

# Understanding audio signals for ML

Valerio Velardo

# Audio signal

---

- Representation of sound
- Encodes all info we need to reproduce sound

# Houston we have a problem!

---

# Houston we have a problem!

---

Analog



VS

Digital



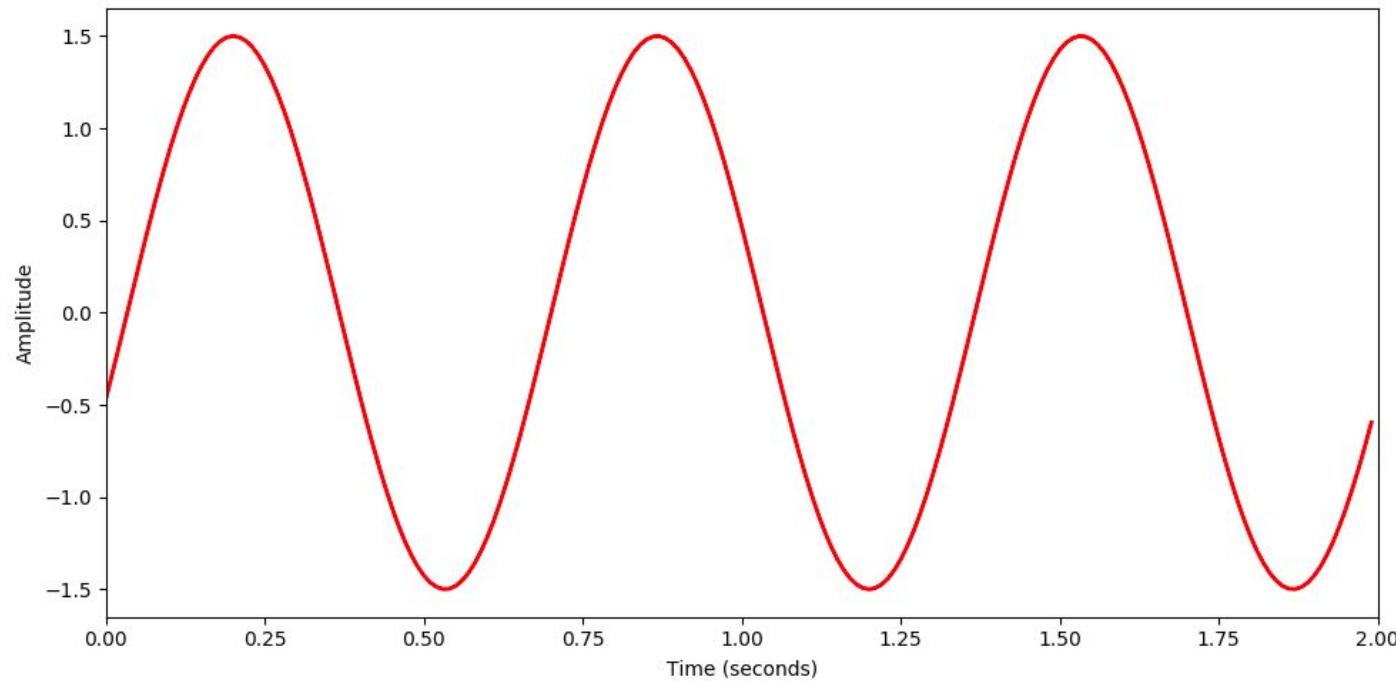
# Analog signal

---

- Continuous values for time
- Continuous values for amplitude

# Analog signal

---



# Digital signal

---

- Sequence of discrete values
- Data points can only take on a finite number of values

