Audio Signal

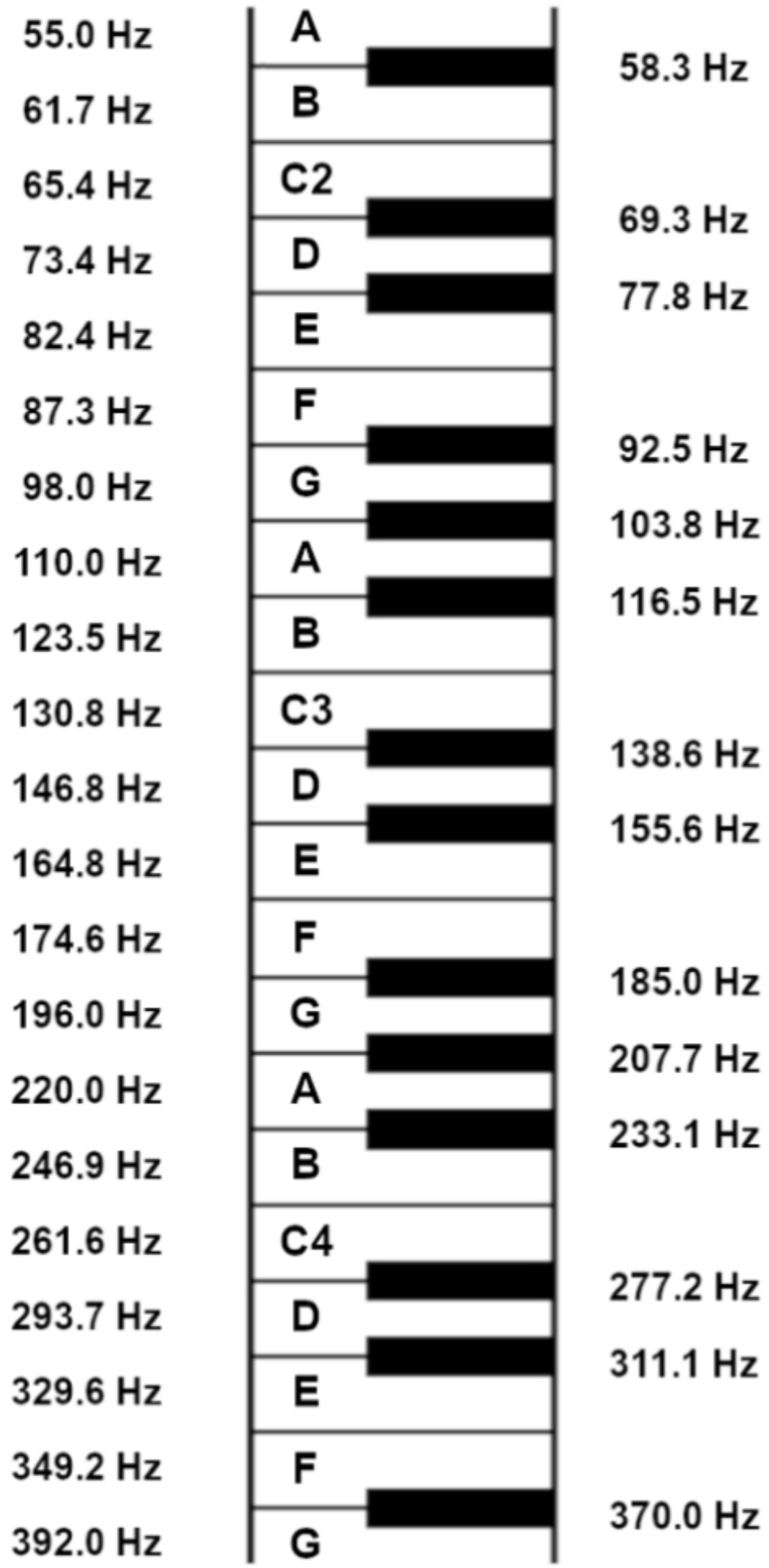
An introduction

Seanghay Yath

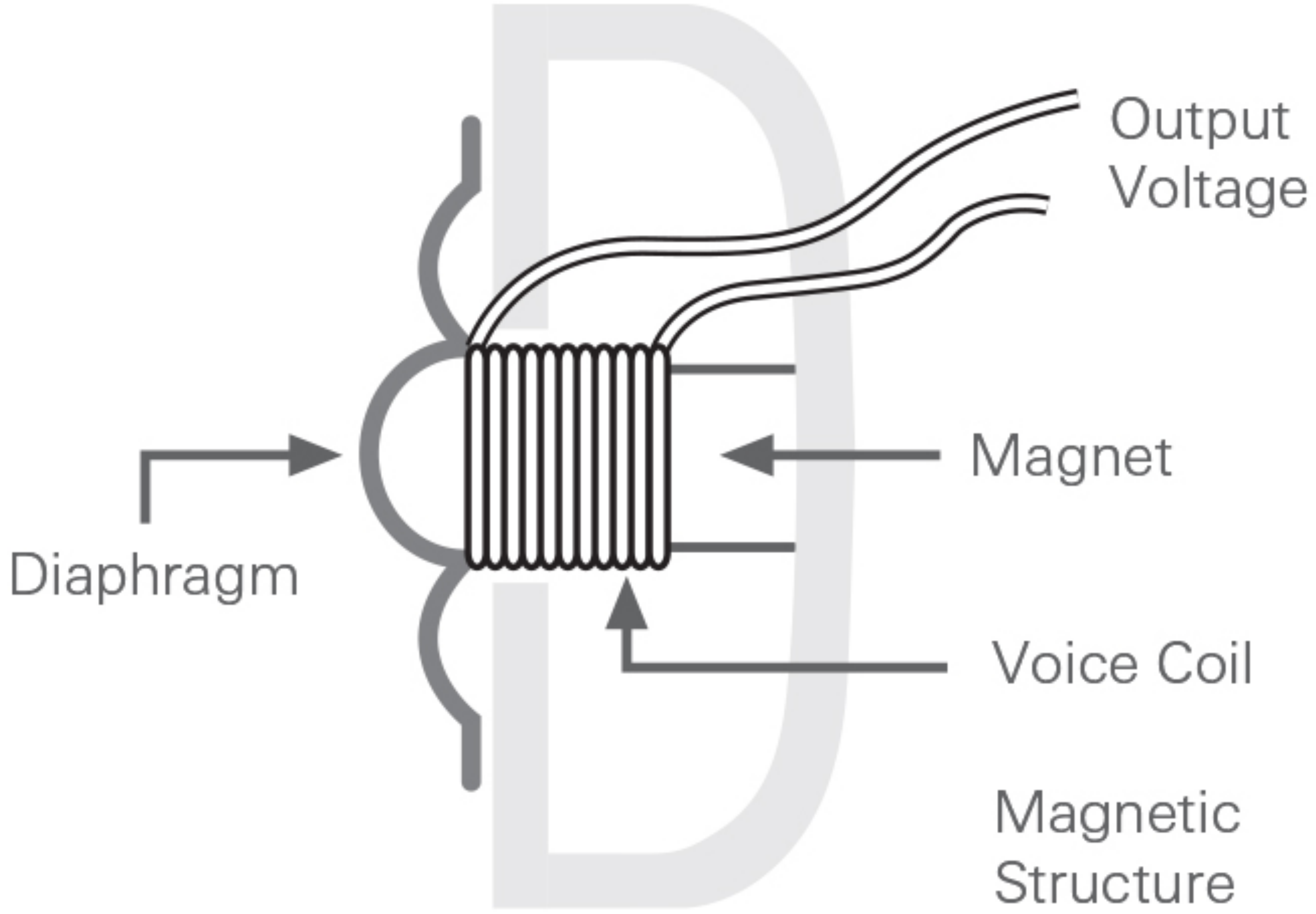
What is sound?

