CSC3100 - Fundamentals of Speech and Language Processing

Lecture 2: Understanding sound and acoustics

Zhizheng Wu

Agenda

- Sound and its journey
- Digital sound wave
- Time domain vs frequency domain
- Quantifying sound
  - Physical property
  - Perceptual property

Sound

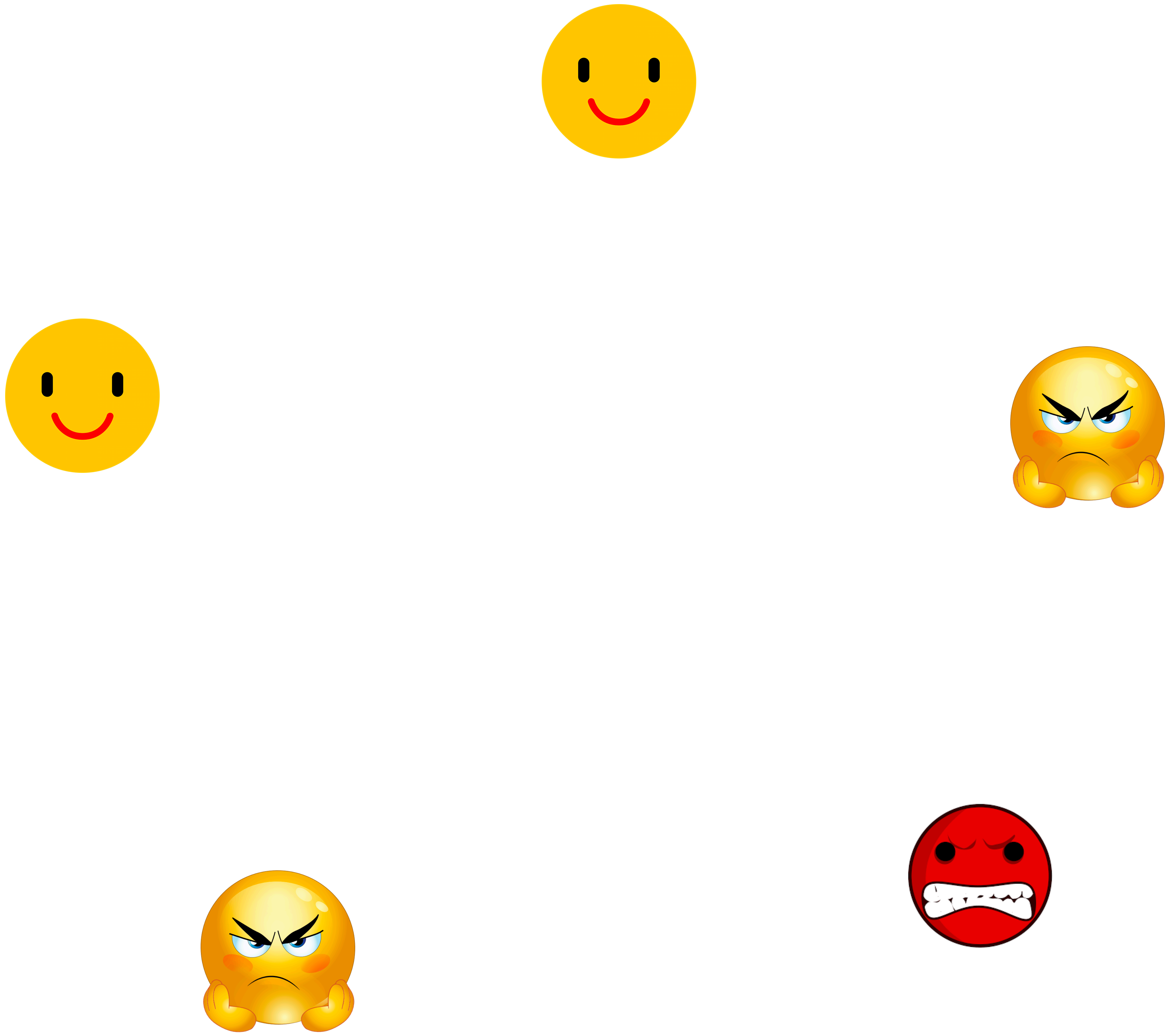
Sound of nature

- The sound of wading through shallow water, picking your feet out and putting them back in. Maybe on a beach or in wetland.

- Gull wheeling overhead on beach with wave sound in background

-

Sound of human



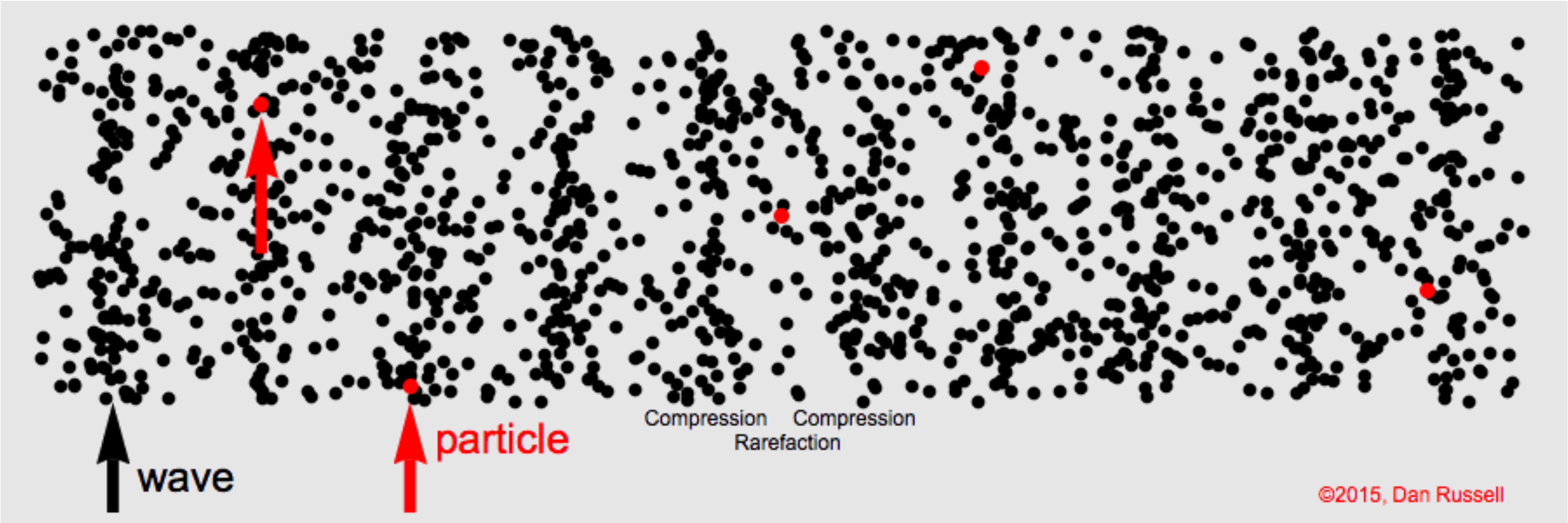
Sound

- Physical definition
  - A **vibration** that propagates as an **acoustic wave**, through a transmission medium such as a gas, liquid or solid.

- Psychophysical definition
  - ***Reception*** of such acoustic waves and their ***perception*** by the brain.

Waves

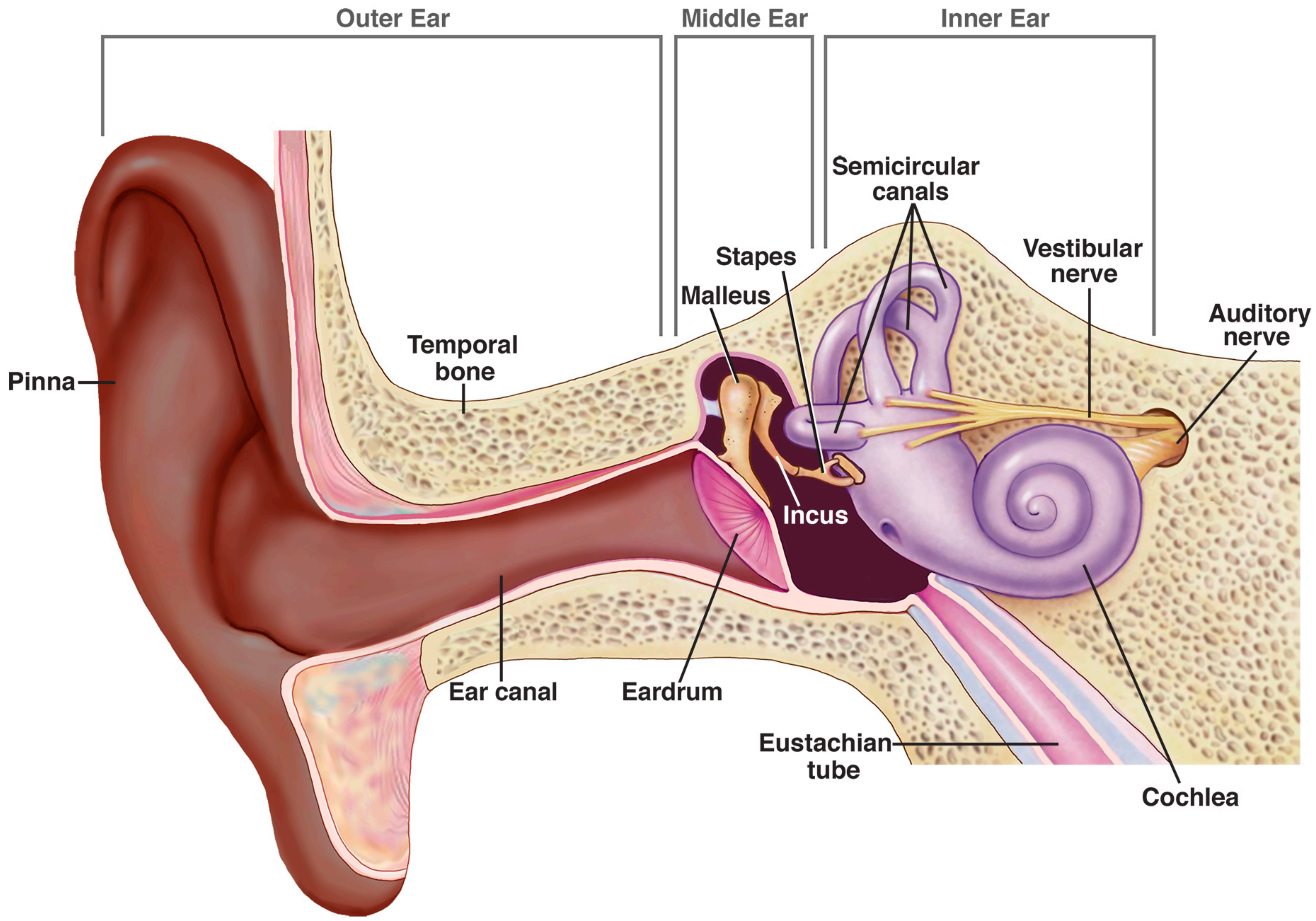
- Sound is transmitted through gases, plasma, and liquids as longitudinal waves, also called compression waves.



https://www.acs.psu.edu/drussell/Demos/waves/wavemotion.html

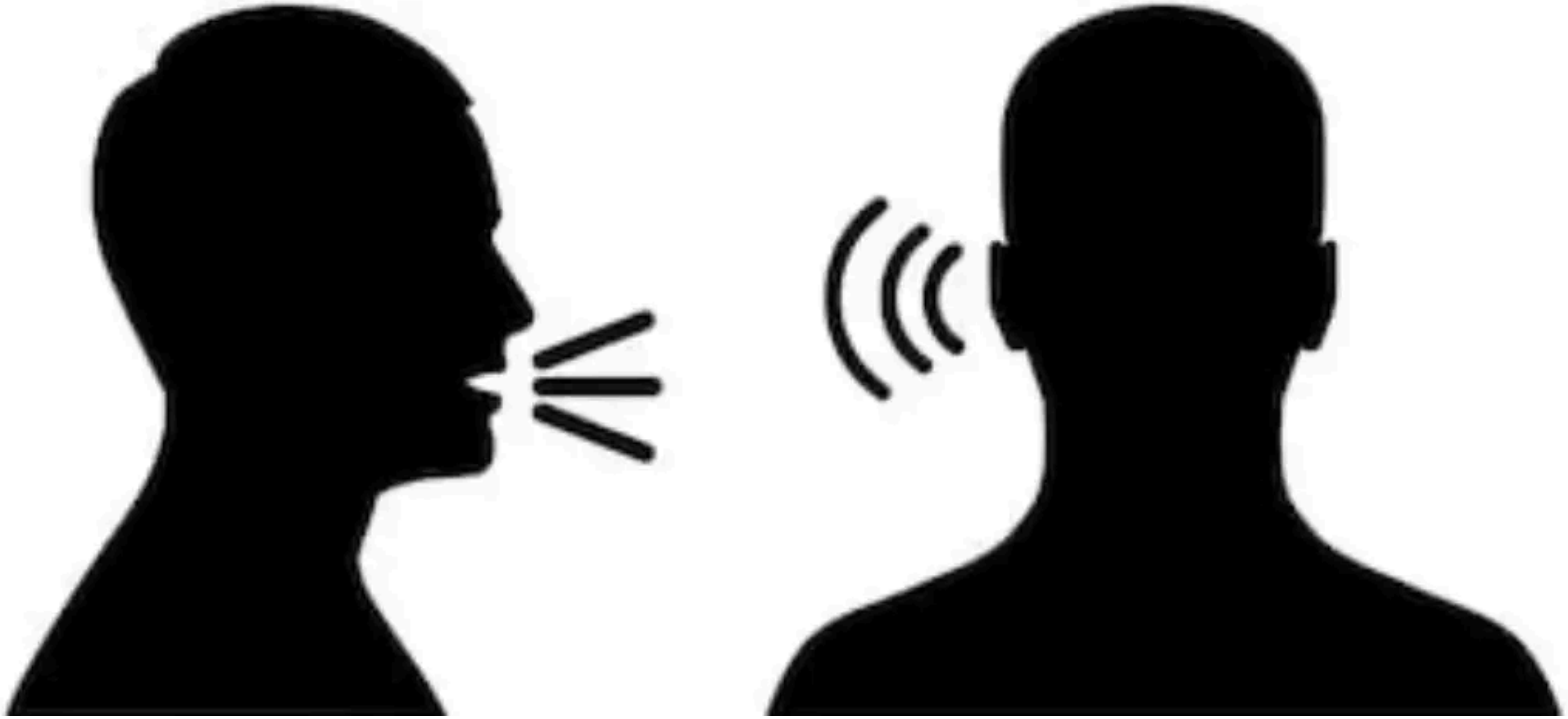
Human ear

- human hearing range: ~20 – 20,000 Hz



https://www.nidcd.nih.gov/health/how-do-we-hear

Speaking vs Listening



Journey of sound to the brain

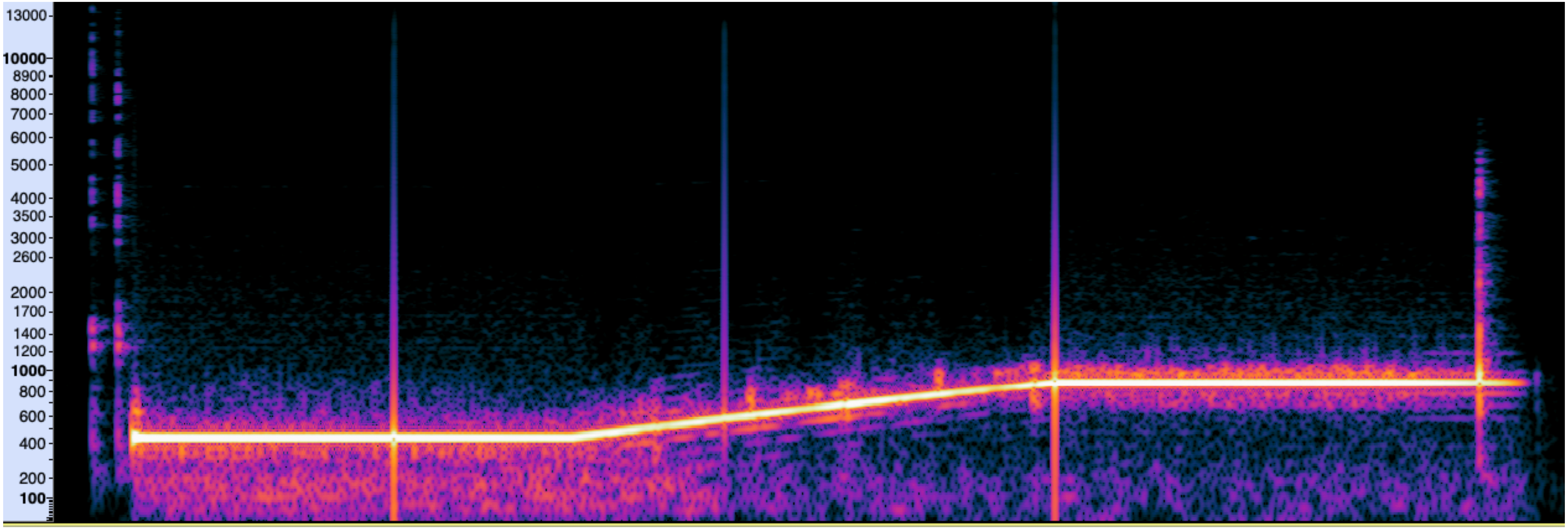
https://www.nidcd.nih.gov/news/multimedia/journey-of-sound-video


What do you hear!?!
YANNY
LAUREL
VOTE

Quantifying sound

- Perceptual characteristics
  - Loudness
  - Pitch
  - Timbre (tone color)
- Physical characteristics
  - Intensity
  - Frequency
  - Time variation and harmonic spectrum

