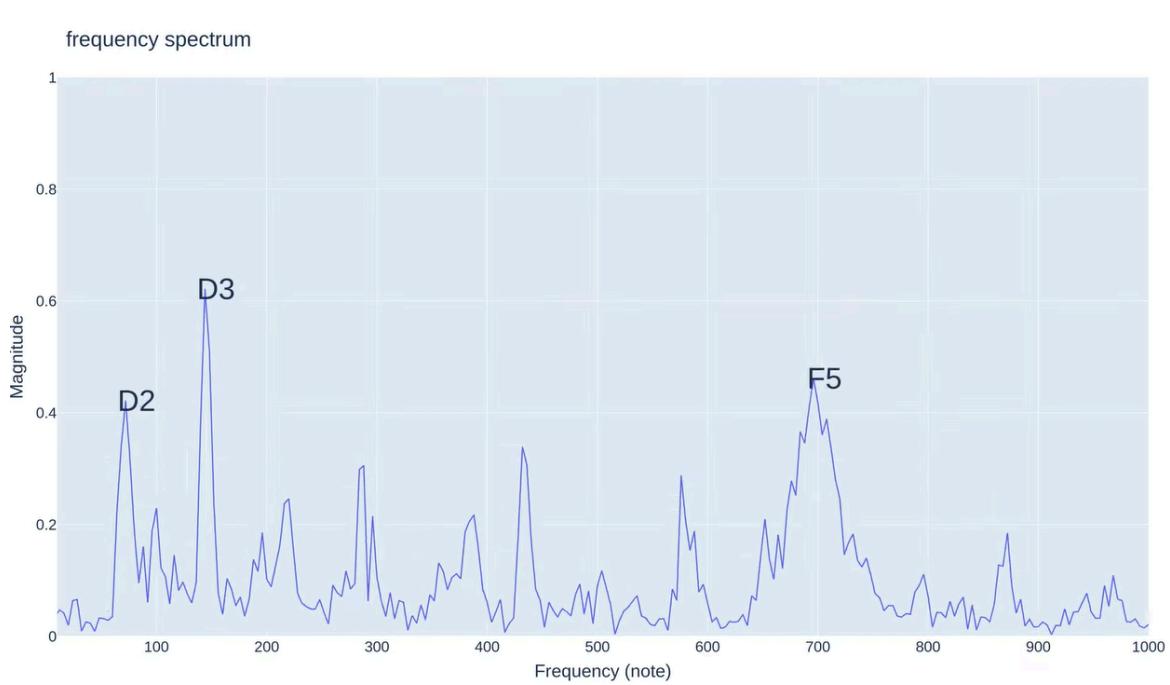
Idea

After Main Project

Harmonikler teoride sonsuz adet olabailir. Bir resimdeki piksellerin belirli bir temel nota ile çarpılarak görüntünden ses dönüşümü yapılabilir. Bu seslerin anlamlı olabilmesi için CLIP, Transformer, LLM gibi yapılarla donatılarak resimlereden çıkarılan özelliklere çeşitli formüller aracılığıyla ses verilebilir. Örneğin aşağıdaki resimden çıkarılan çeşitli notalarla, resim içeriğinin semantik kombinlenmesi sonucunda temaya uygun müzikler oluşturulabilir. Bu fikir, ana projeden farklı olarak daha sonra implemente edilebilir.



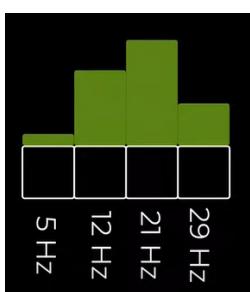
Initial Work



I have implemented “audio to graph” as initial work. This process based on FFT method. Assume that we are working with piano which has 27.5 (A0) to 4186 (C8) frequency interval. FFT breaks complex audio-frequencies to consistent signals.

```
fs, data = wavfile.read(os.path.join(PATH,AUDIO_FILE)) # Load the data
audio = data.T[0] # this is a two channel soundtrack, get the first track
FRAME_STEP = (fs / FPS) # audio samples per video frame
```

1. Audio file has two dimension while we reading wavfile in Python.
2. Firstly take the transpose two dimensional matrix for FFT library.
3. Then visualize the values returned from FFT library with FFMEG.



-> **FFT** Stores how many time coming same frequency basically

Constraints of Initial Work: This is **not** real time process.

Code Available At : <https://github.com/Umit-Yilmaz/GraphAuido>