In physics, sound is a vibration that propagates as an acoustic wave through a transmission medium such as a gas, liquid or solid. Sound is measured in Hertz.

https://ciechanow.ski/sound



Digital Sound Recording

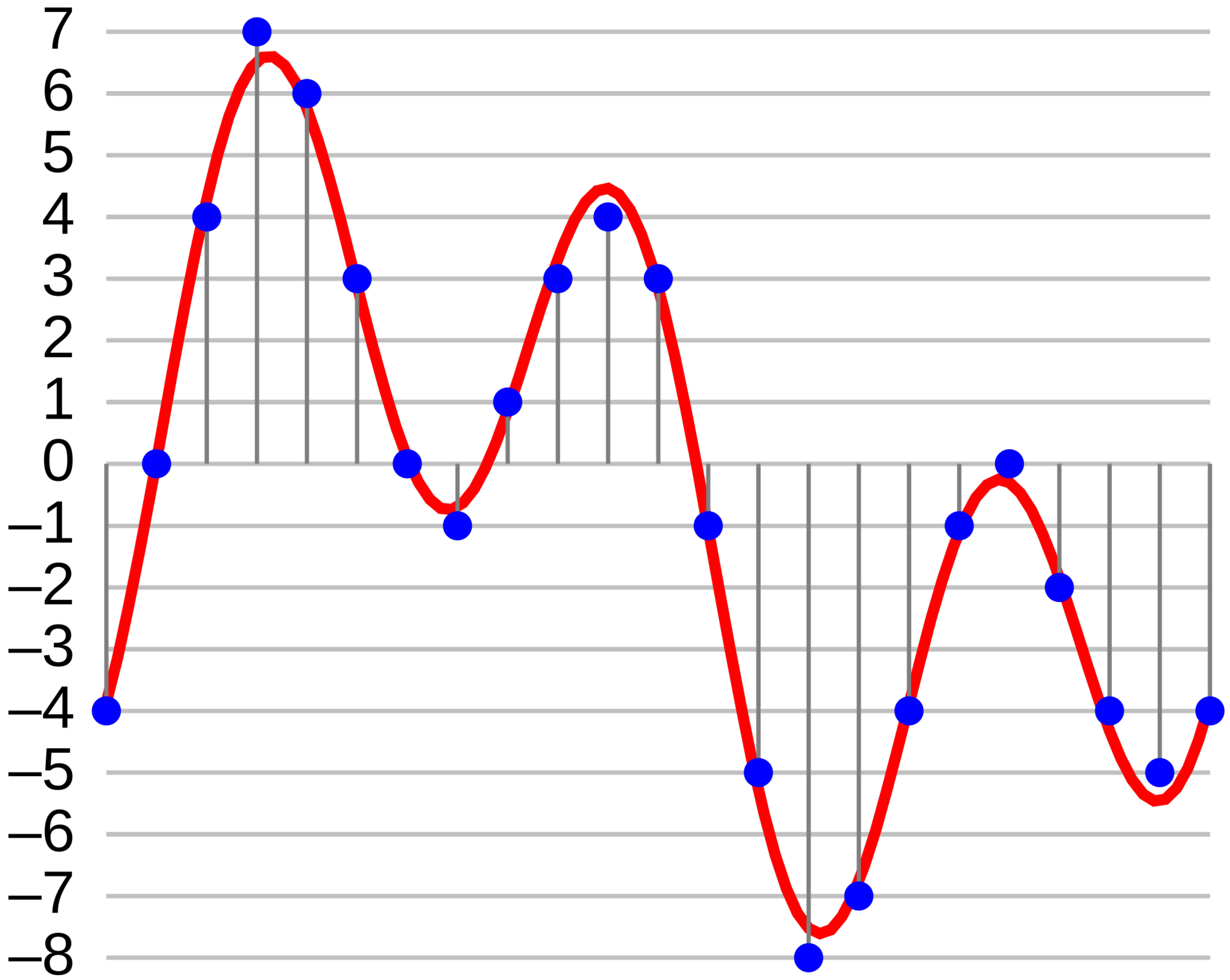


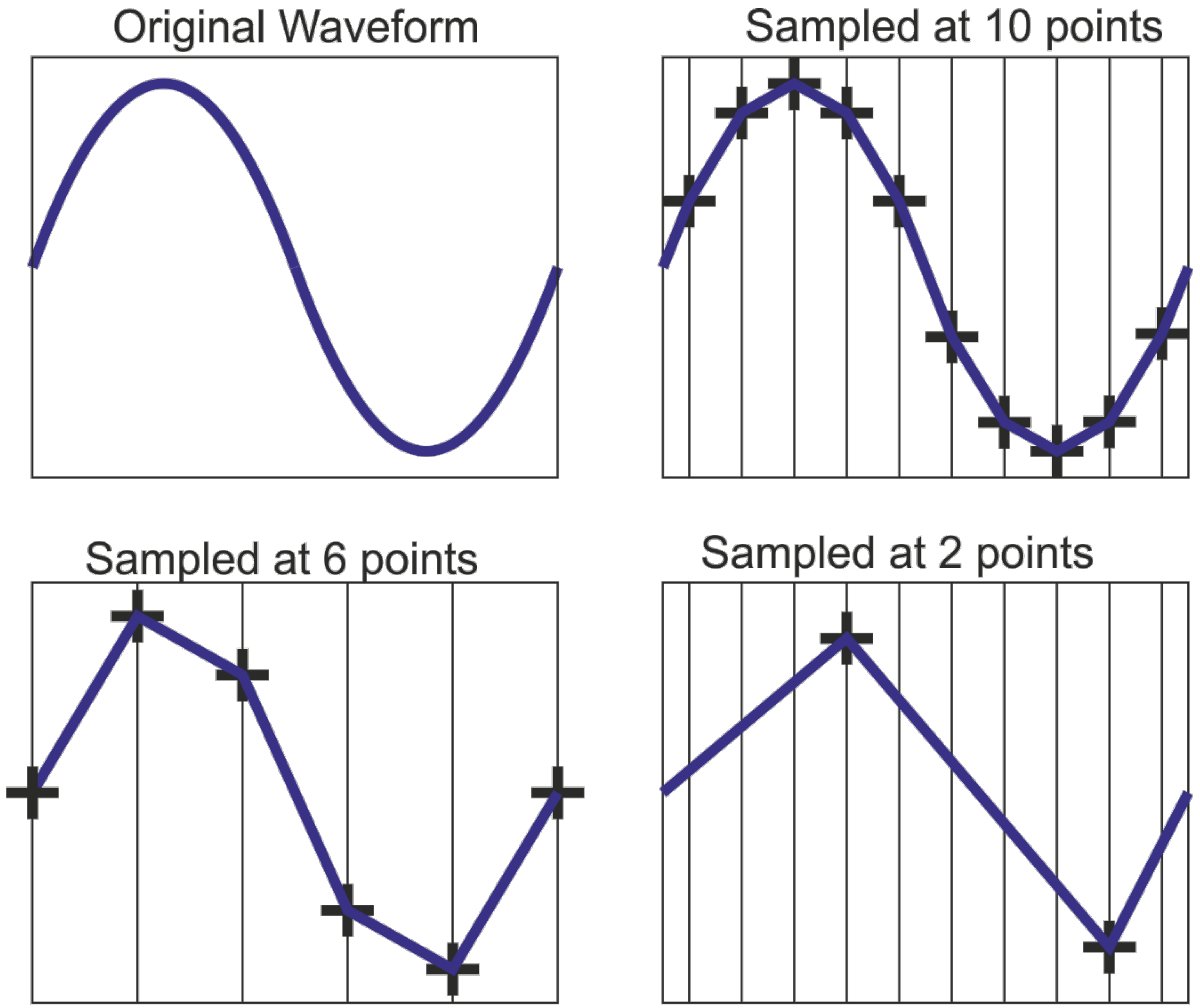
Output voltage from microphone is AC (Analog Signal)

It’s not possible to store analog signal digitally. So we need to get creative.