Frequency and pitch



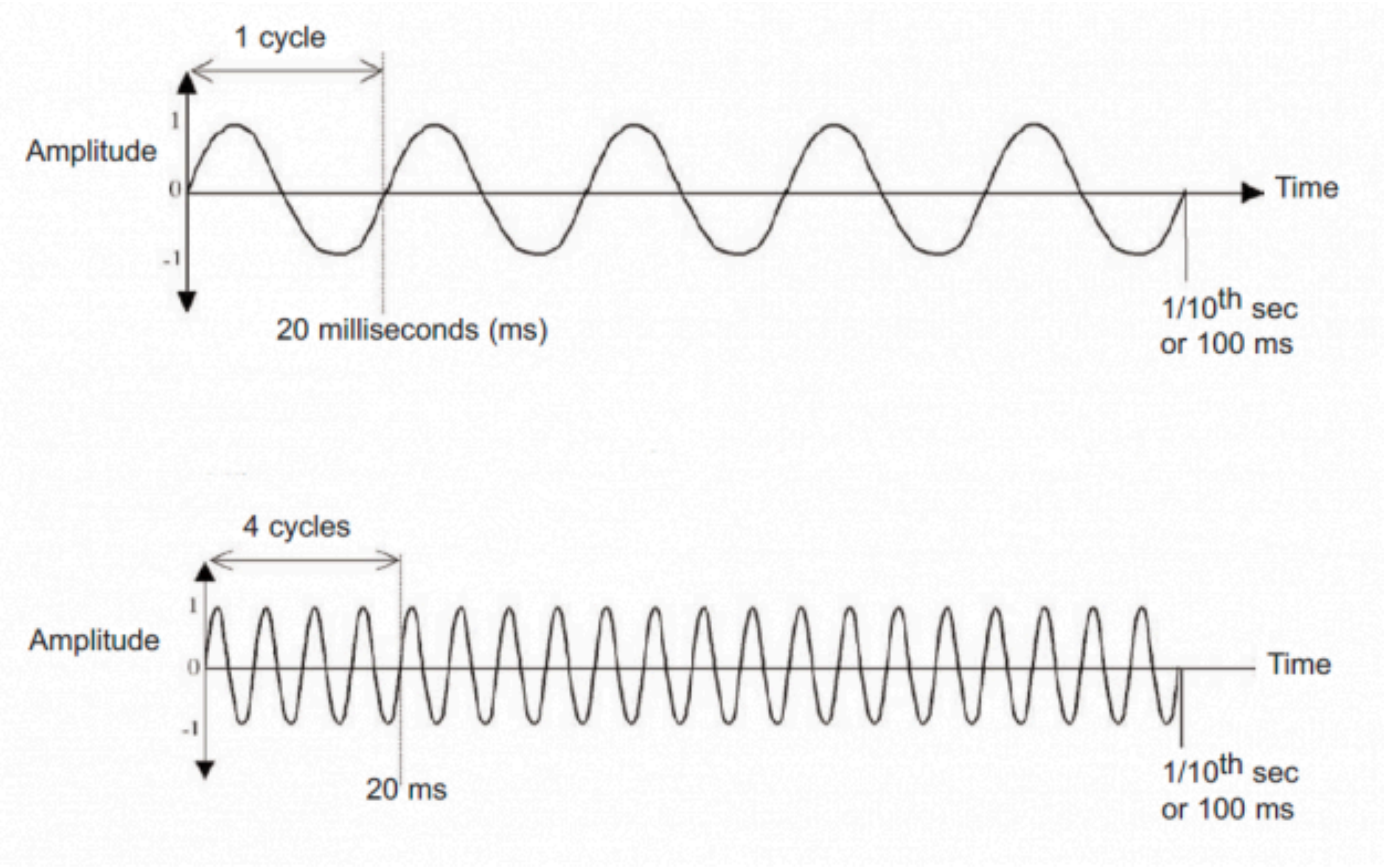
Pitch

- Pitch: Perceptual property
  - Low pitch <=> low frequency of vibration/oscillation
  - High pitch <=> high frequency of vibration/oscillation

Frequency

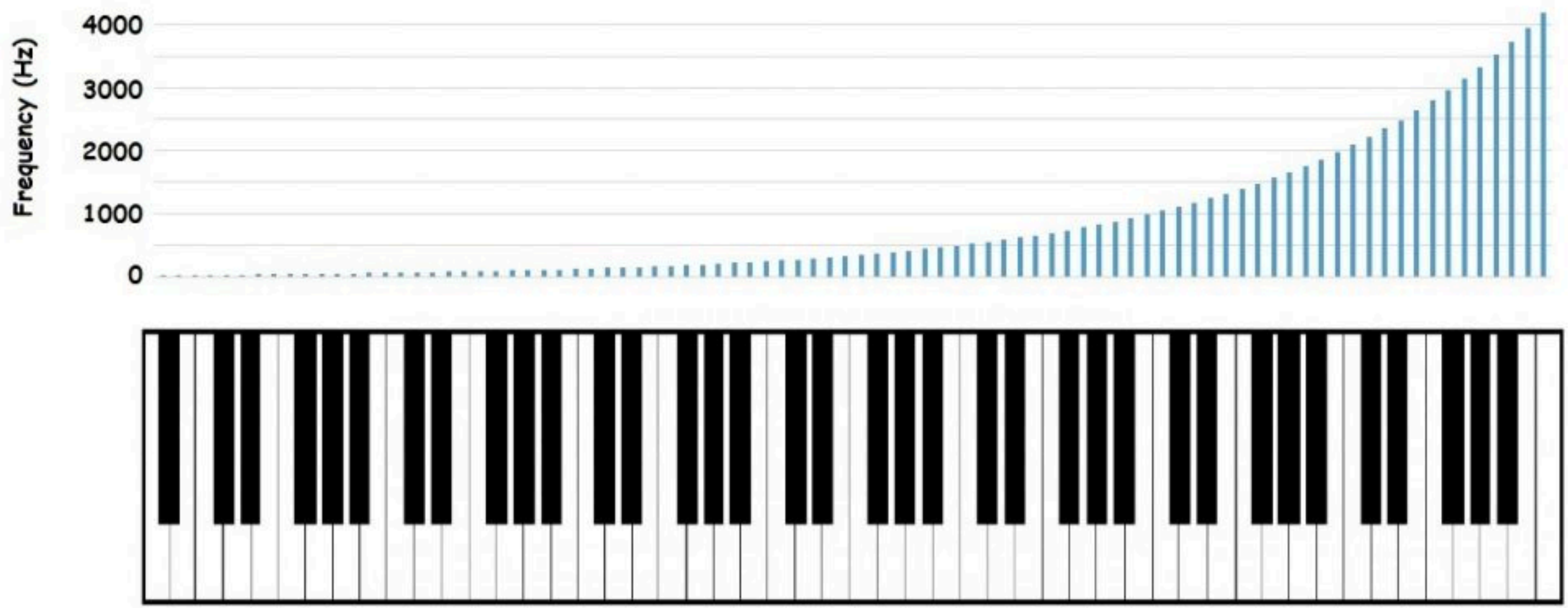
- Frequency: Physical property
  - An expression of how frequently a periodic wave form or signal repeats itself at a given amplitude

$$f = \frac{1}{T}$$



Frequency

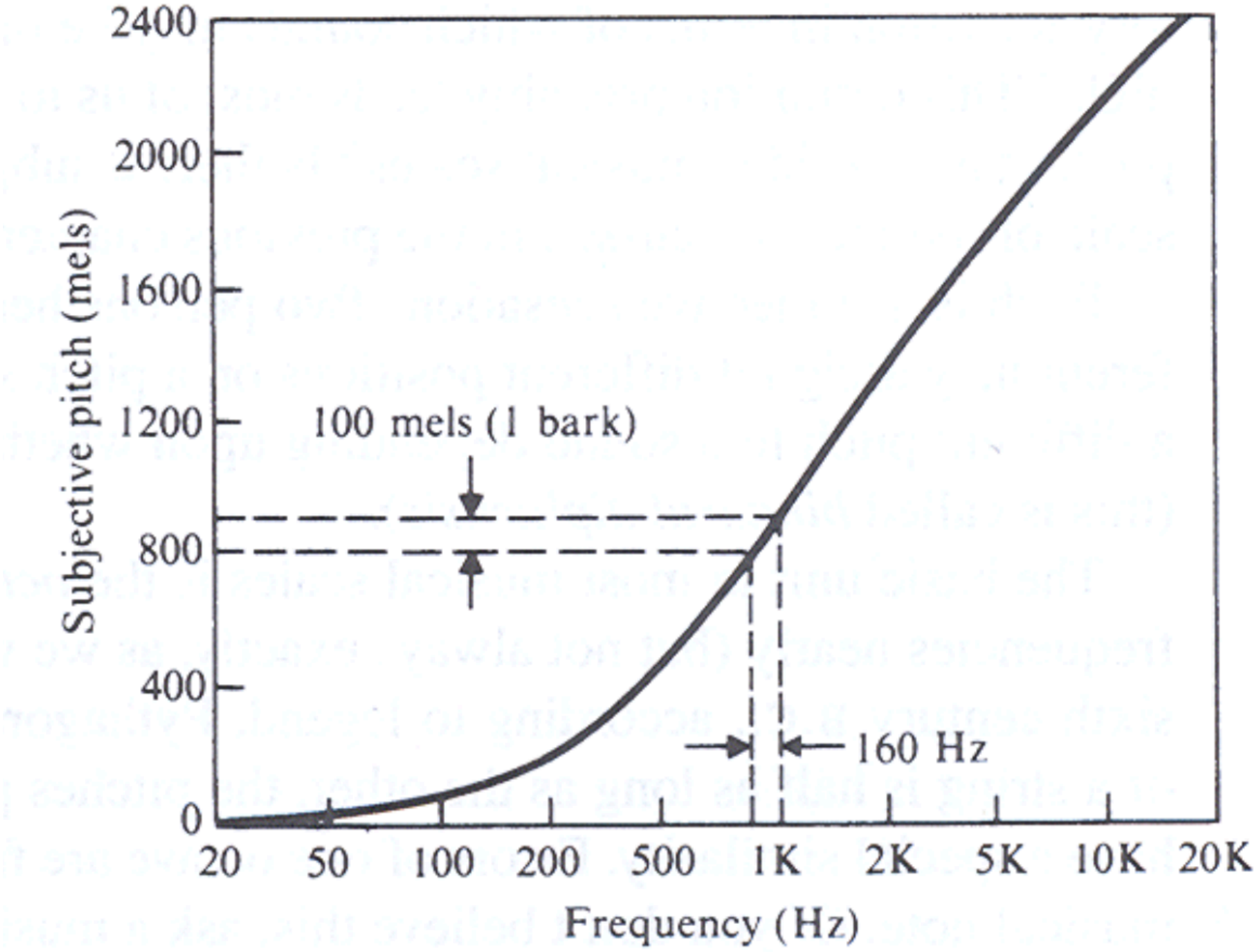
A0 is equal to 27.5 Hz while A4 is equal to 440 Hz. Notes start from A0 and go to C8 from left to right.



27.5 Hz 4,186 Hz

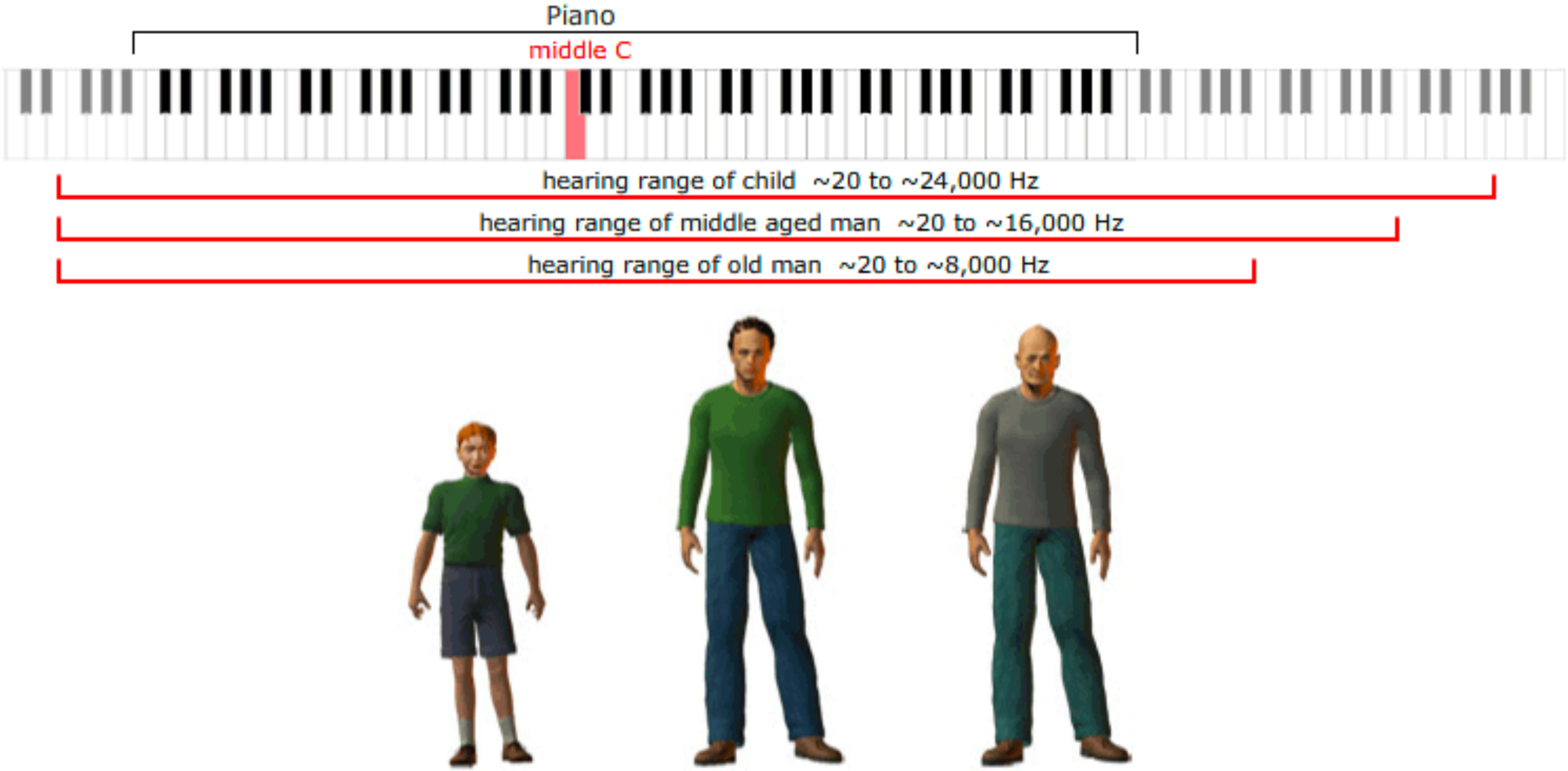
Frequency and pitch

- Pitch depends primarily (approximately) logarithmically on frequency

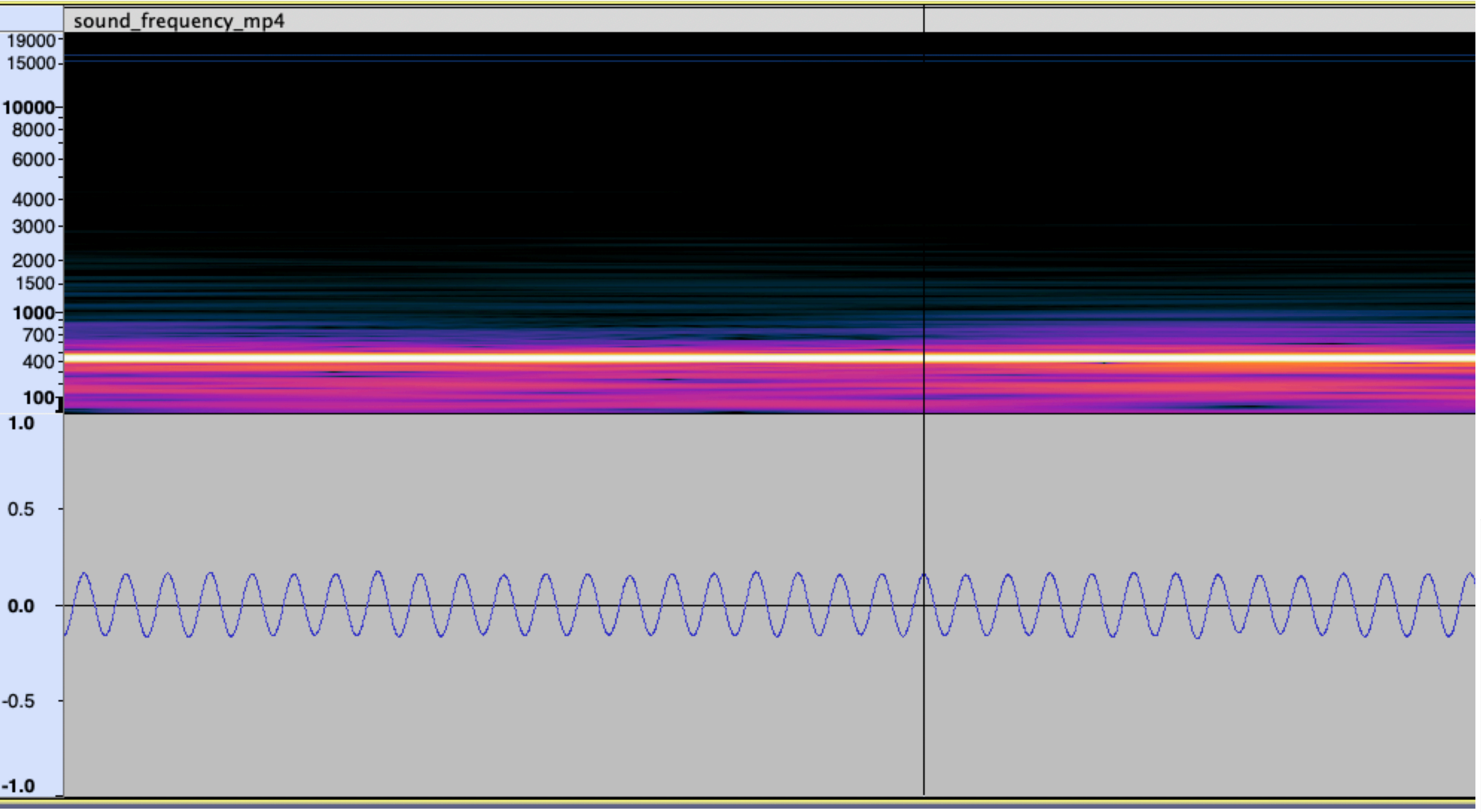


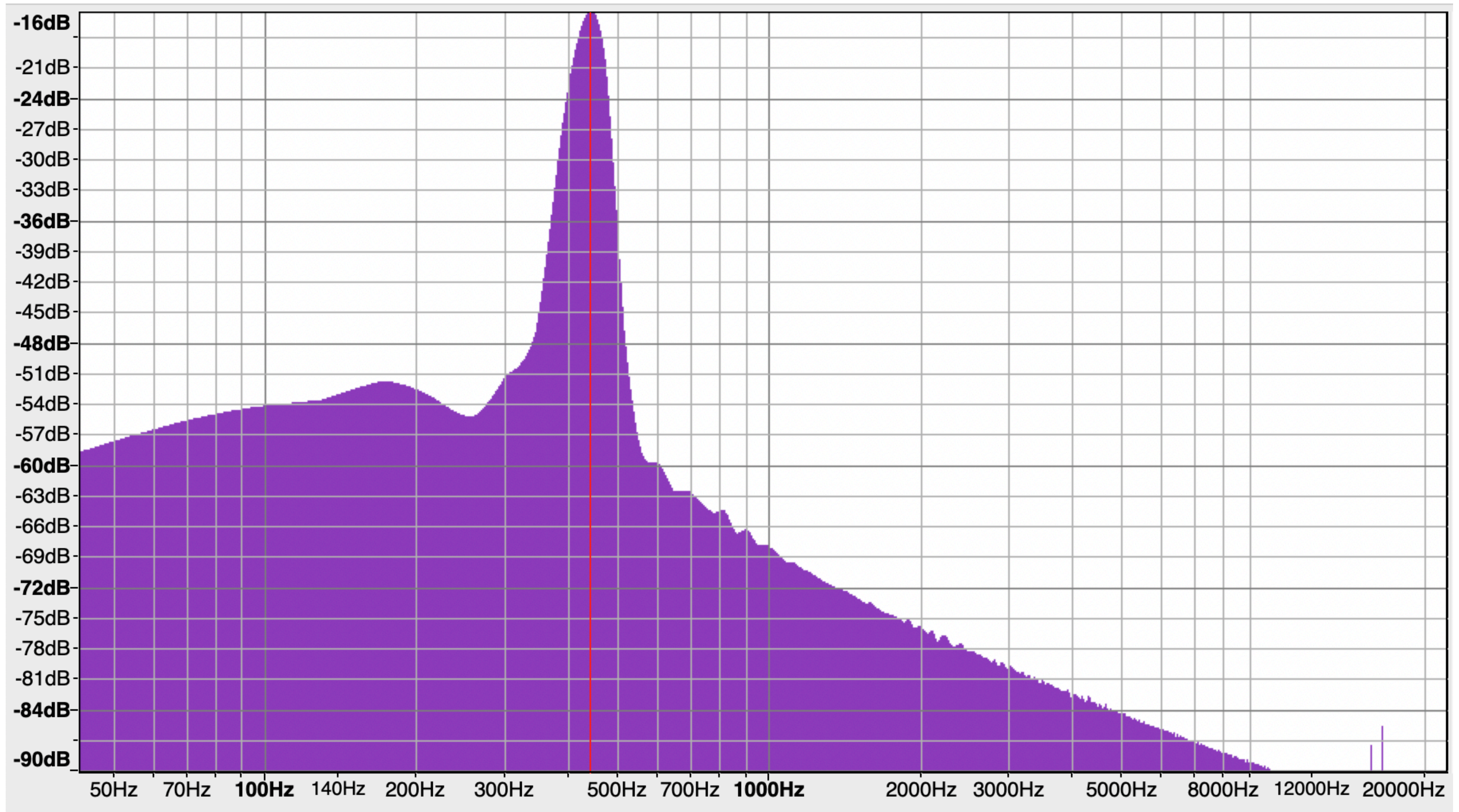
Human hearing range

- Human can hear frequencies above about 20 Hz



Frequency and pitch

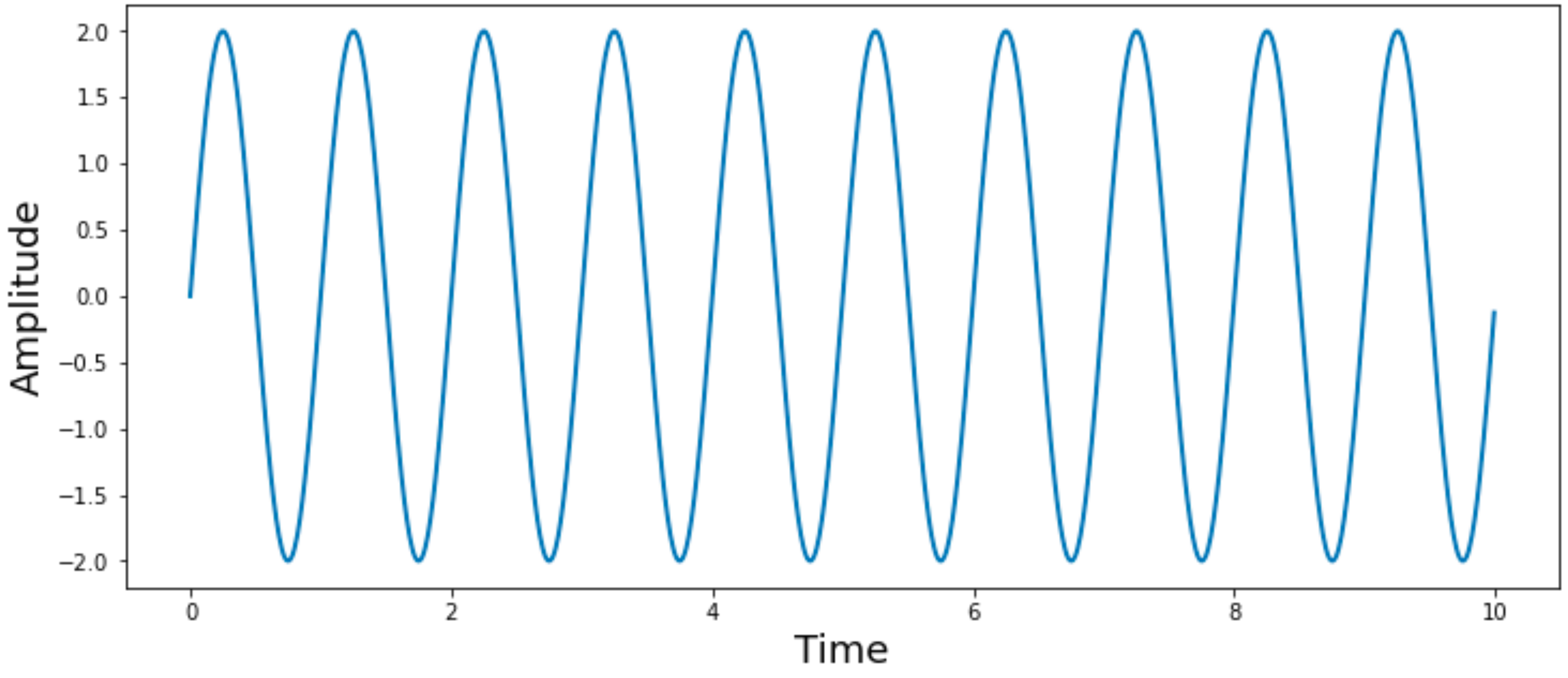


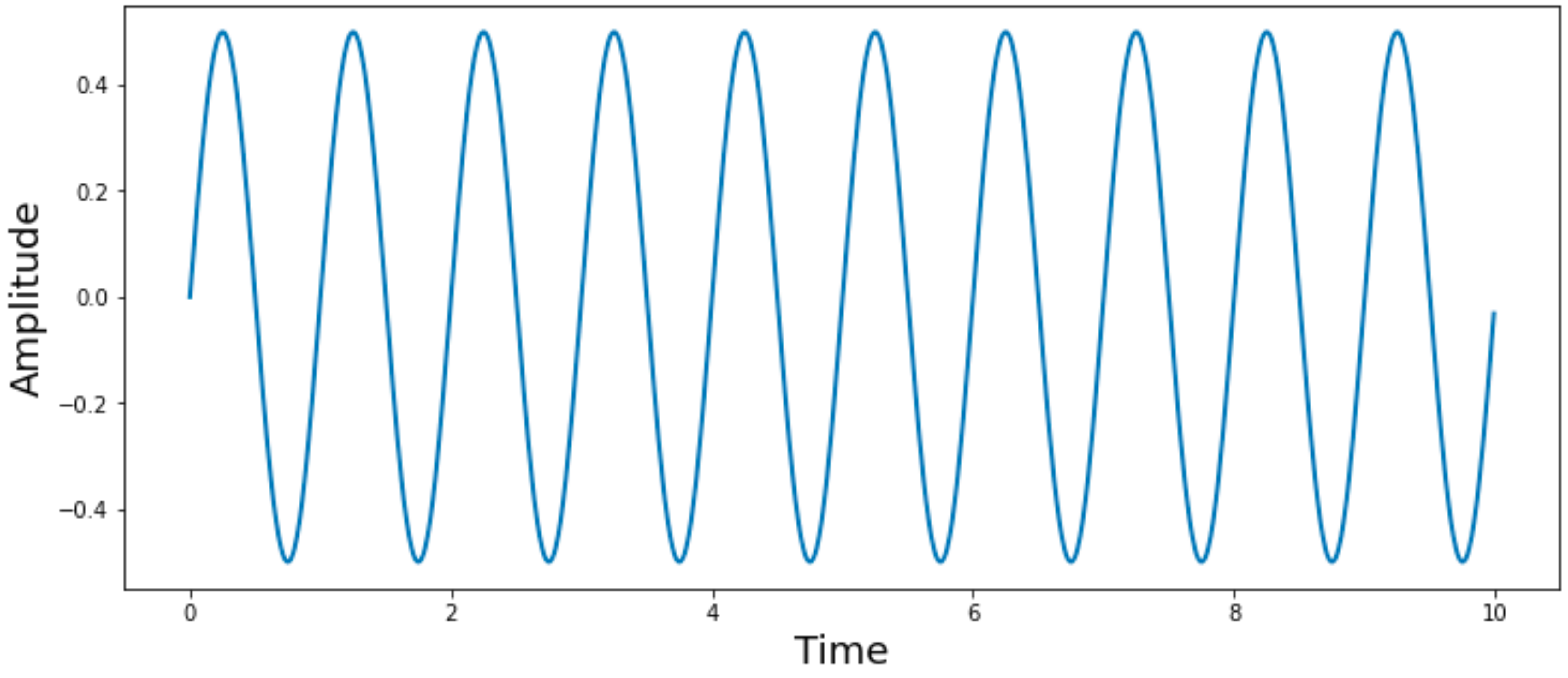


Intensity and loudness

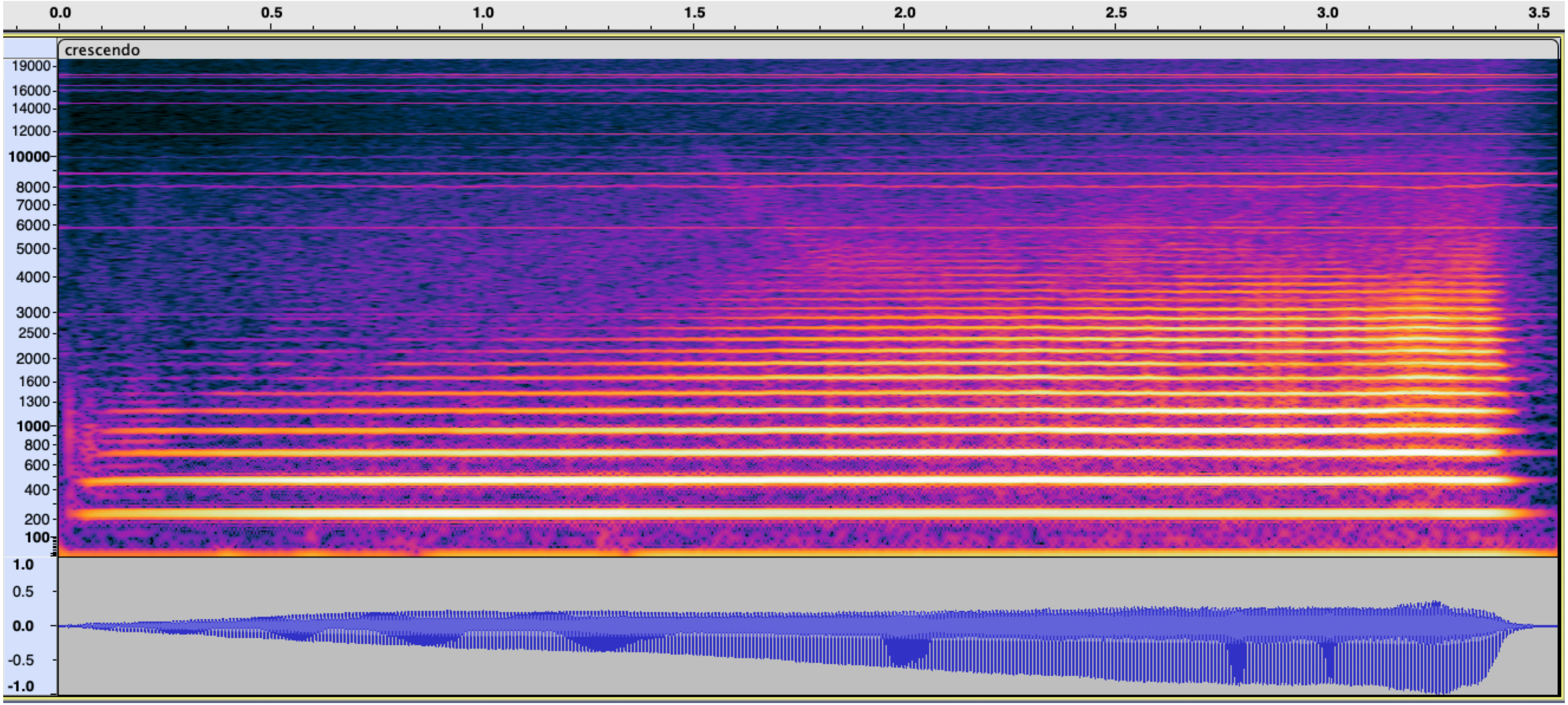
- Intensity is an objective comparison of sound power per unit area. But the ear responds in a non-linear way to that sound intensity.

- Loudness is the strength of the ear's perception of the sound. It is a subjective measurement of perception.



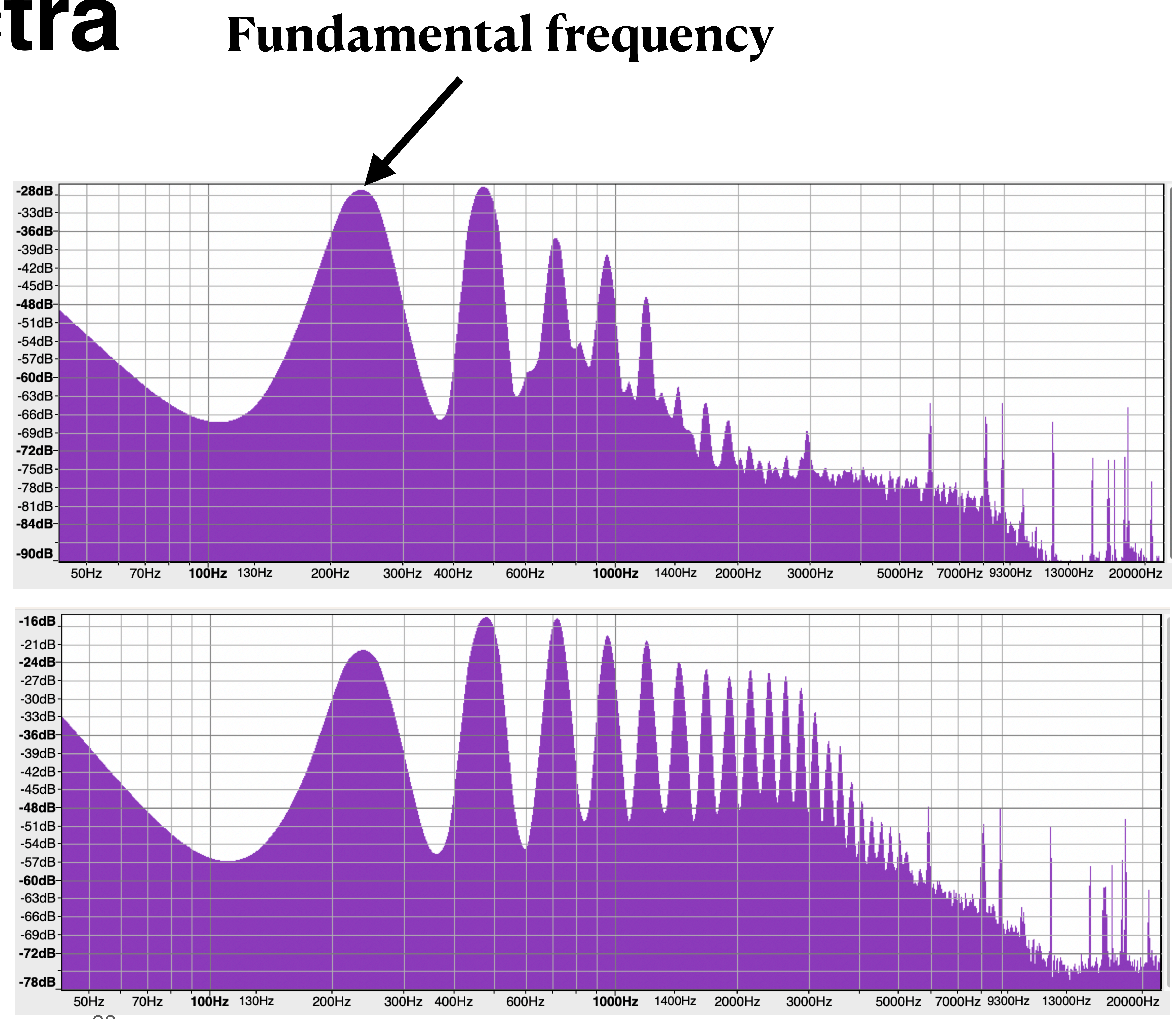


Loudness vs intensity



Loudness and spectra

- Upper spectrum: Spectrum of the ***first*** 0.3 seconds
- Lower spectrum: Spectrum of the ***last*** 0.3 seconds



- Observations
  - Fundamental frequency is hardly changed
  - Higher harmonics make the note sound louder

Decibel (dB)

- Decibel: a logarithmic unit used to measure sound level - difference as a ratio

- Example: One loudspeaker plays a sound with power P_1, and another plays a louder version of the same sound with power P_2, but everything else (how far away, frequency) kept the same

$$10 \log(P_2/P_1) \text{dB}$$

  - P_2 is twice as much power than P_1

$$10 \log(P_2/P_1) = 10 \log(2) \approx 3\text{dB}$$

  - P_2 has ***a million times*** the power of P_1

$$10 \log(P_2/P_1) = 10 \log 1{,}000{,}000 \approx 60\text{dB}$$

Loudness = volume?