Digital Sound Recording

Digital Sound Recording

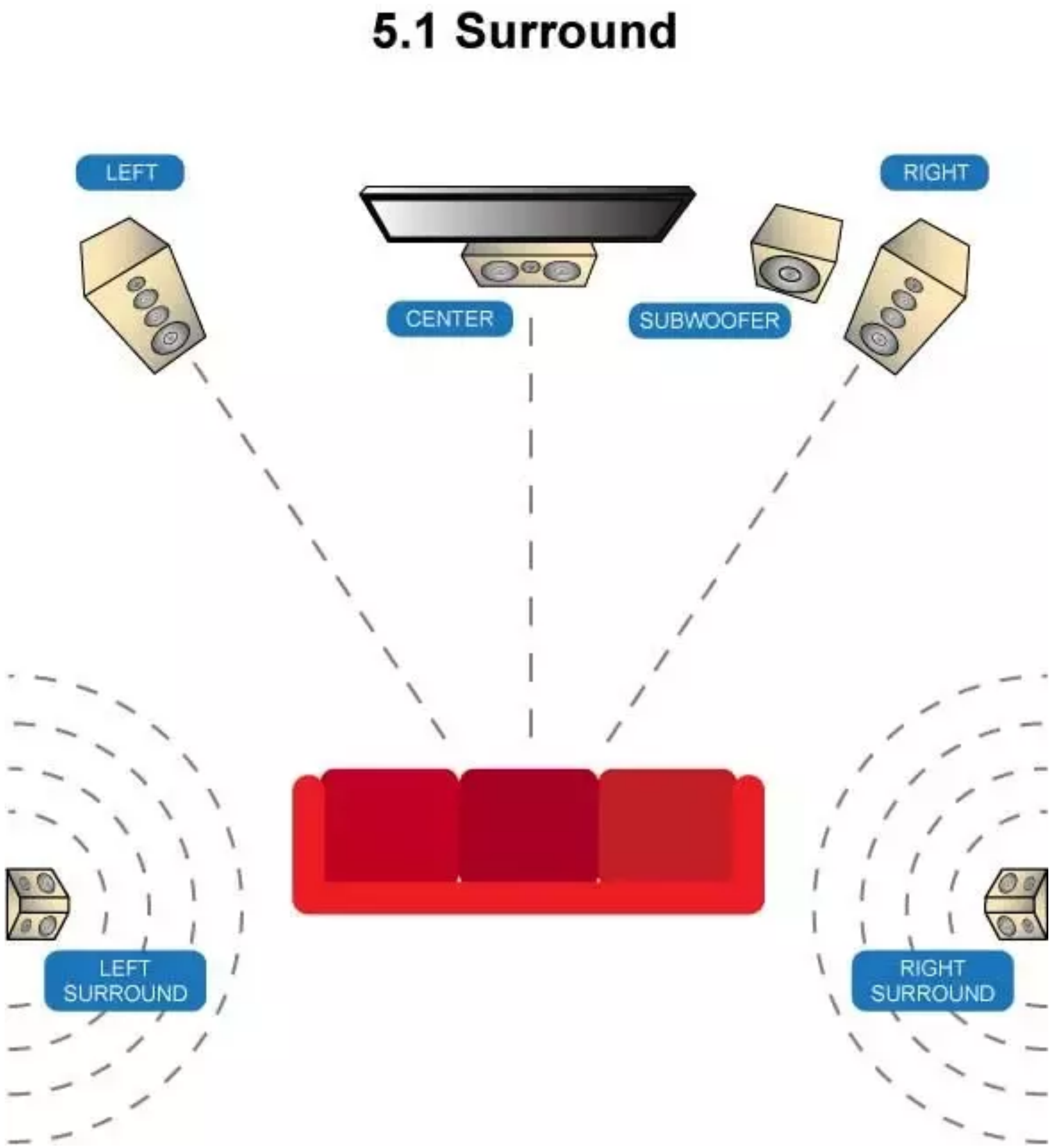




Audio Channels

1-Channel = Mono

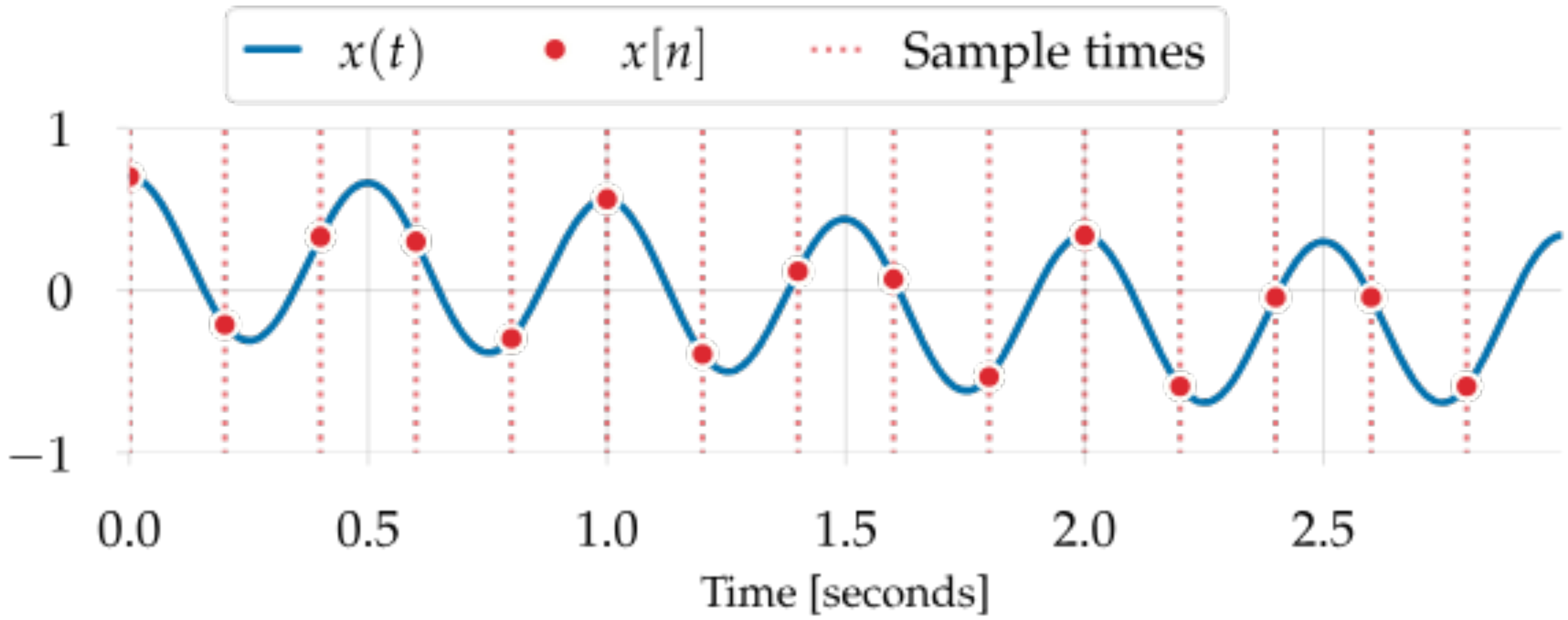
2-Channel = Stereo



Sample Rate

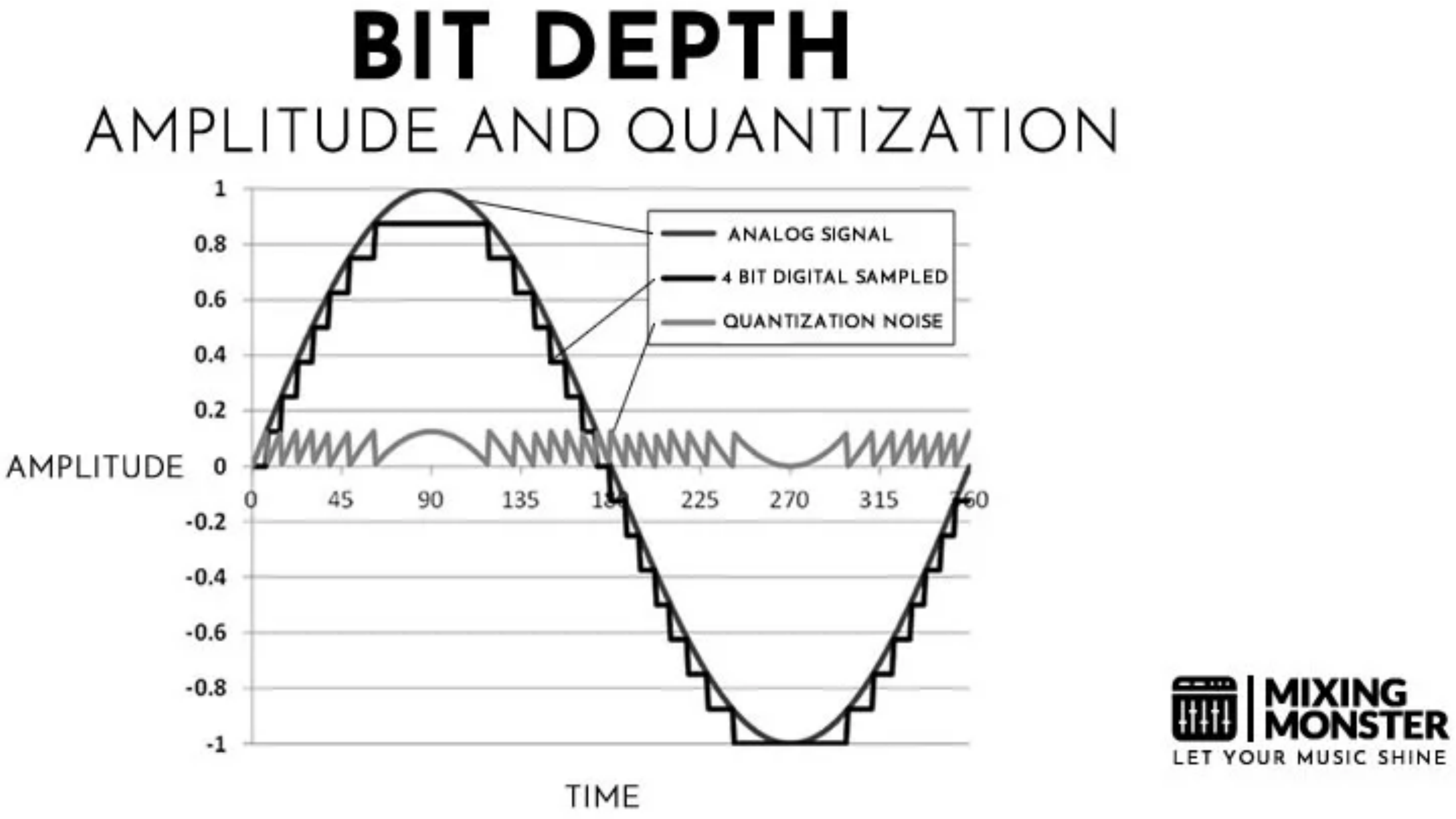
Sampling rate or sampling frequency defines the number of samples per second (or per other unit) taken from a continuous signal to make a discrete or digital signal. e.g. 16kHz, 22.05kHz, 44.1kHz 48kHz

```
Audio Duration In Seconds = Number of Samples / Sample Rate
```



Bit Depth / Sample Width