- Loudness is the noise level perceived by an individual, whereas volume is an absolute noise level that can be scientifically measured. For example, if your family is watching a movie together, the TV volume is the same for everyone in the room. However, the TV's loudness may be much less for a person with a hearing impairment than it is for a person with normal hearing.

- If you increase the volume on a television, it will also incrementally increase the loudness of the noise. However, increasing the volume will not increase the loudness to the same degree for every person.

Live demo

https://colab.research.google.com/drive/1yk5HOi2bpzQ3MDzBTwGdLZ0JvwILIelZ?usp=sharing

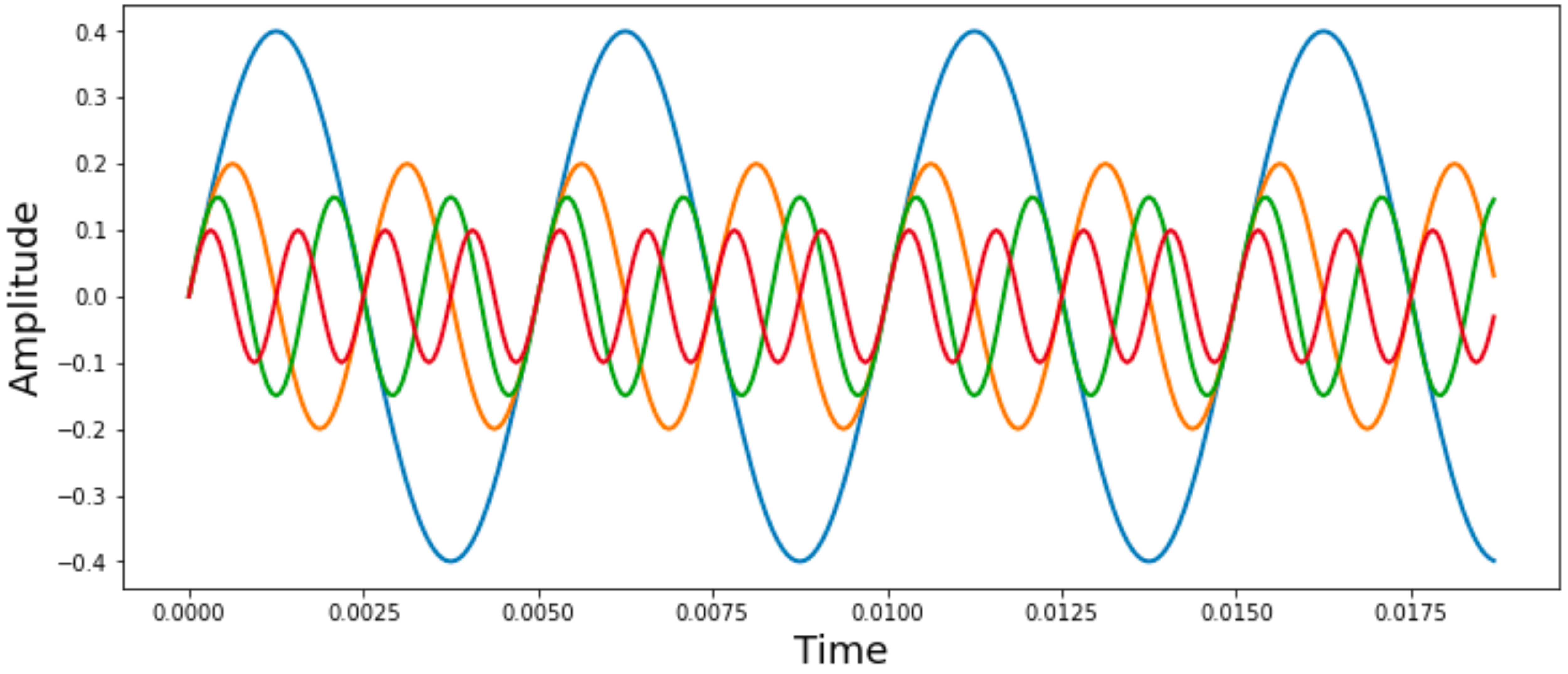
Timbre (also known as tone color/quality)

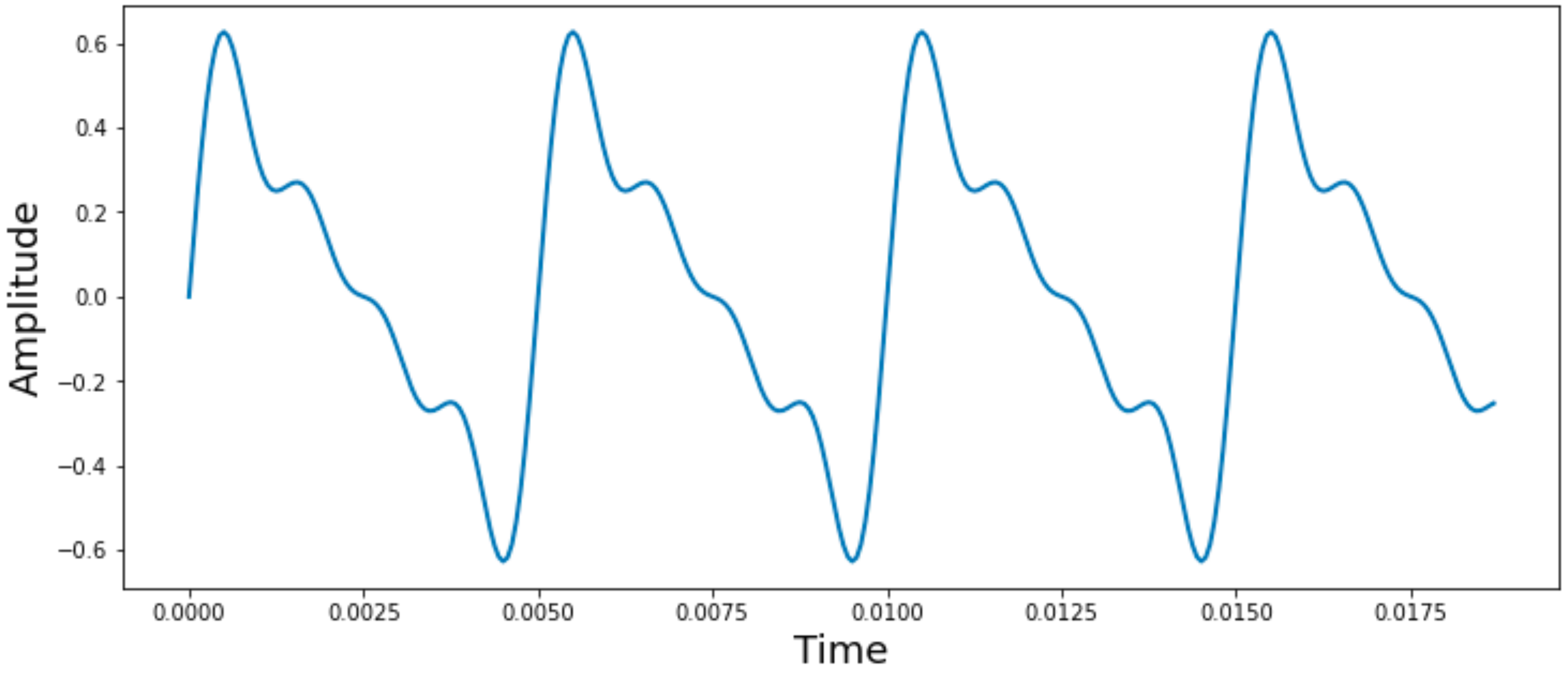
- Depends strongly on ***envelope (time variation)*** and also depends on ***spectrum***

Timbre: Spectrum and harmonics

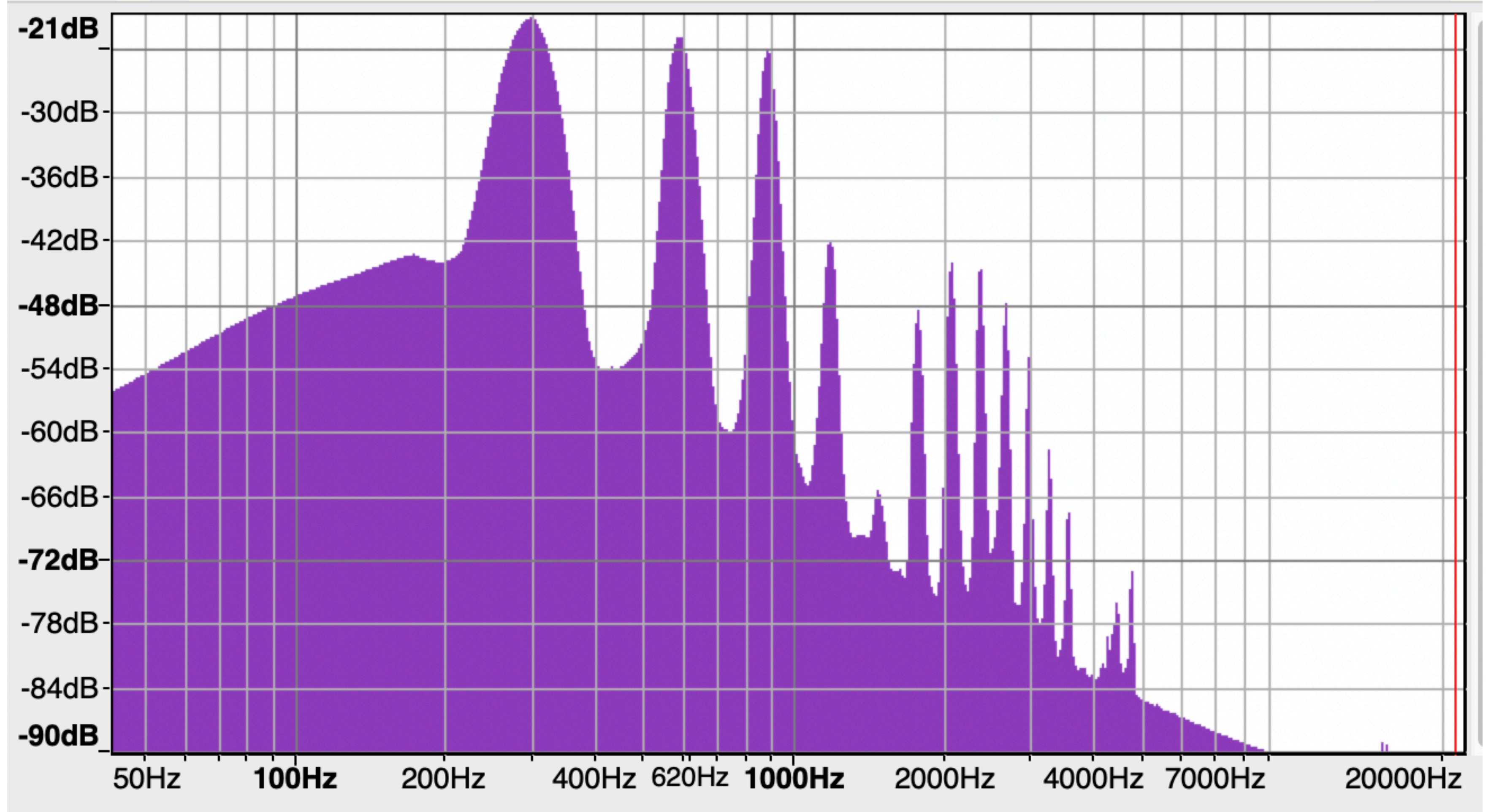
- A periodic wave has a harmonic spectrum
  - Spectrum includes both magnitude of the harmonics and not their relative phases
  - Our ears are not very sensitive to relative phase

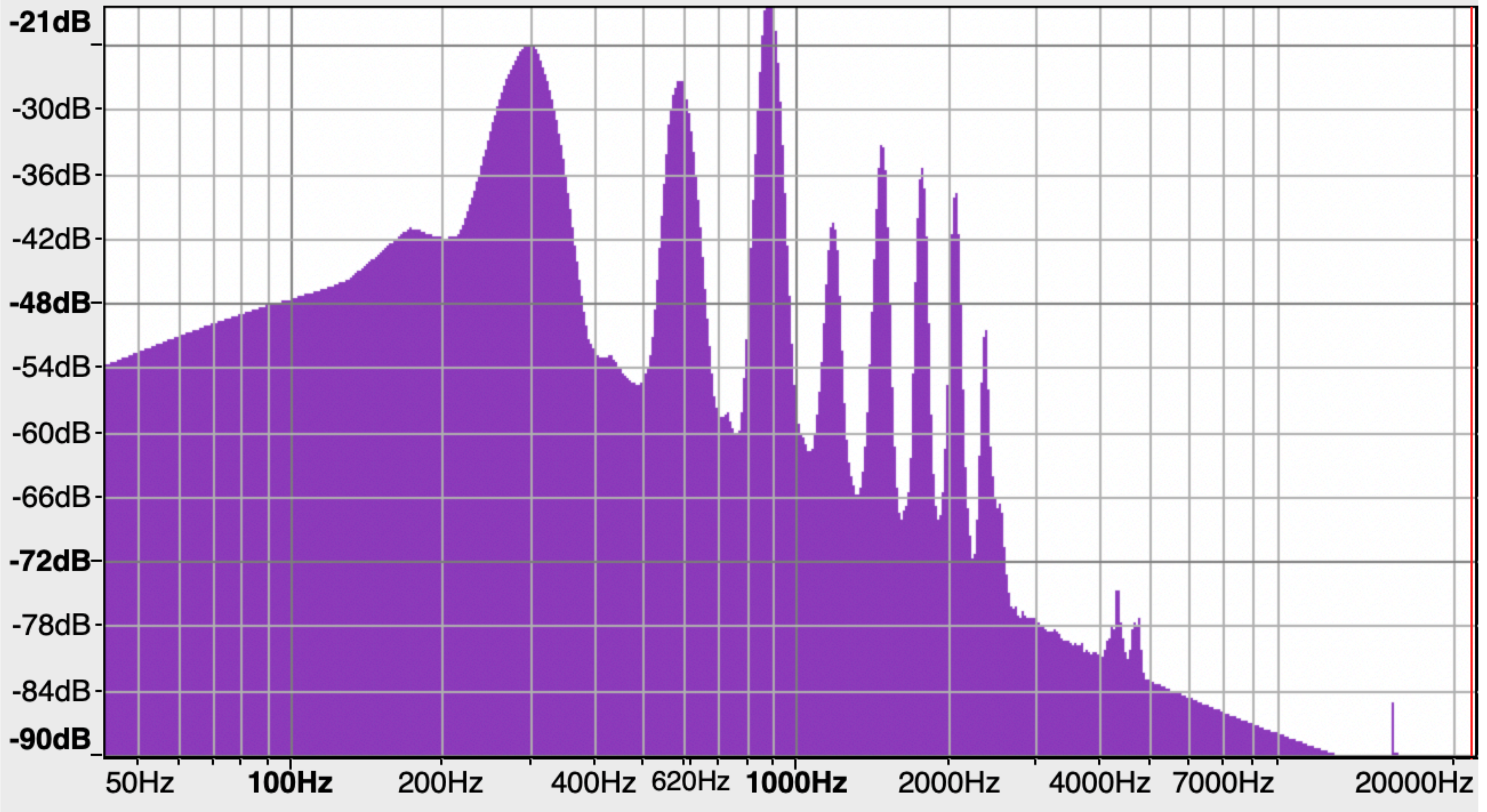
- Harmonic series: a set of frequencies *f*, 2*f*, 3*f*, 4*f*, *f* is the fundamental frequency

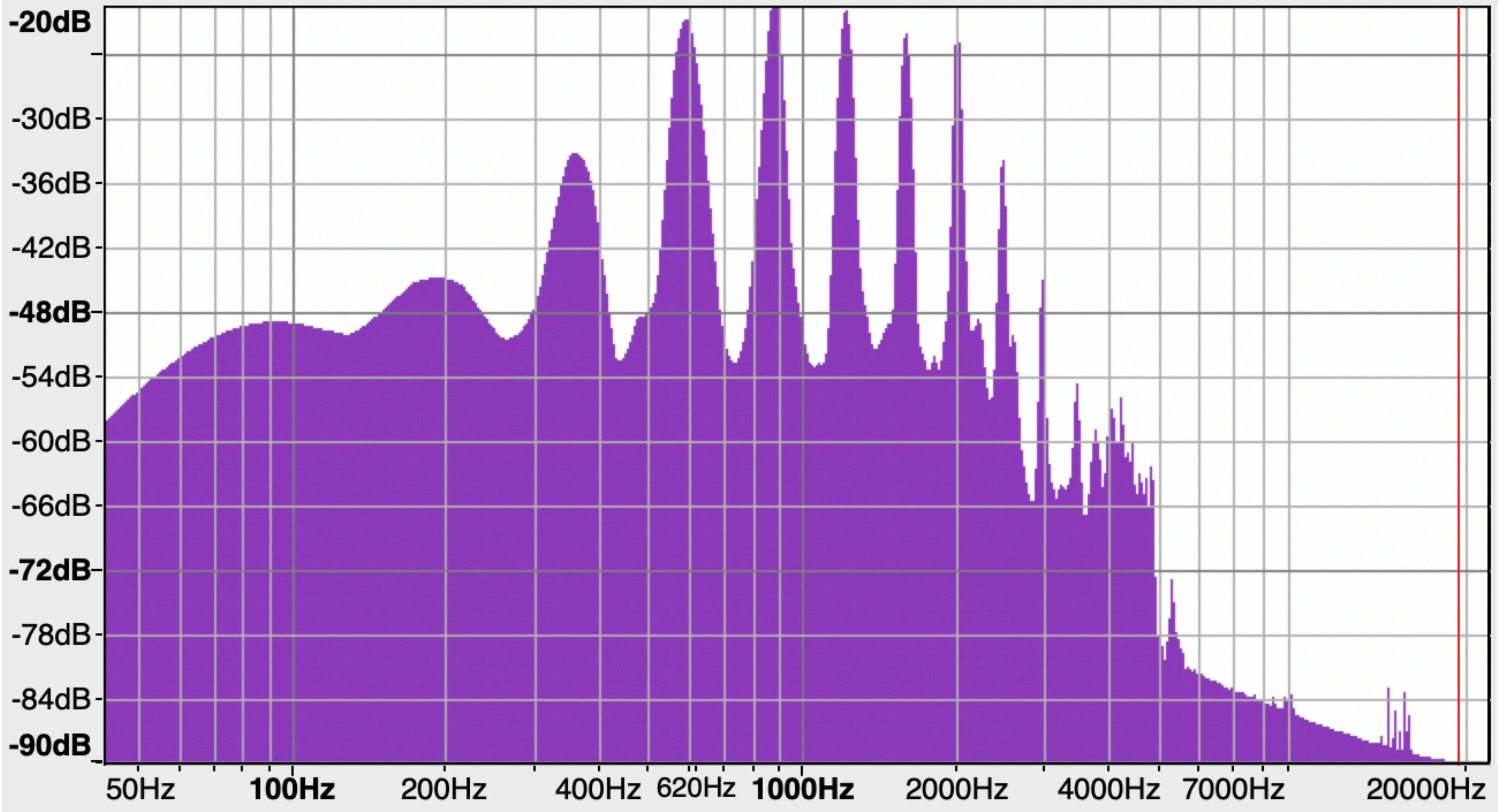


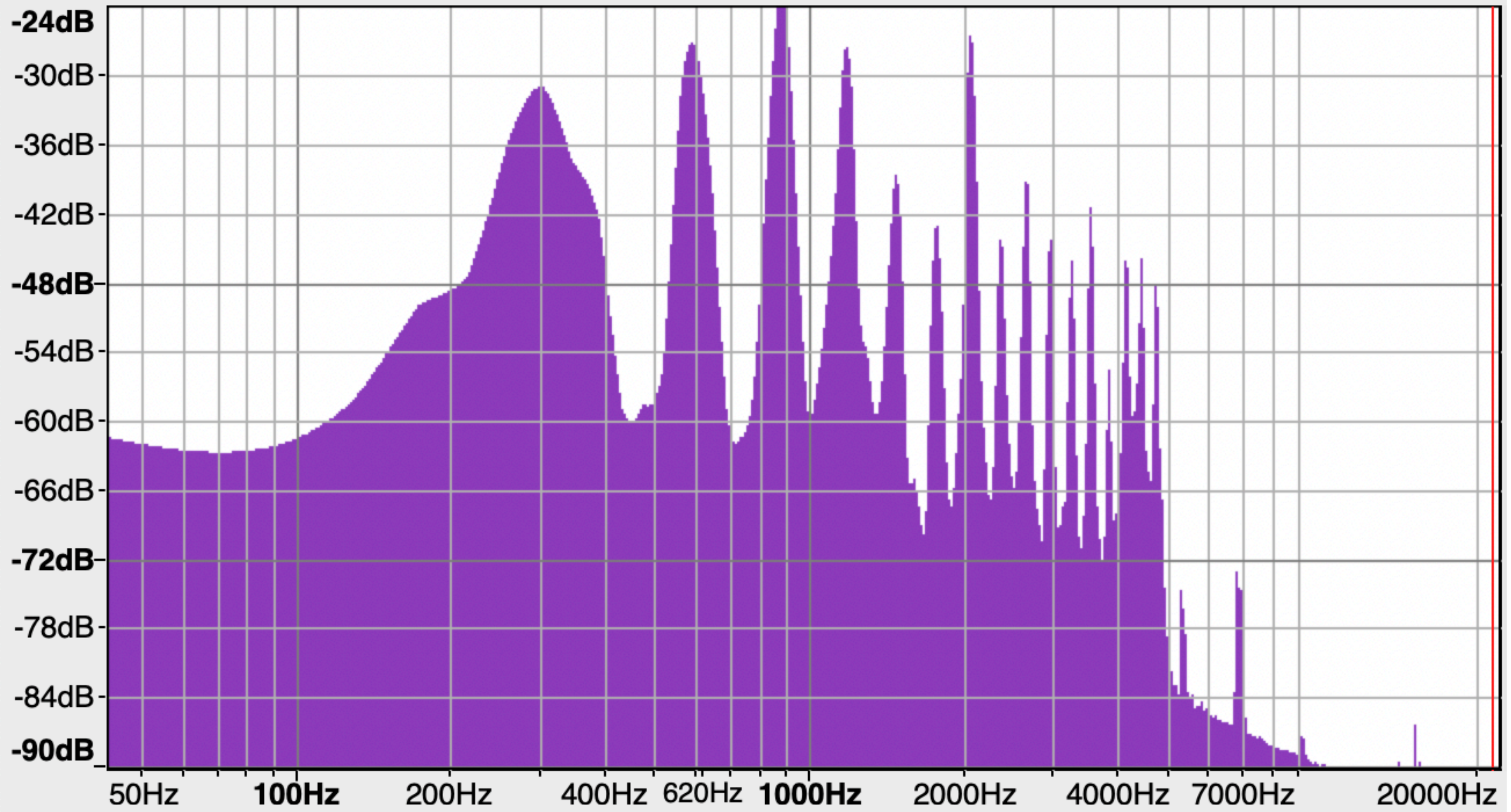


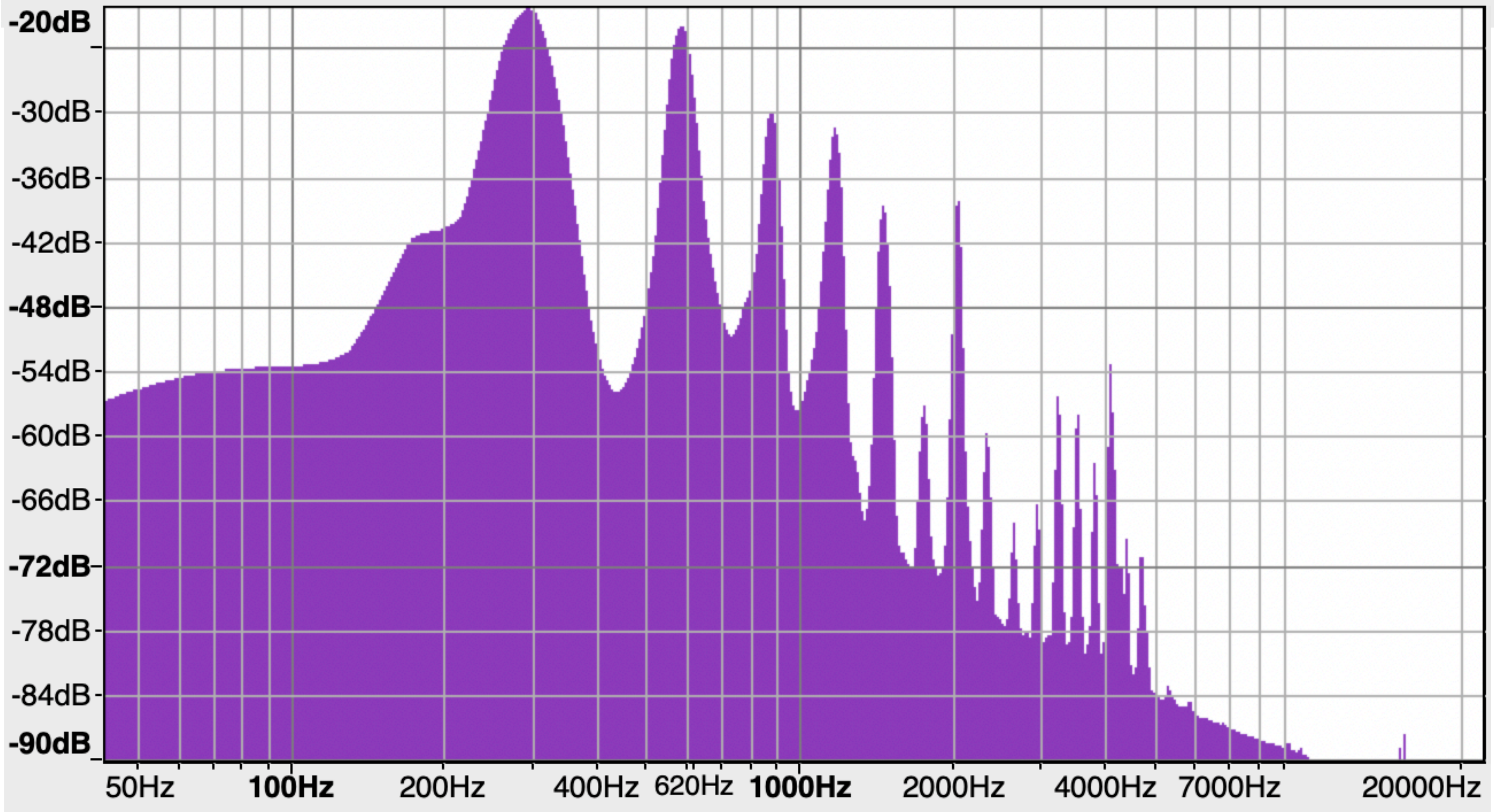
https://colab.research.google.com/drive/1yk5HOi2bpzQ3MDzBTwGdLZoJvwILIelZ?usp=sharing

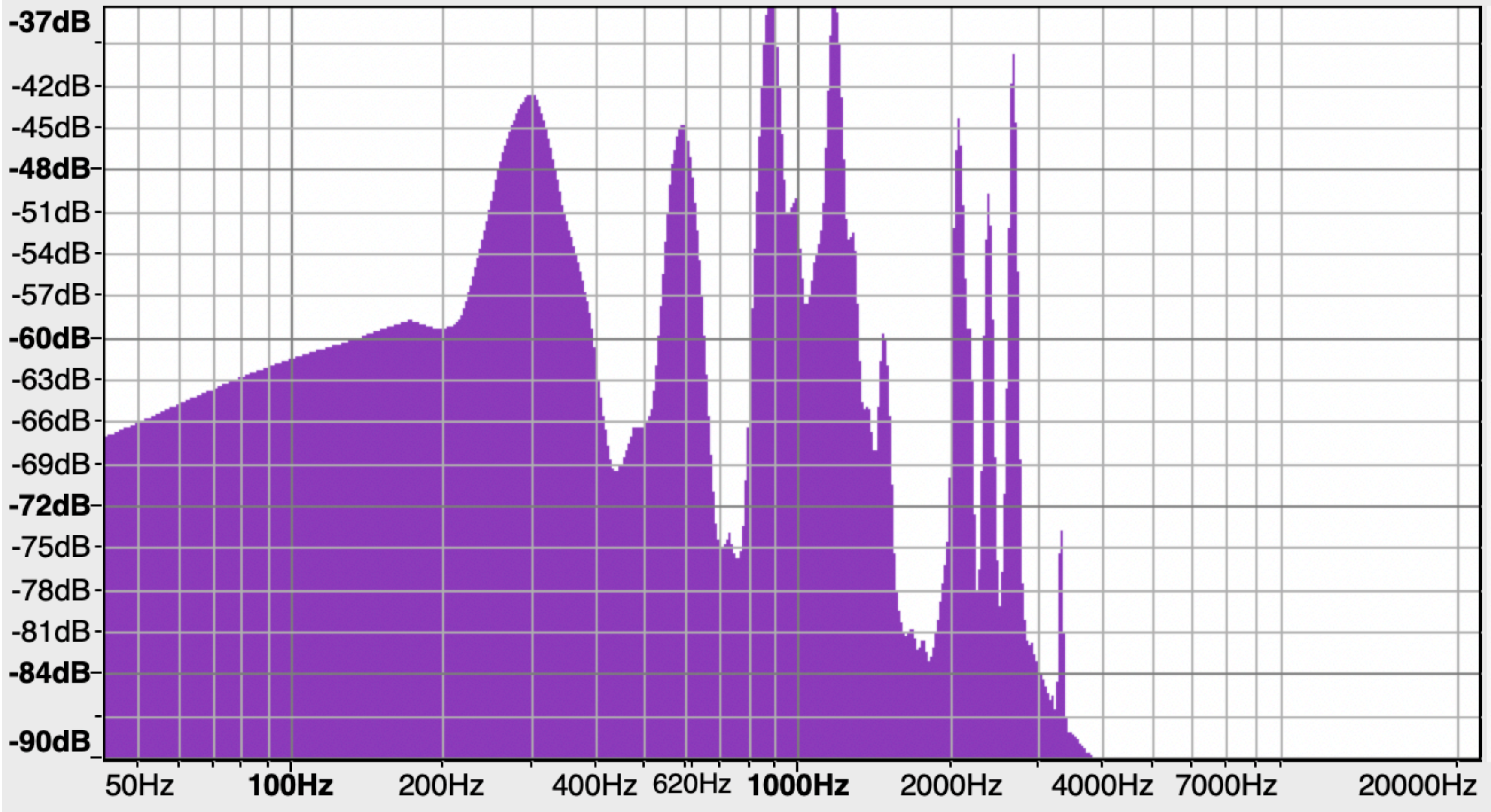

-21dB
-30dB
-36dB
-42dB
-48dB
-54dB
-60dB
-66dB
-72dB
-78dB
-84dB
-90dB
50Hz
100Hz
200Hz
400Hz
620Hz
1000Hz
2000Hz
4000Hz
7000Hz
20000Hz


-21dB
-30dB
-36dB
-42dB
-48dB
-54dB
-60dB
-66dB
-72dB
-78dB
-84dB
-90dB
50Hz
100Hz
200Hz
400Hz
620Hz
1000Hz
2000Hz
4000Hz
7000Hz
20000Hz

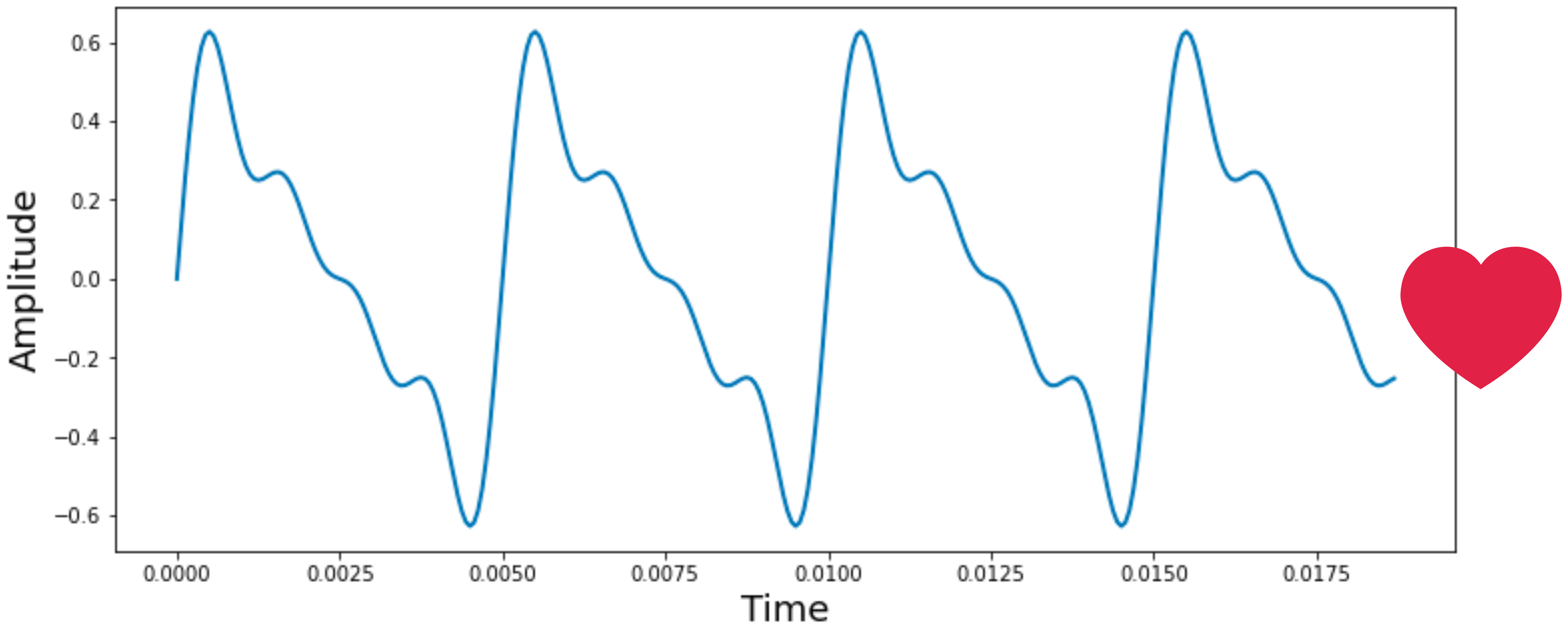

-20dB
-30dB
-36dB
-42dB
-48dB
-54dB
-60dB
-66dB
-72dB
-78dB
-84dB
-90dB
50Hz
100Hz
200Hz
400Hz
620Hz
1000Hz
2000Hz
4000Hz
7000Hz
20000Hz