Bit depth is the number of bits of information in each sample, and it directly corresponds to the resolution of each sample. e.g. 32-bit 16-bit, 8-bit,



Commonly used

32-Bit float32

16-Bit int16

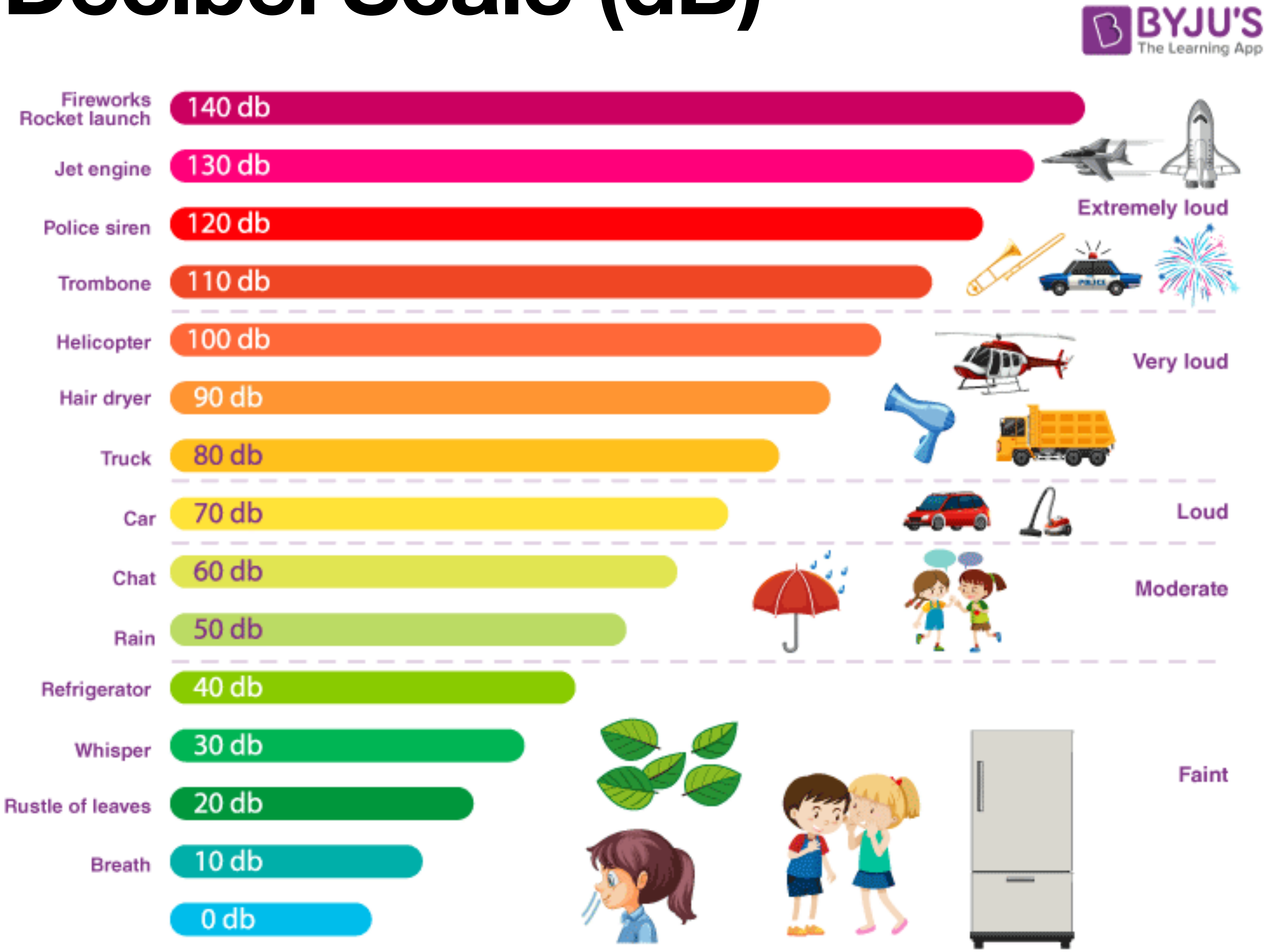
How does it look like?

```
audio = [
  [0.3, 0.1, ..., 0.1, .8, .75],
  [0.45, 0.2, ..., 0.3, .23, .1],
]
```

Pulse-code modulation (PCM)

Pulse-code modulation is a method used to digitally represent analog signals. It is the standard form of digital audio in computers, compact discs, digital telephony and other digital audio applications.

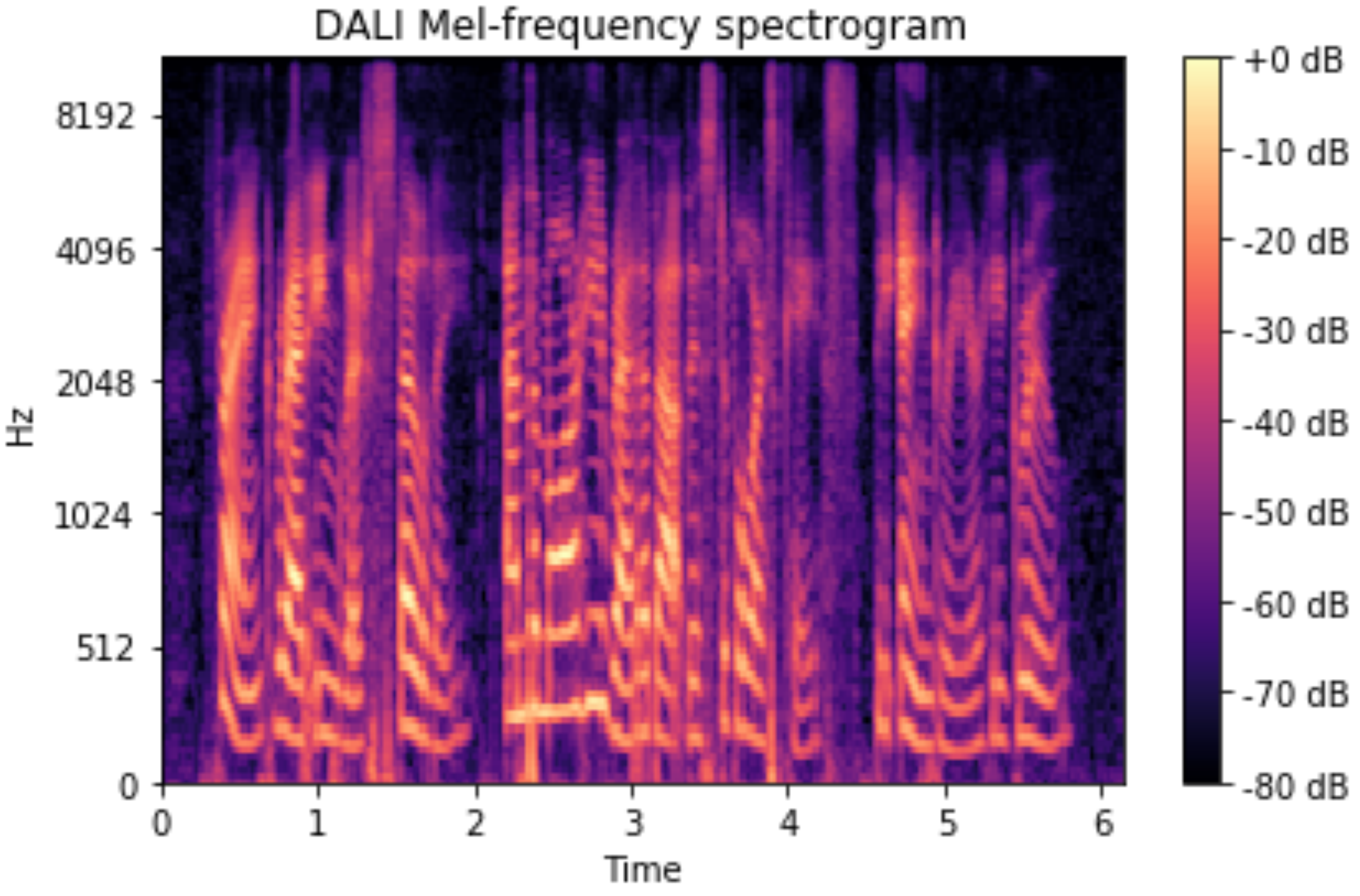
Decibel Scale (dB)



Mel-scale (Melody Scale)

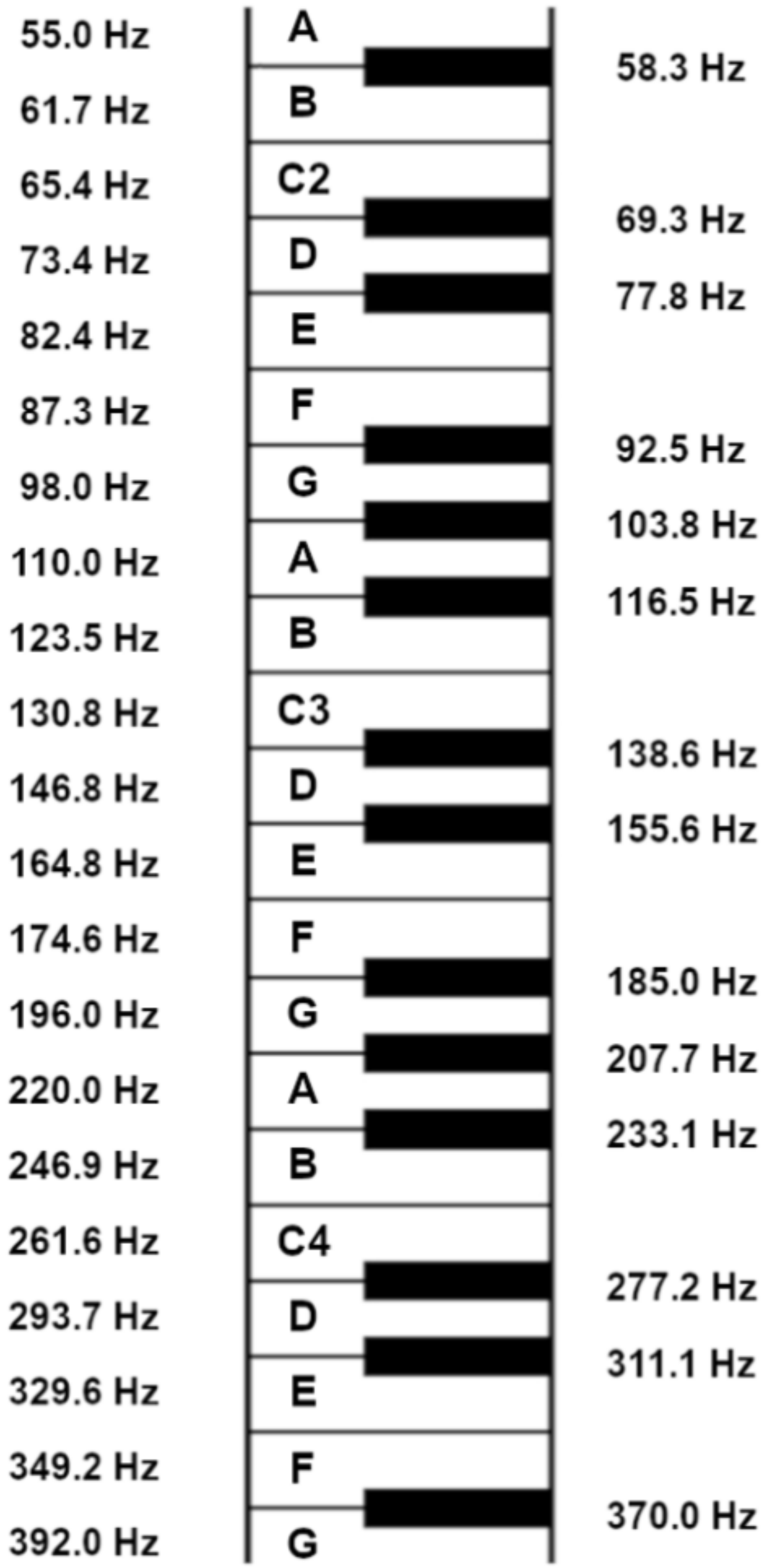
$$m = 1127 \cdot \log\left(1 + \frac{f}{700}\right)$$

The mel scale is a (human) perceptual scale of pitches judged by listeners to be equal in distance from one another. The reference point between this scale and normal frequency measurement is defined by assigning a perceptual pitch of 1000 mels to a 1000 Hz tone, 40 dB above the listener's threshold.



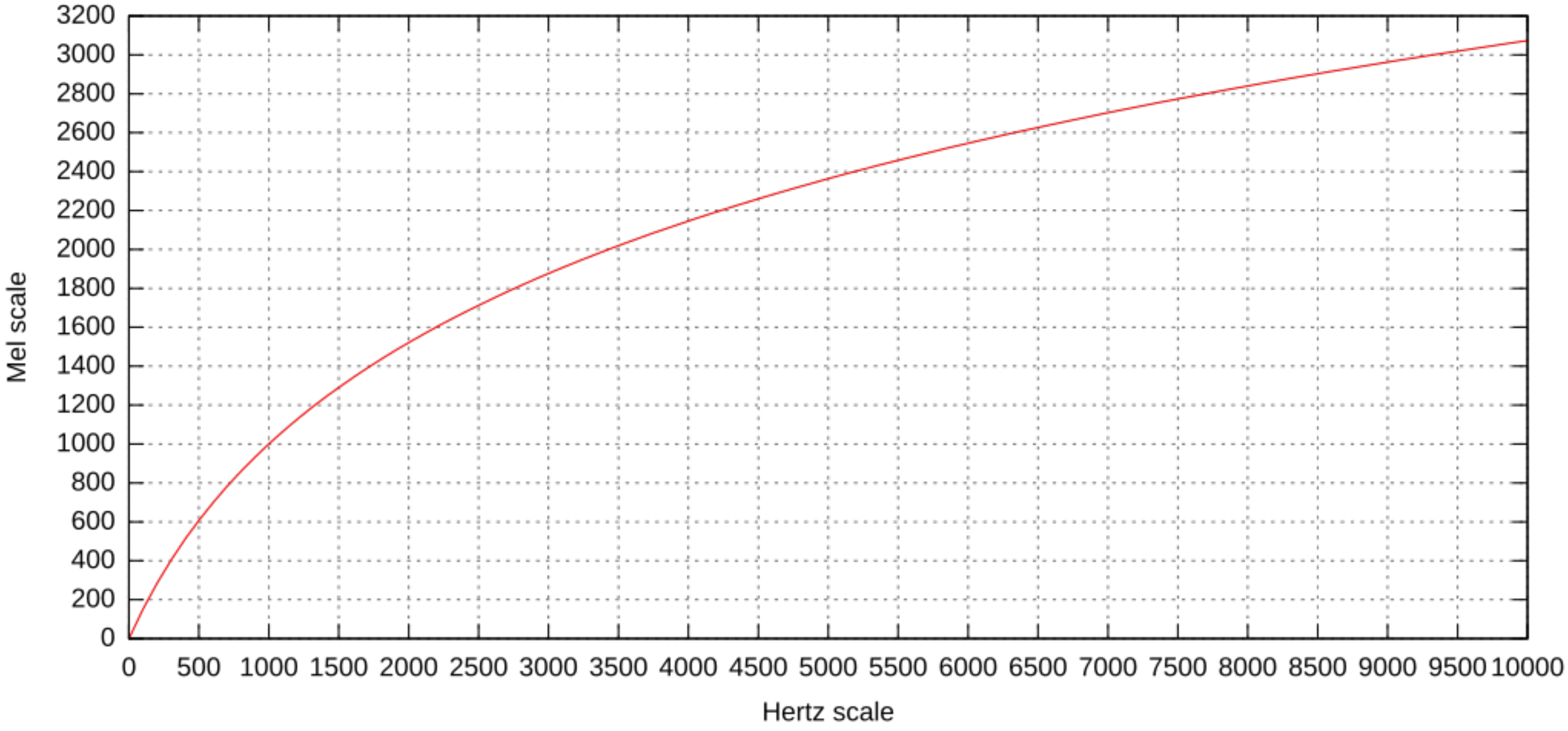
Mel-scale (Melody Scale)