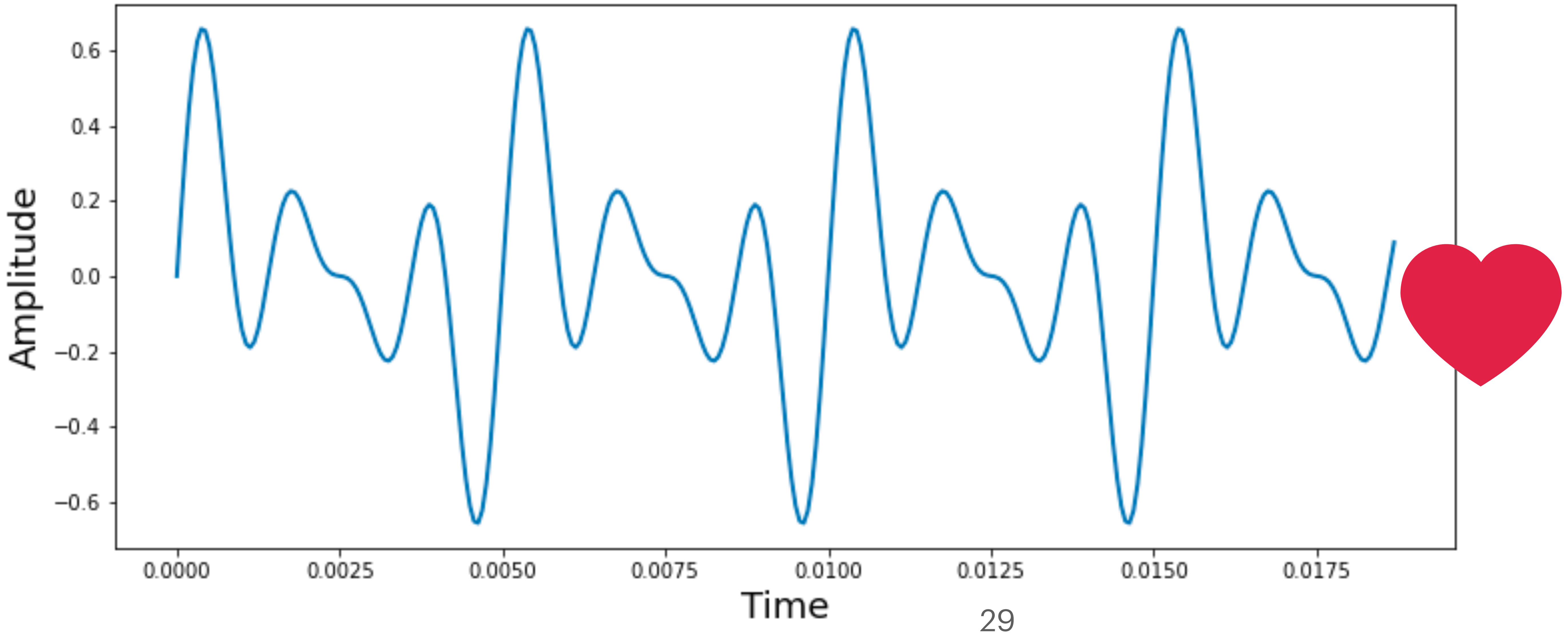

-24dB
-30dB
-36dB
-42dB
-48dB
-54dB
-60dB
-66dB
-72dB
-78dB
-84dB
-90dB
50Hz
100Hz
200Hz
400Hz
620Hz
1000Hz
2000Hz
4000Hz
7000Hz
20000Hz


-20dB
-30dB
-36dB
-42dB
-48dB
-54dB
-60dB
-66dB
-72dB
-78dB
-84dB
-90dB
50Hz
100Hz
200Hz
400Hz
620Hz
1000Hz
2000Hz
4000Hz
7000Hz
20000Hz

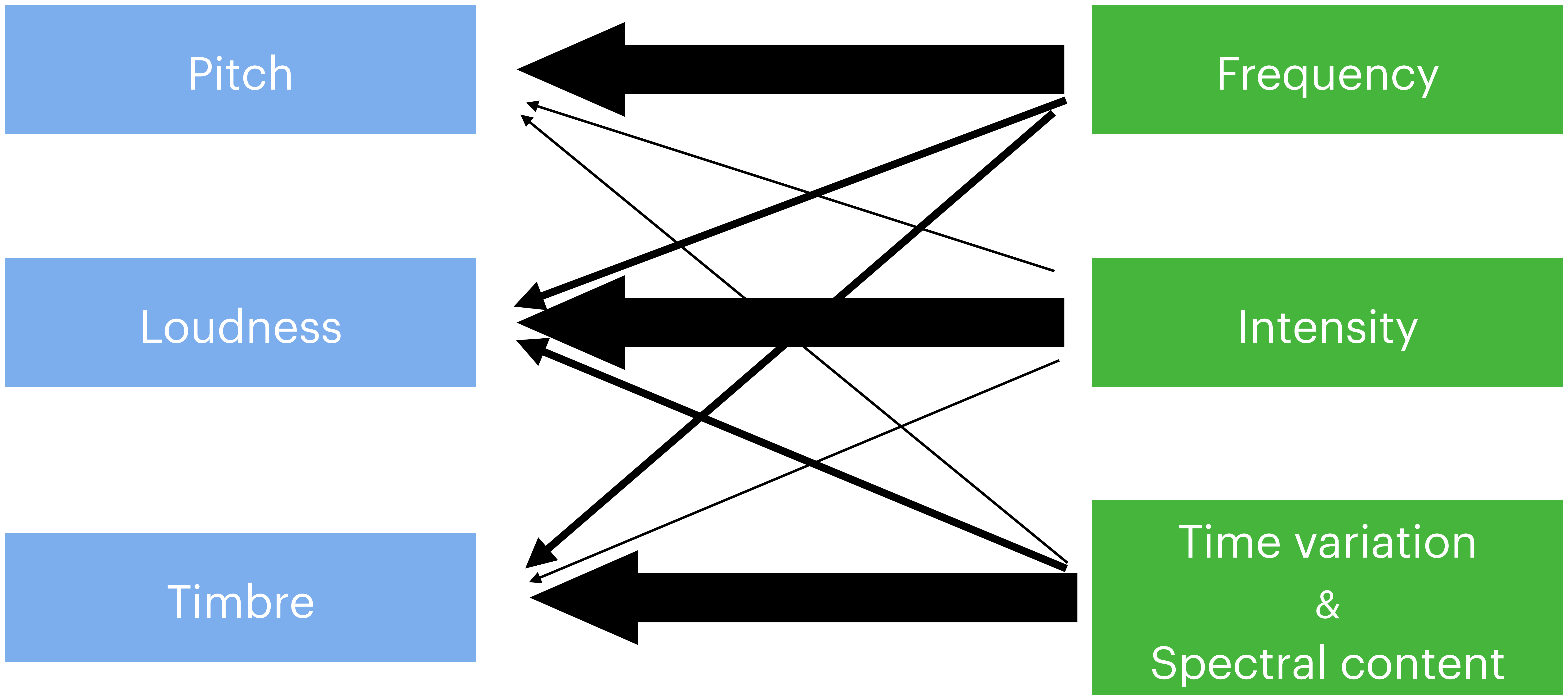

-37dB
-42dB
-45dB
-48dB
-51dB
-54dB
-57dB
-60dB
-63dB
-66dB
-69dB
-72dB
-75dB
-78dB
-81dB
-84dB
-90dB
50Hz
100Hz
200Hz
400Hz
620Hz
1000Hz
2000Hz
4000Hz
7000Hz
20000Hz

Timbre: Time variations (envelope)

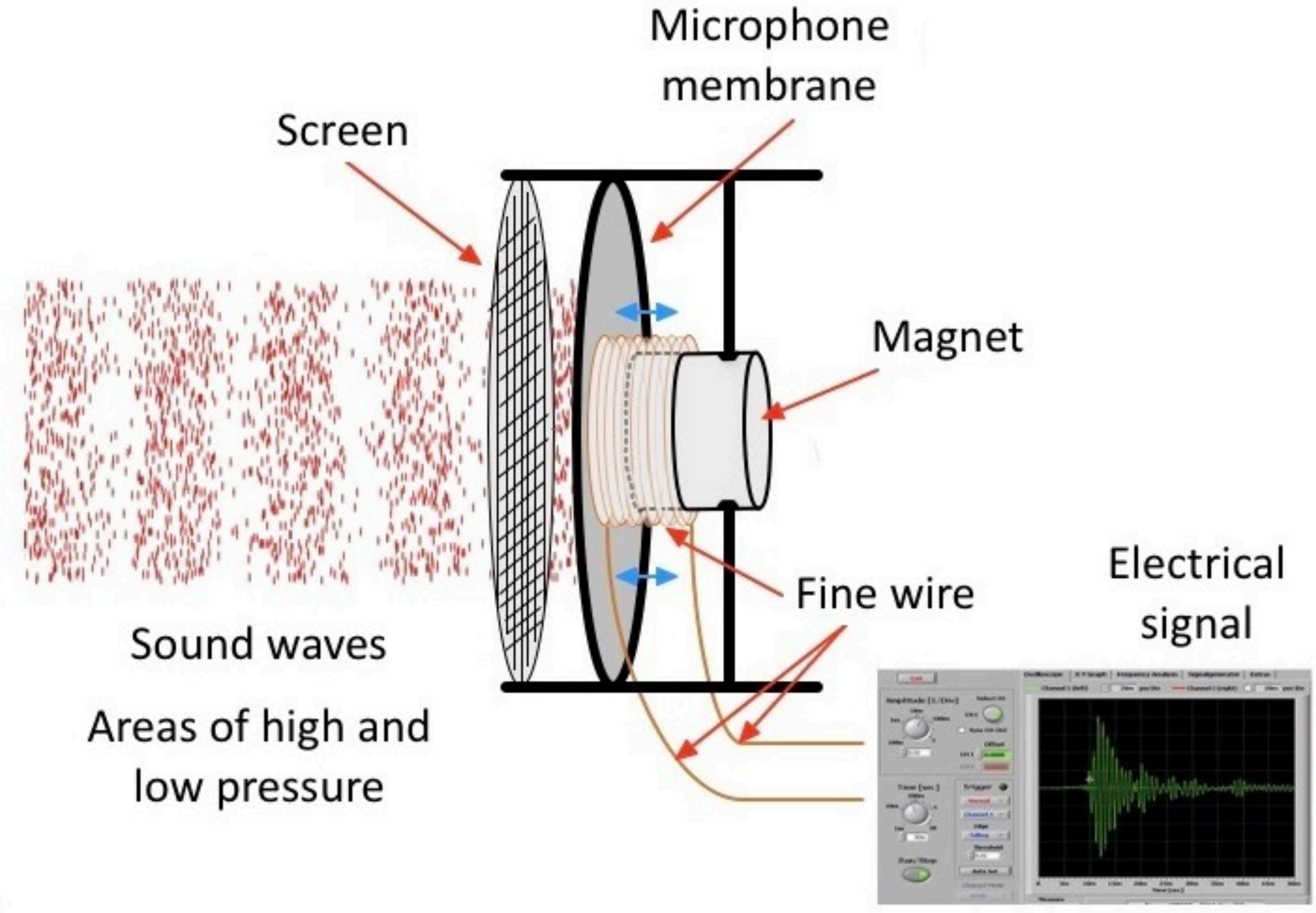




Physical property vs perceptual property

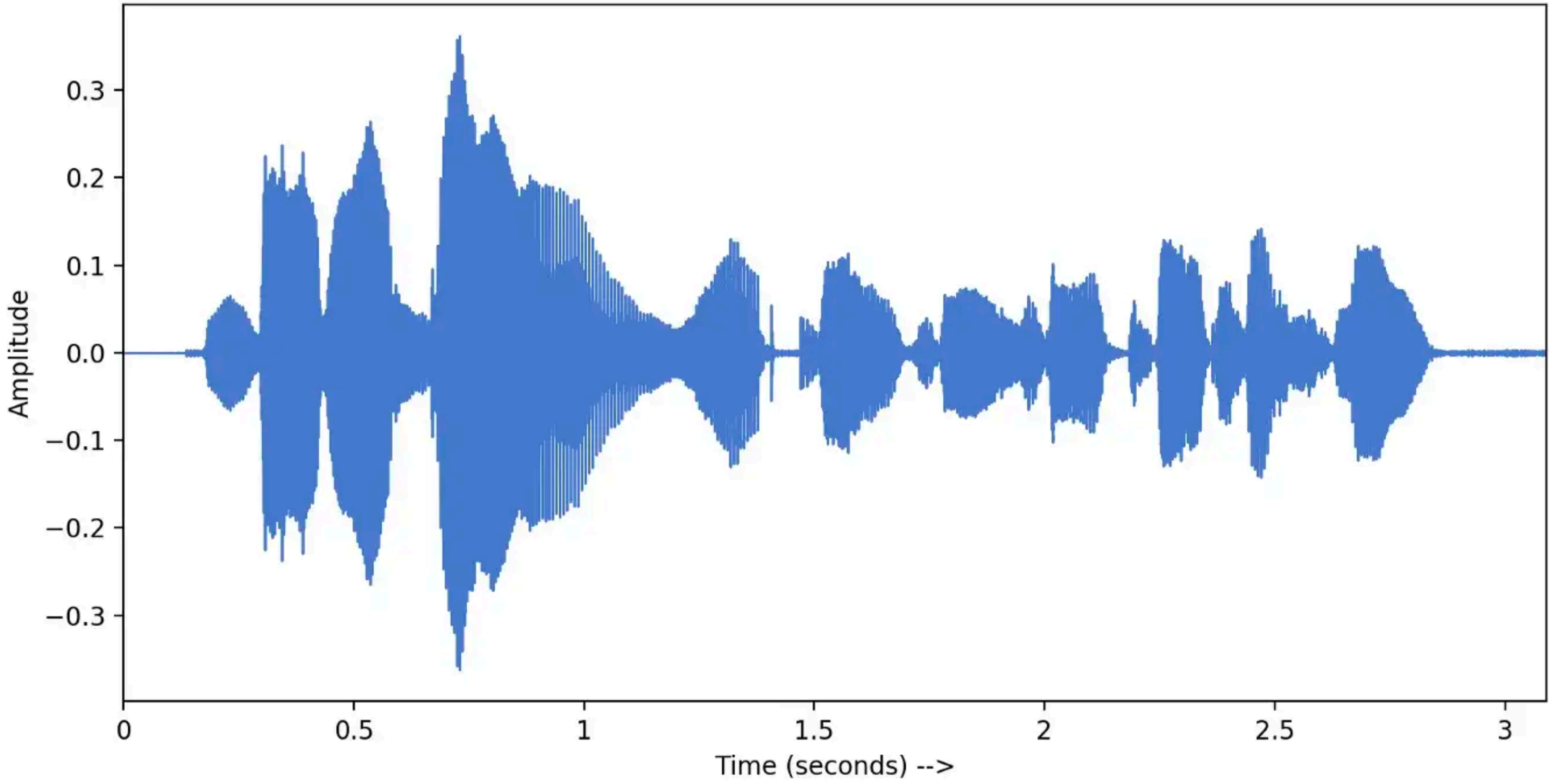


Digital sound waves



Digital sound waves

- Microphones convert sound pressure variations into changes in continuous electrical voltage
  - They capture changes in air pressure to record sound (continuous electrical signal)



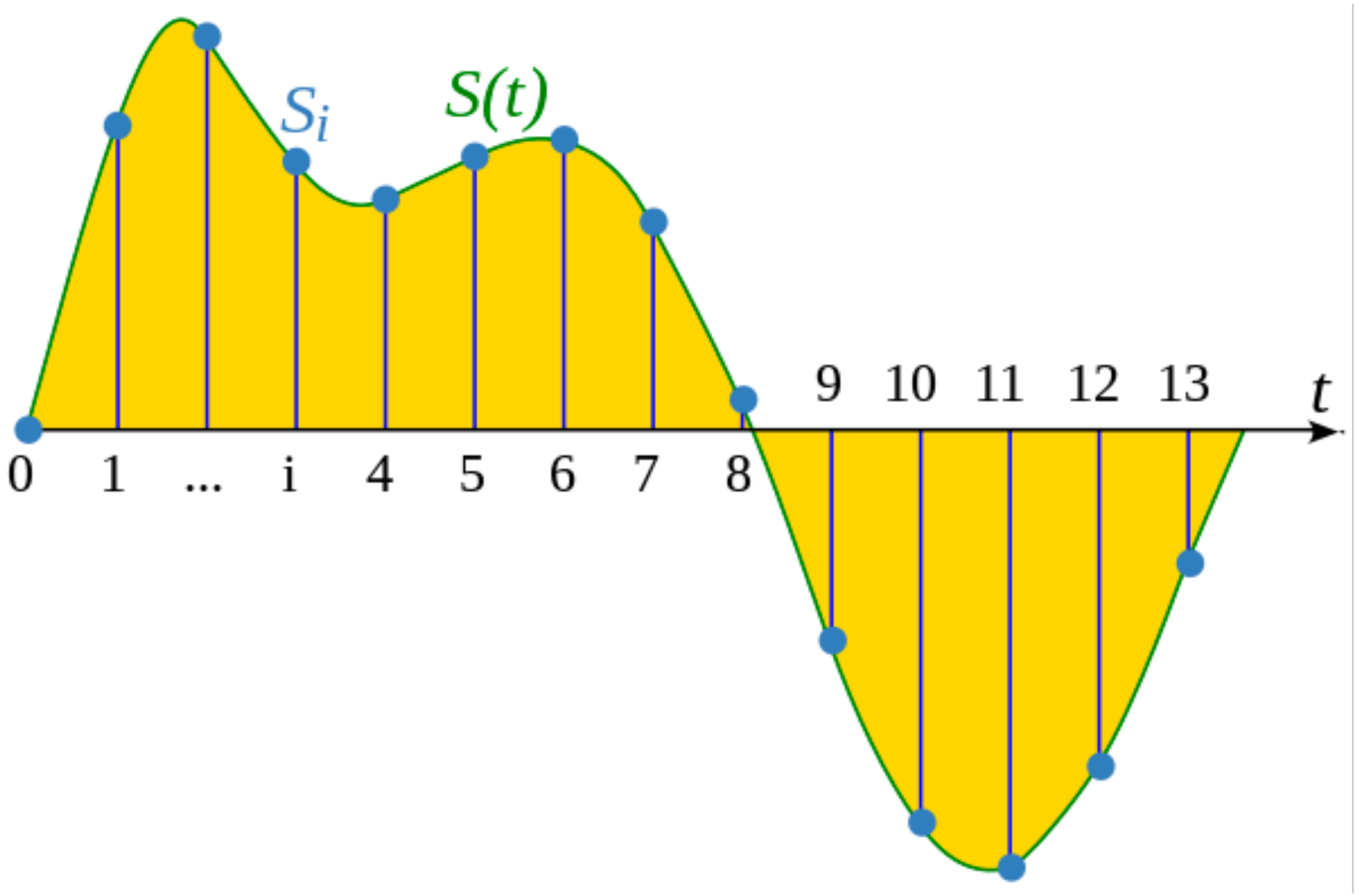
Digital sound waves

- Problem: Computers deal with discrete data (zeros and ones)

  - We need to convert (sample) the continuous signal into digital presentation
    - **Sampling** converts a time-varying voltage signal into a discrete-time signal, a sequence of real numbers.
    - **Quantization** replaces each real number with an approximation from a finite set of discrete values.