The mel scale is a (human) perceptual scale
of pitches judged by listeners to be equal in distance from one another. The reference point between this scale and normal frequency measurement is defined by assigning a perceptual pitch of 1000 mels to a 1000 Hz tone, 40 dB above the listener's threshold.

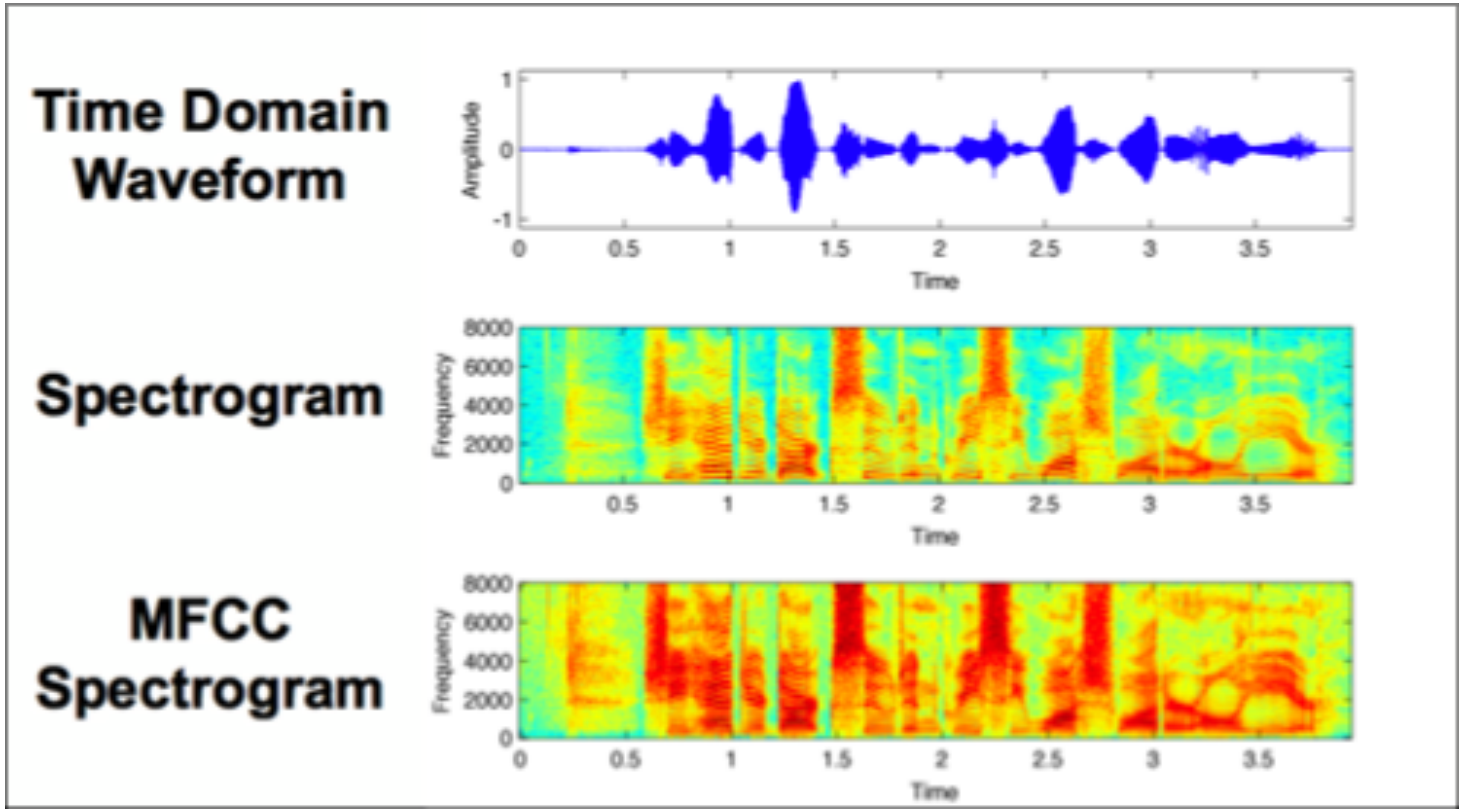


Mel-scale (Melody Scale)

$$m = 1127 \cdot \log\left(1 + \frac{f}{700}\right)$$



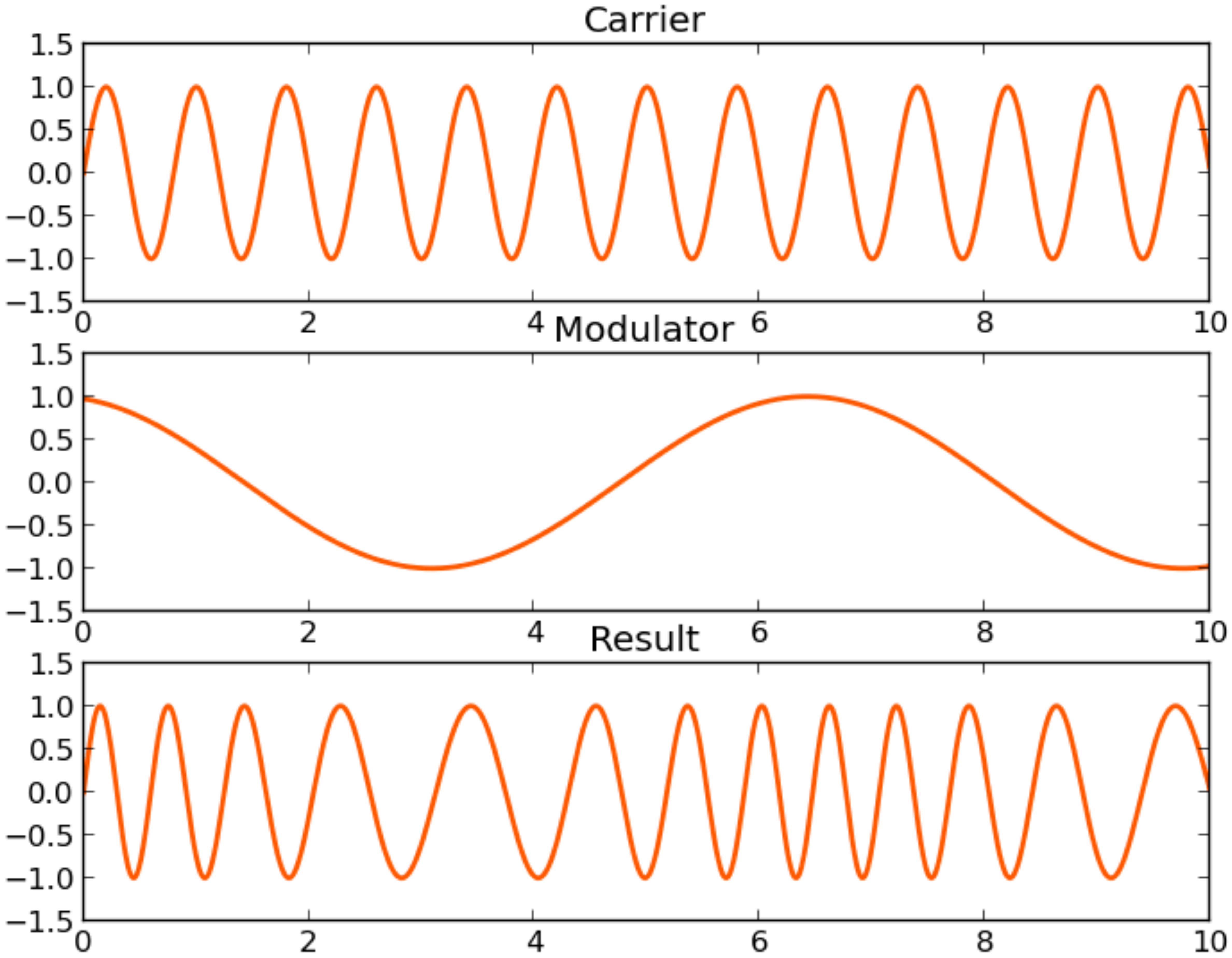
Mel-frequency cepstral coefficients (MFCCs)



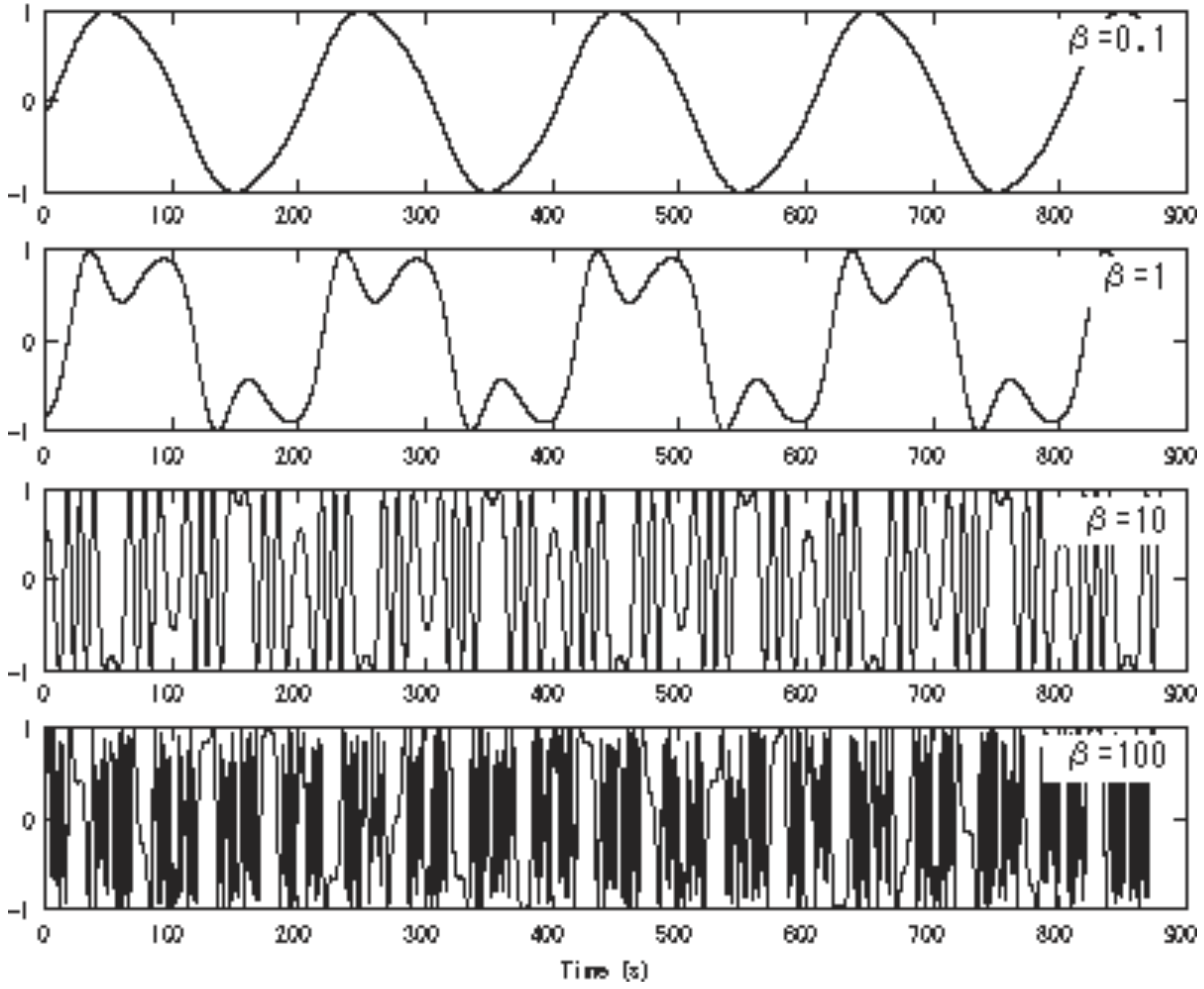
Mel-spectrograms are often used as inputs to deep learning models, which can learn relevant features directly. MFCCs are typically used in more traditional machine learning models where dimensionality reduction is beneficial.