Analog signal to digital signal: Sampling

- Sampling period = 1/sampling rate (seconds)



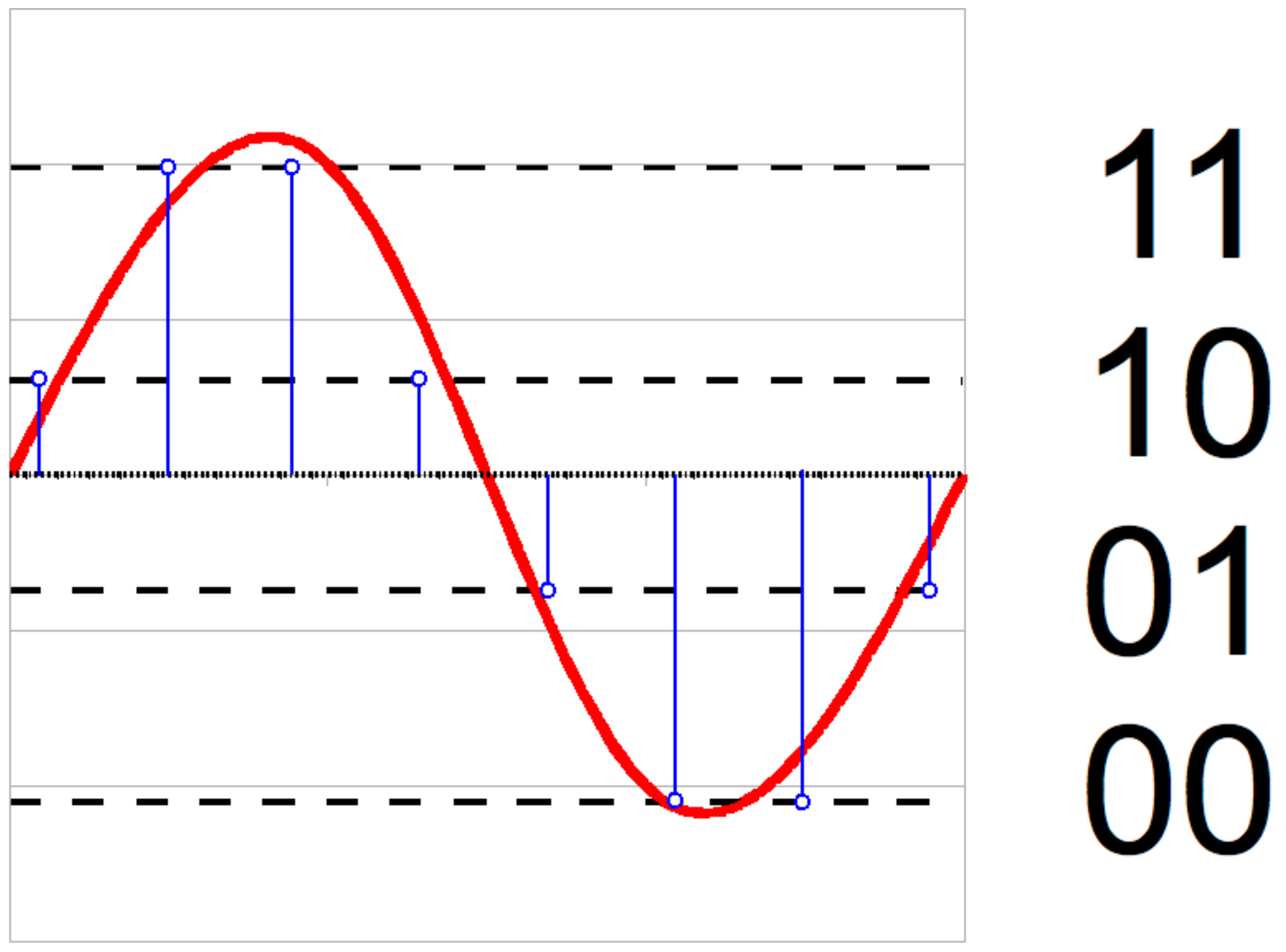
Signal sampling

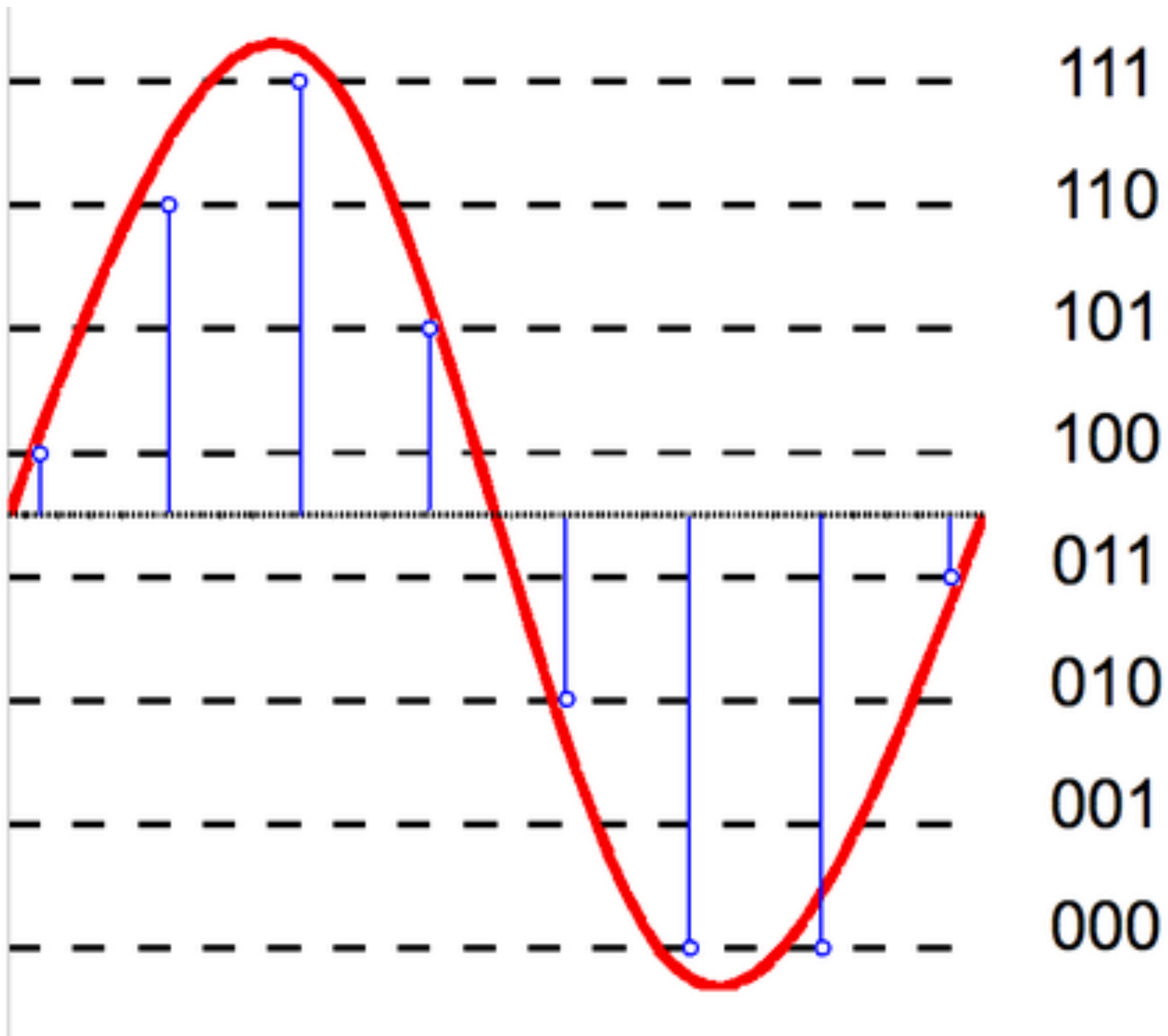
- Typical sampling rates and samples

Sampling rate	Use cases
8 kHz	Telephone and encrypted walkie-talkie, wireless intercom and wireless microphone transmission
16 kHz	Used in most modern VoIP and VVoIP communication products. Wideband extension over standard telephone narrowband.
22.05 kHz	One half the sampling rate of audio CDs; used for lower-quality PCM and MPEG audio.
44.1 kHz	Audio CD, also most commonly used with MPEG-1 audio (VCD, SVCD, MP3).

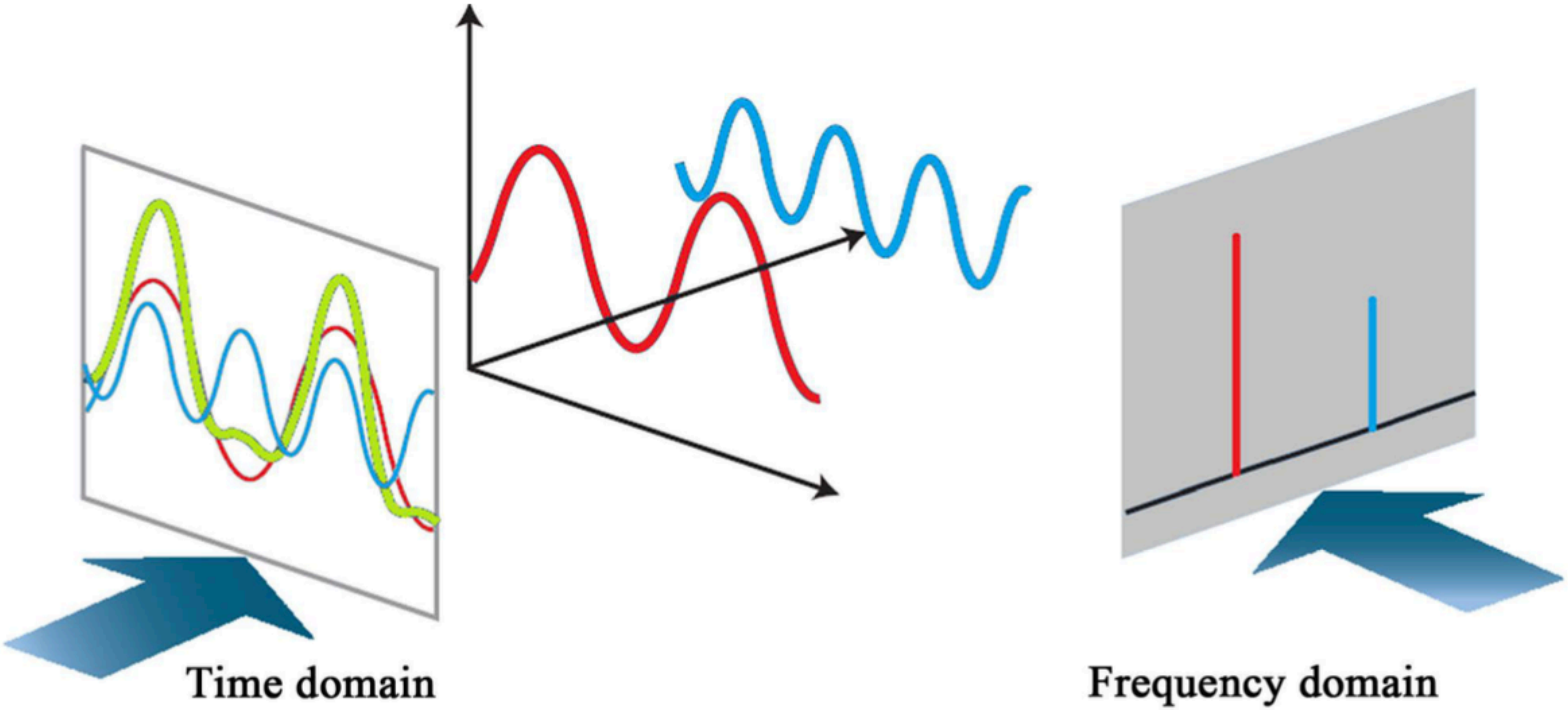
https://en.wikipedia.org/wiki/Sampling_(signal_processing)