https://maelfabien.github.io/machinelearning/Speech9/

Modulation

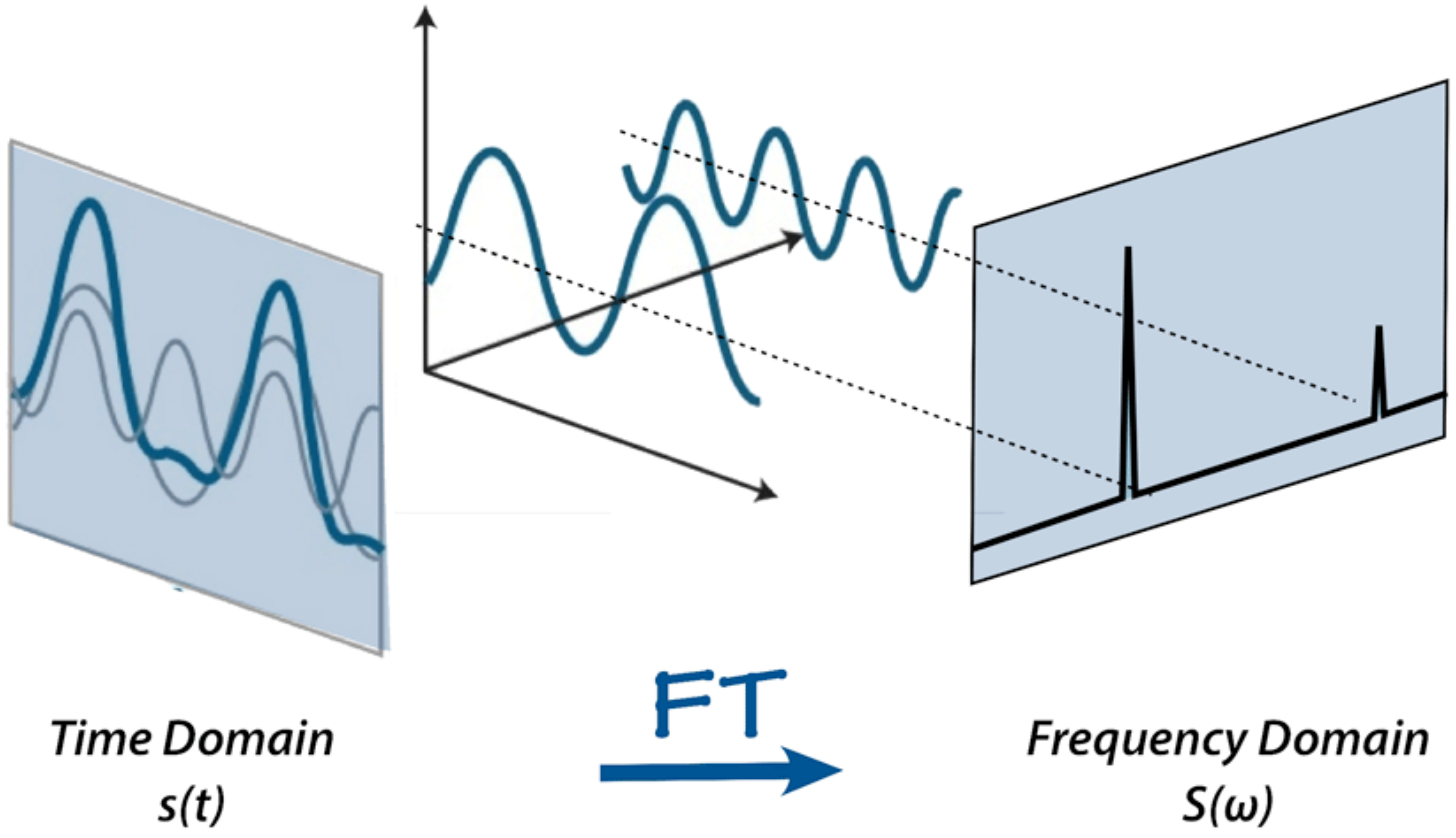


Carrier
Modulator
Result
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0
2
4
6
8
10



β =0.1
β =1
β =10
β =100
1
0
−1
0
100
200
300
400
500
600
700
800
900
Time [s]

Fourier Transform (FFT, STFT, InverseSTFT)

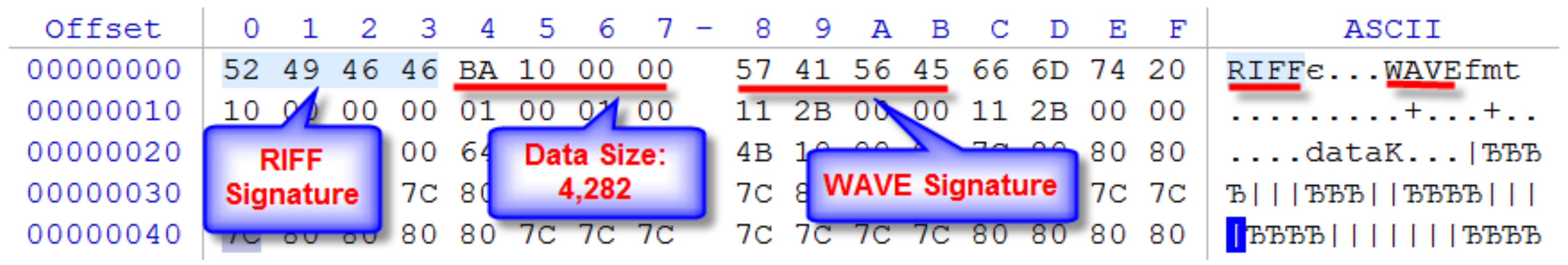


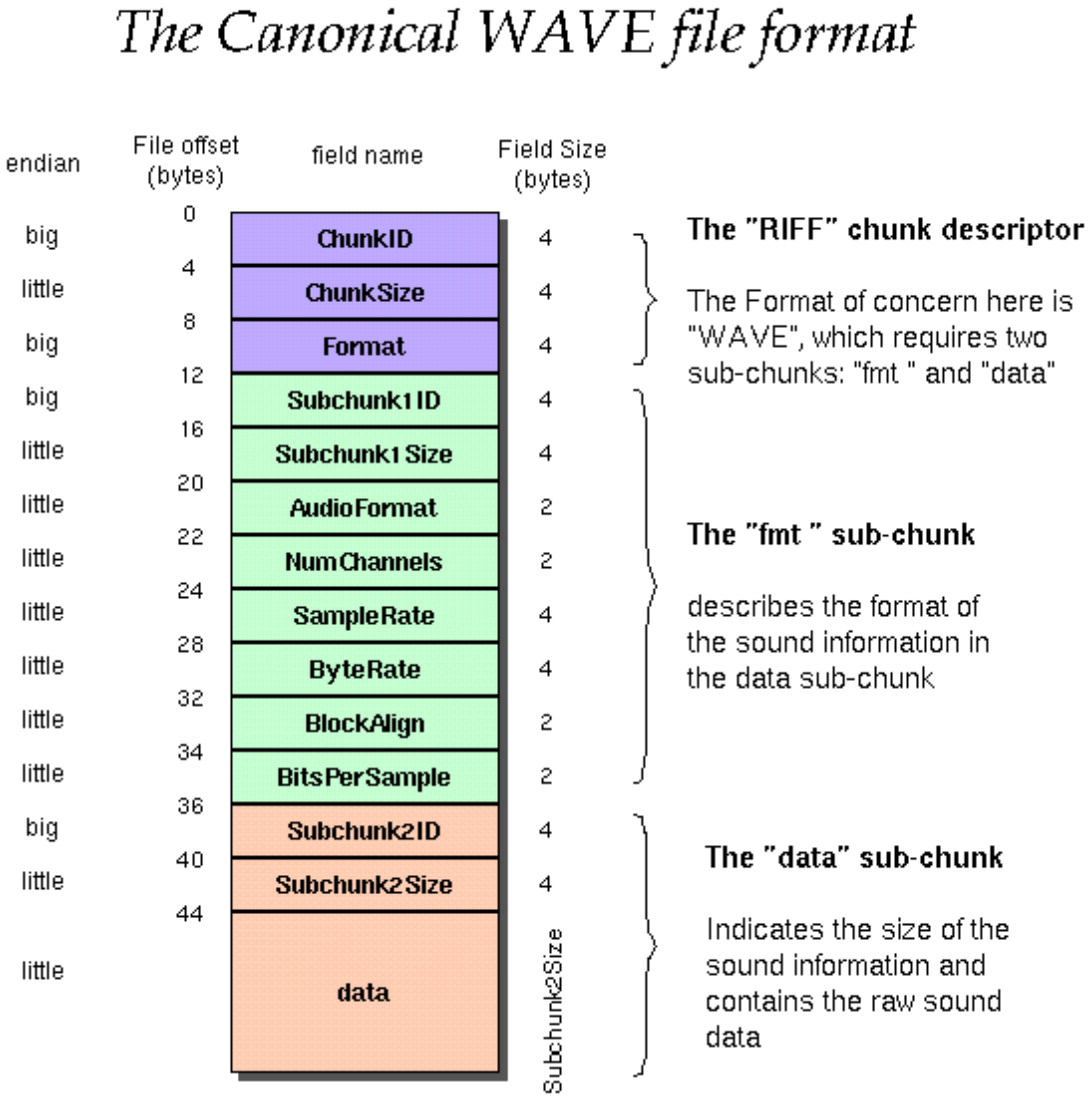
https://www.youtube.com/watch?v=spUNpyF58BY

Sound Storage

Choose the Right!

Waveform Audio File Format (WAV, WAVE)





FLAC - Free Lossless Audio Codec

FLAC is an audio coding format for lossless compression of digital audio, developed by the Xiph.Org Foundation.

FLAC compresses files at a ratio of approximately 2:1, reducing the size of the original by **50%**.

MP3 - MPEG-1/2 Audio Layer III

Audio coding format for digital audio that offers **Lossy Compression.**
Think of it as a zip file.

Requires LAME MP3 Audio Encoder / Decoder

Bit Rate is the amount of data processed per second in an audio file, measured in kilobits per second (kbps). It directly affects the file size and the audio quality. Higher is better. Higher results in larger file size.

Advanced Audio Coding (AAC)

an audio coding standard for lossy digital audio compression.

OGG

Ogg Vorbis is a fully open, non-proprietary, patent-and-royalty-free, general-purpose compressed audio format for mid to high quality (8kHz-48.0kHz, 16+ bit, polyphonic) audio and music at fixed and variable bitrates from 16 to 128 kbps/channel. This places Vorbis in the same competitive class as audio representations such as MPEG-4 (AAC), and similar to, but higher performance than MPEG-1/2 audio layer 3, MPEG-4 audio (TwinVQ), WMA and PAC.

Audio Library for ML

- librosa
- torchaudio
- audiomentations
- scipy
- numpy
- pydub
- sox
- soundfile
- pedalboard

Audio Signal Processing Tools

- FFmpeg
- Audacity
- Praat

https://github.com/Yuan-ManX/audio-development-tools

End.