Analog signal to digital signal: Quantization

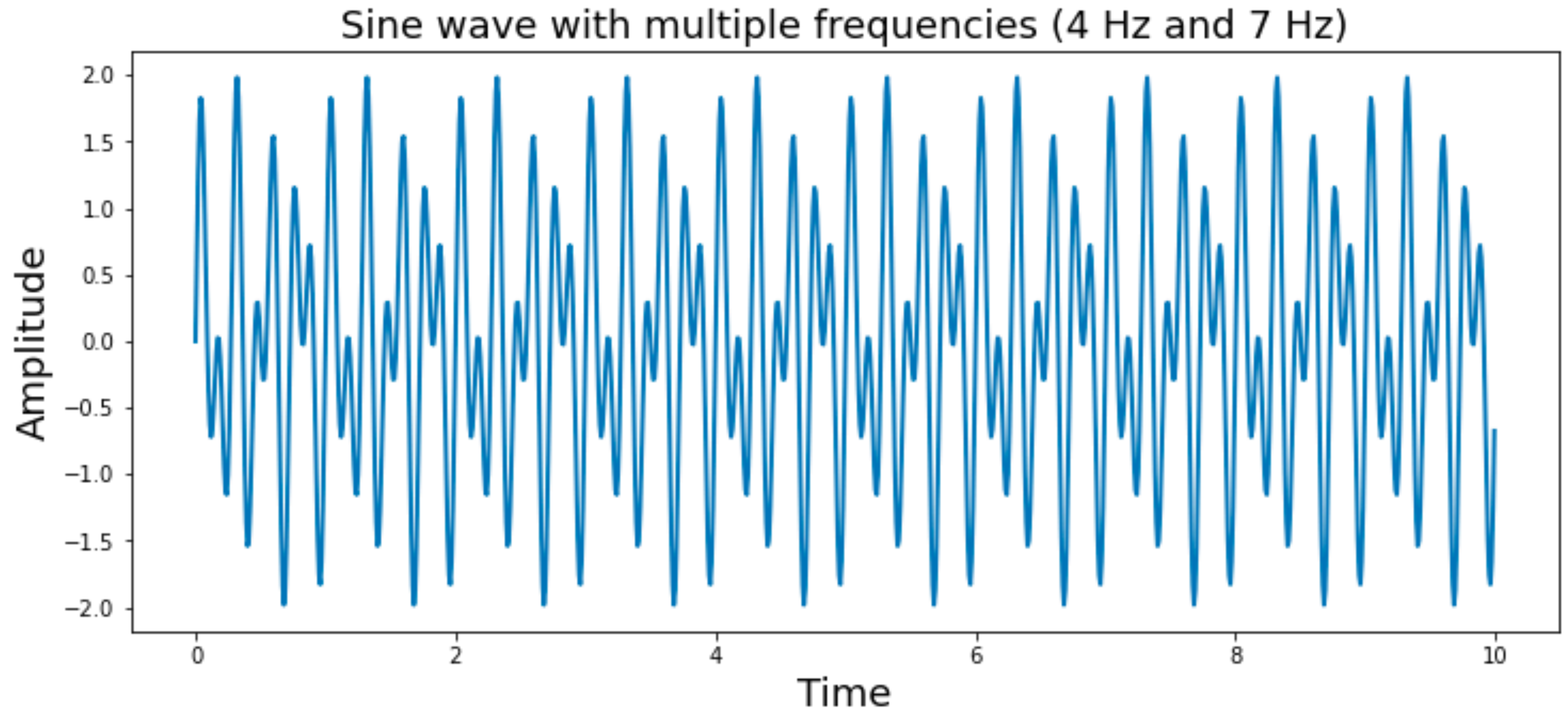


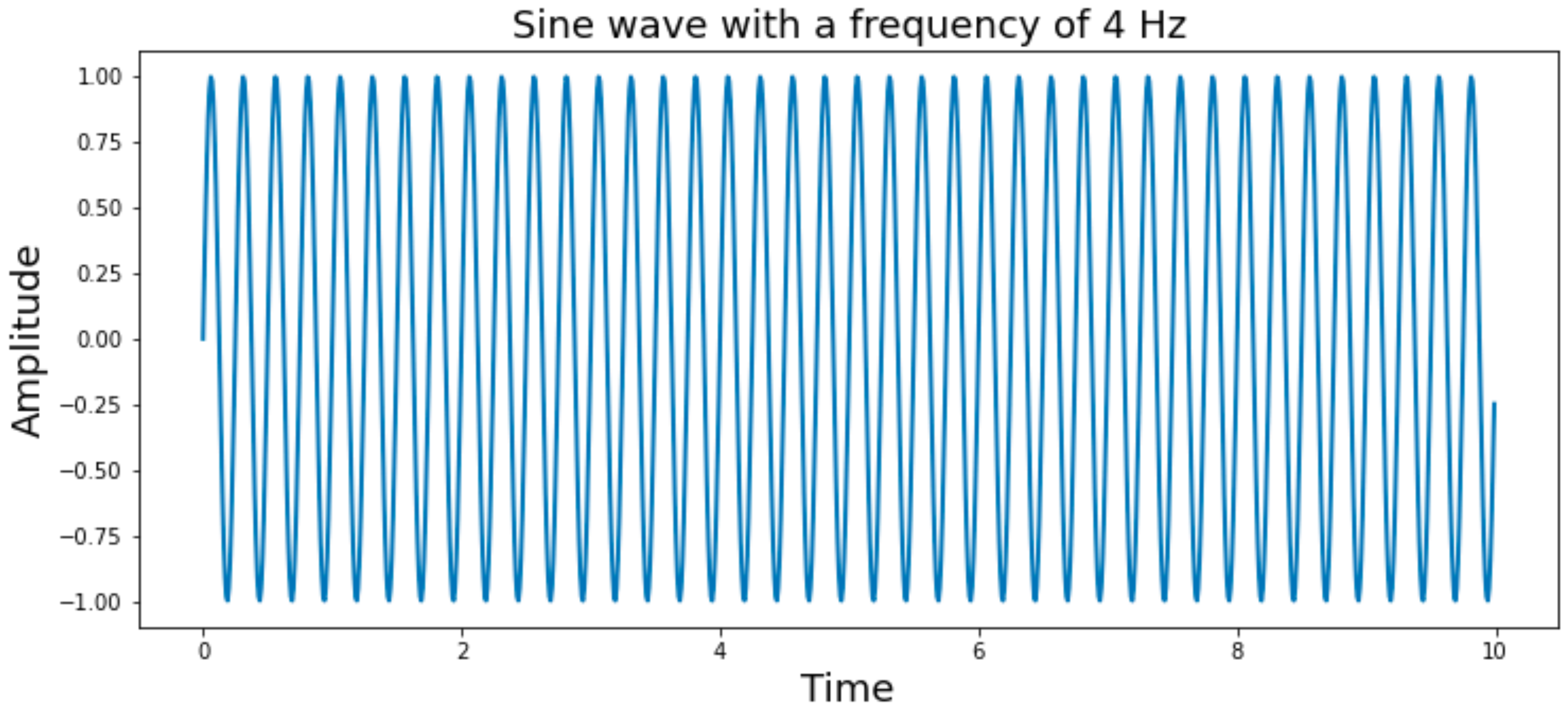


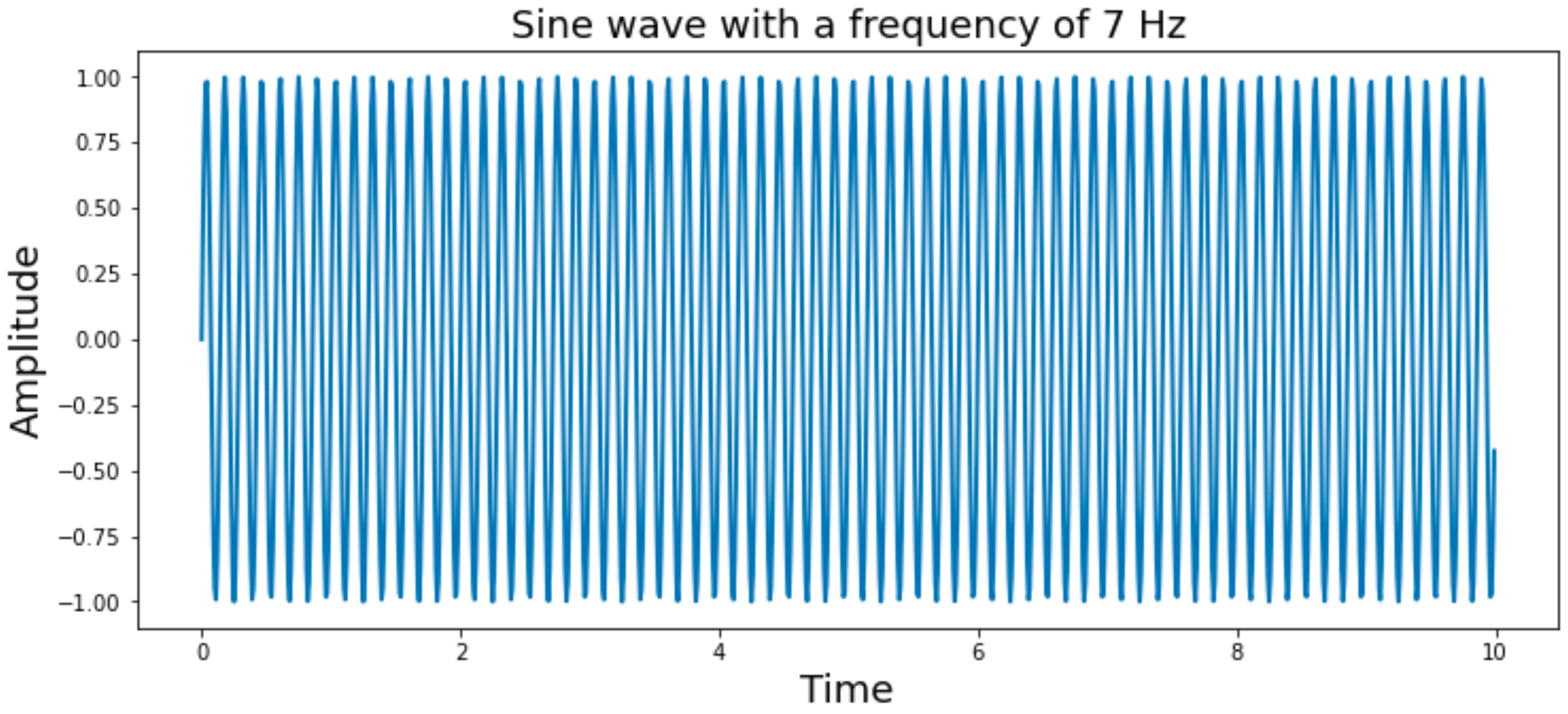
Time domain vs frequency domain



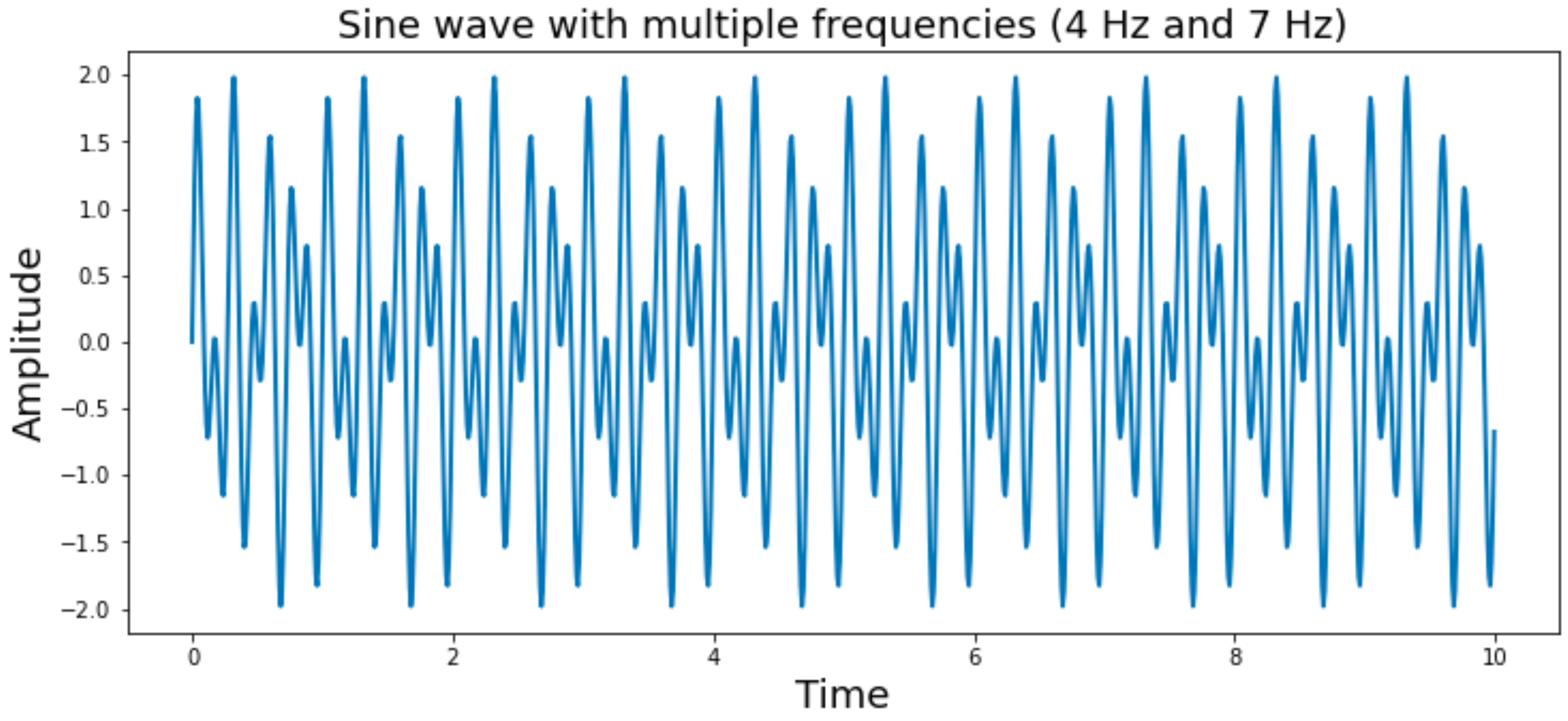
A signal in time domain

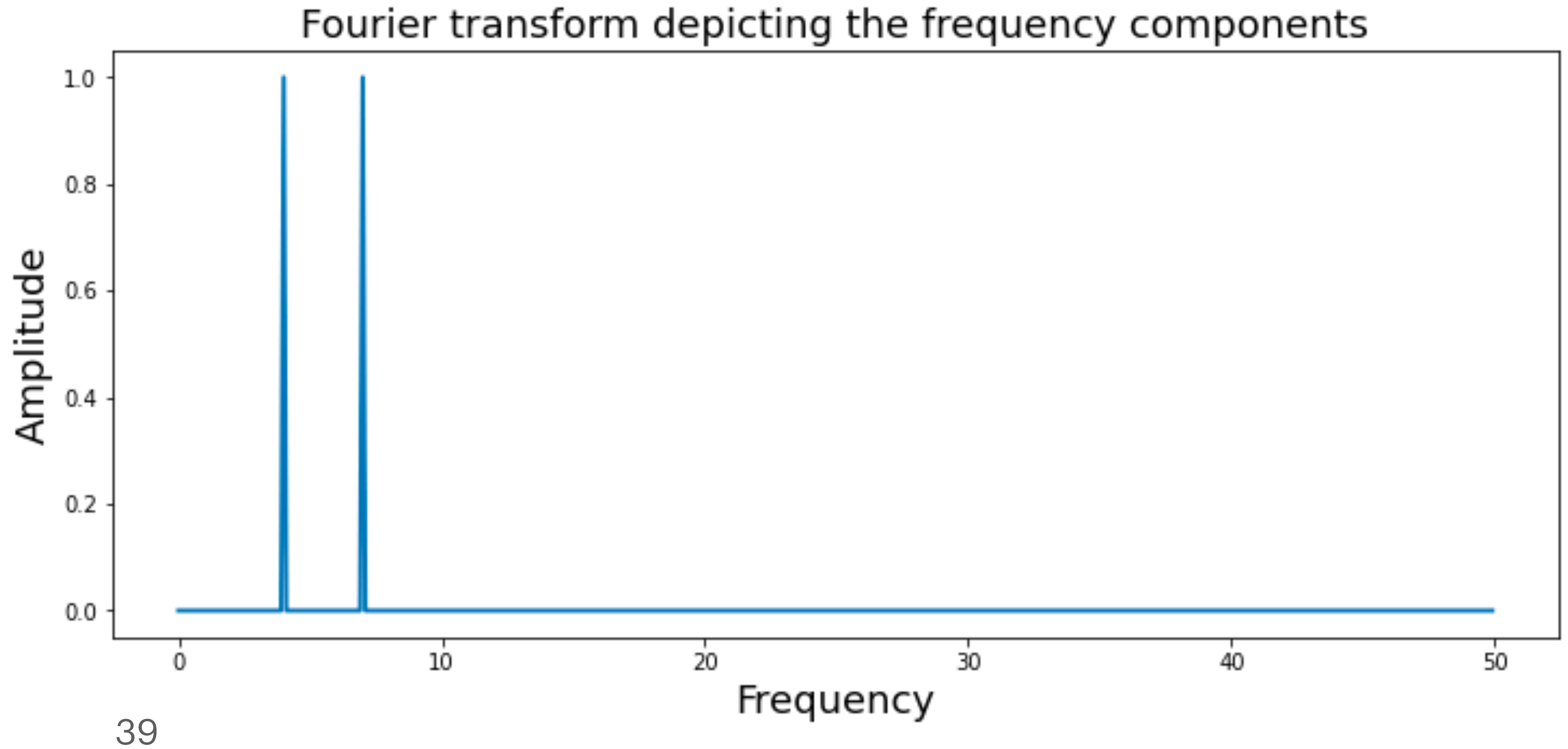






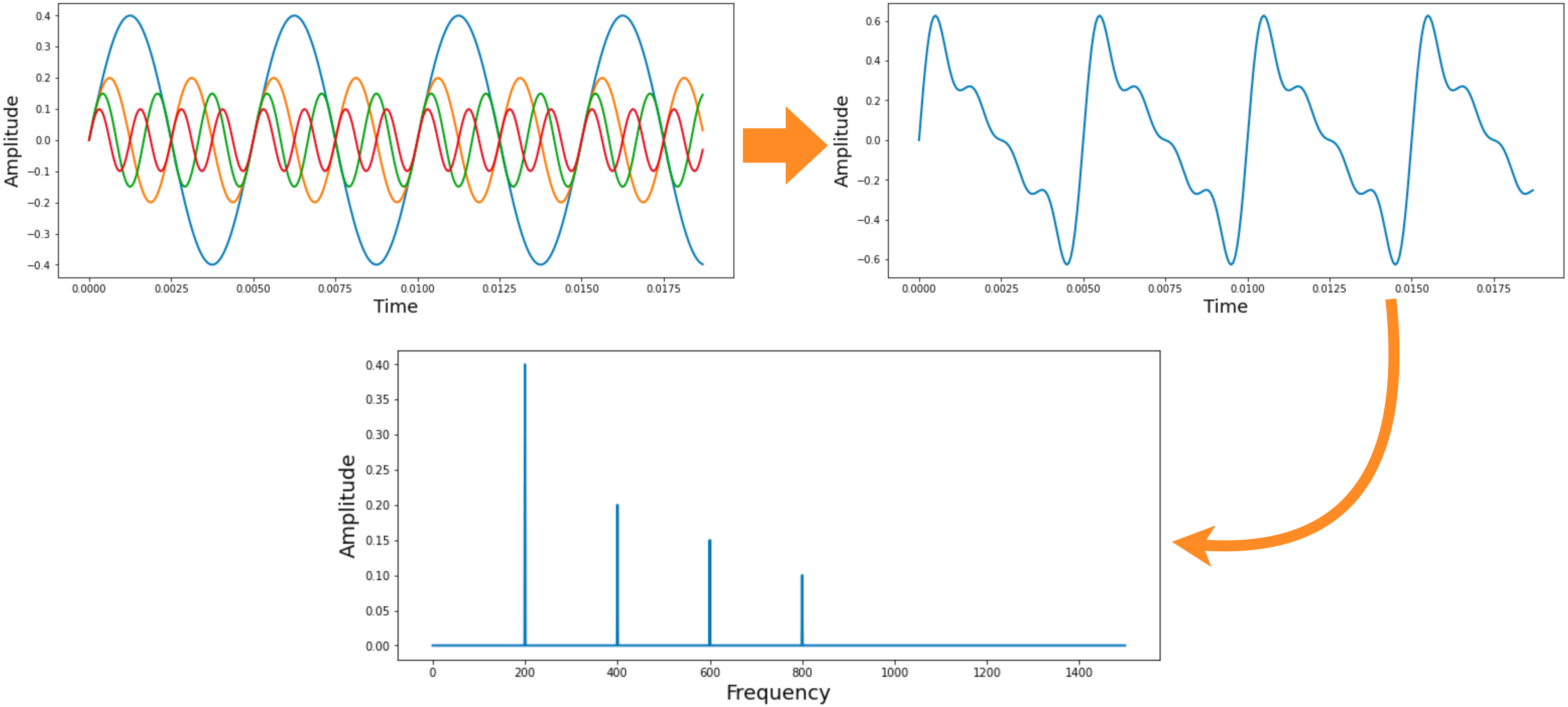
Frequency-domain representation





https://colab.research.google.com/drive/1yk5HOi2bpzQ3MDzBTwGdLZoJvwILIelZ?usp=sharing

Frequency-domain representation



Frequency analysis

Summary

- Quantifying sound
  - Physical property: Frequency, intensity, time variation and spectrum
  - Perceptual property: Pitch, loudness and timbre
- Digital sound wave
  - Sampling and quantization
- Time domain vs frequency domain
  - Frequency domain representation and frequency analysis