# Analog to digital conversion

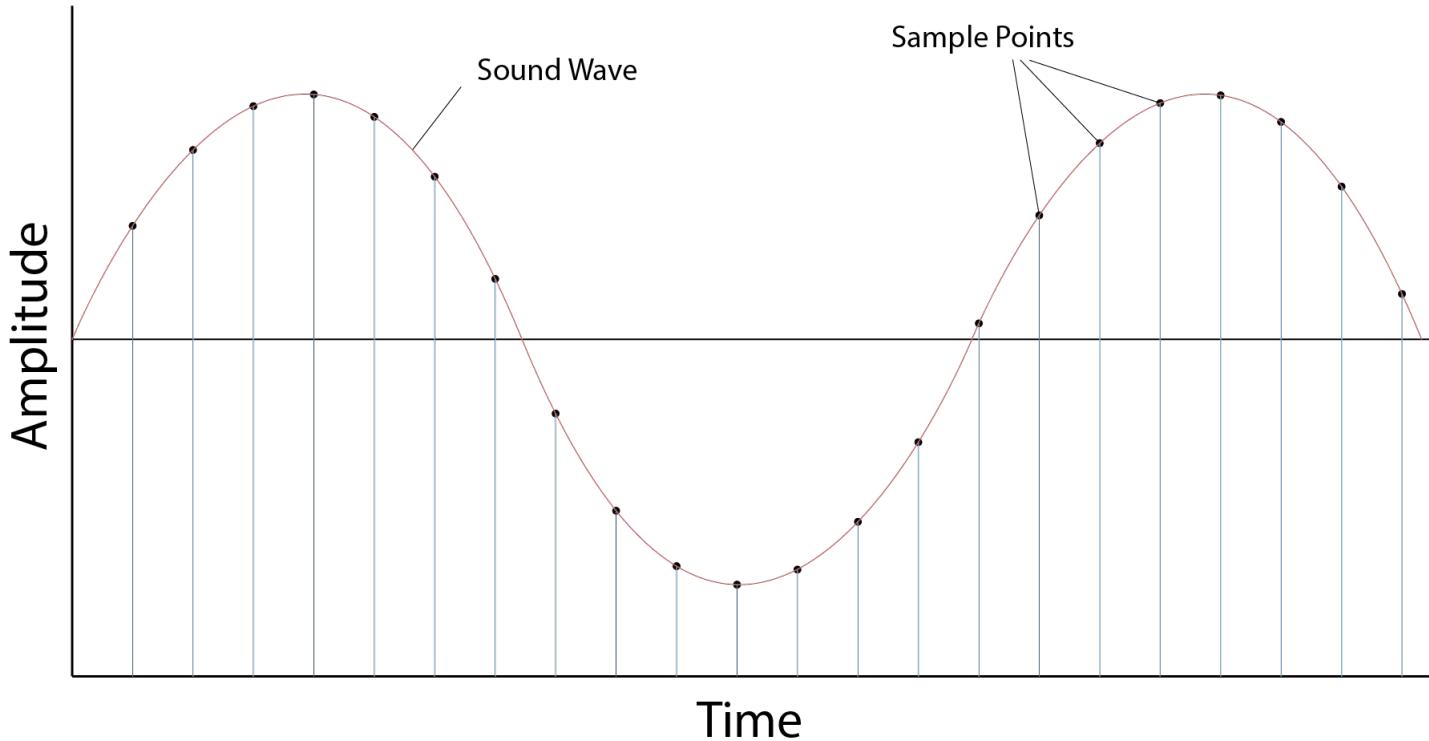
---

- Sampling
- Quantization

# Pulse-code modulation

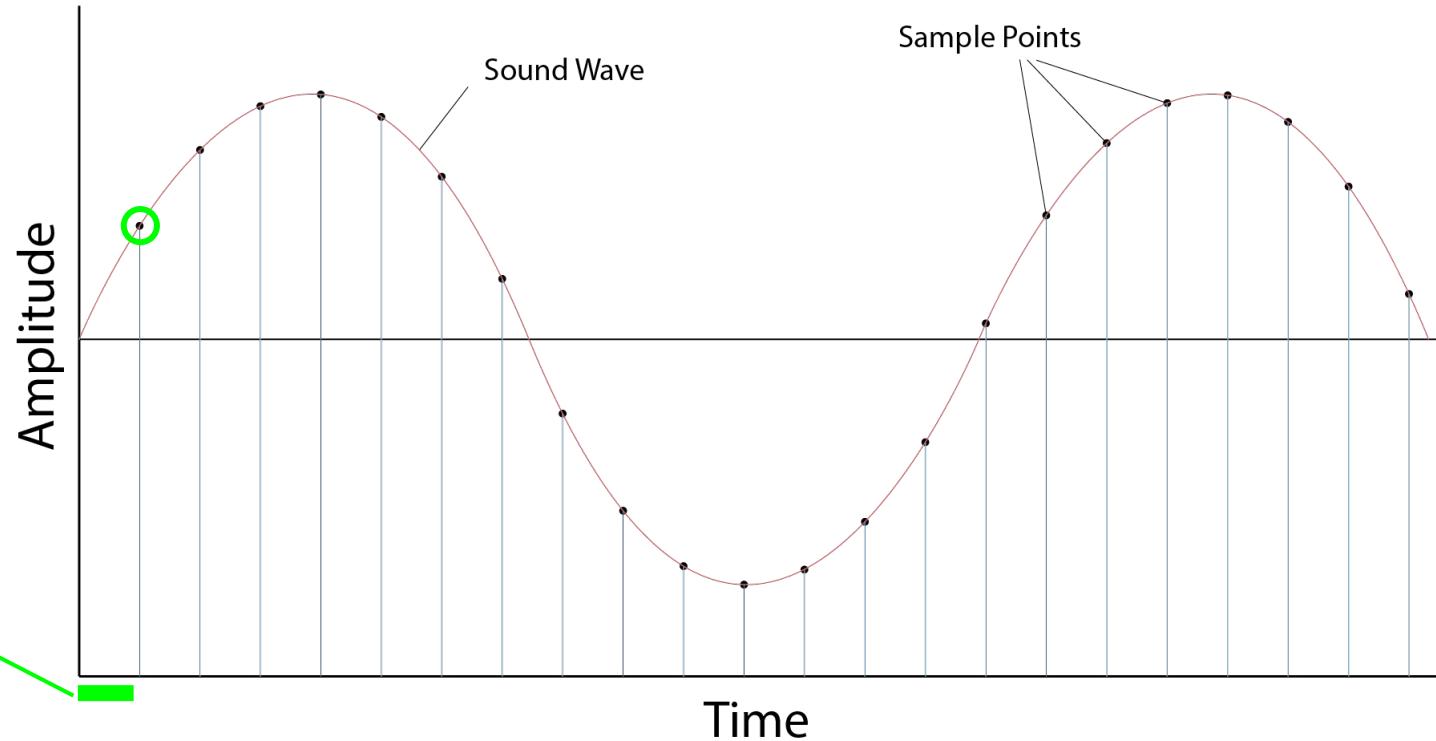
# Sampling

---



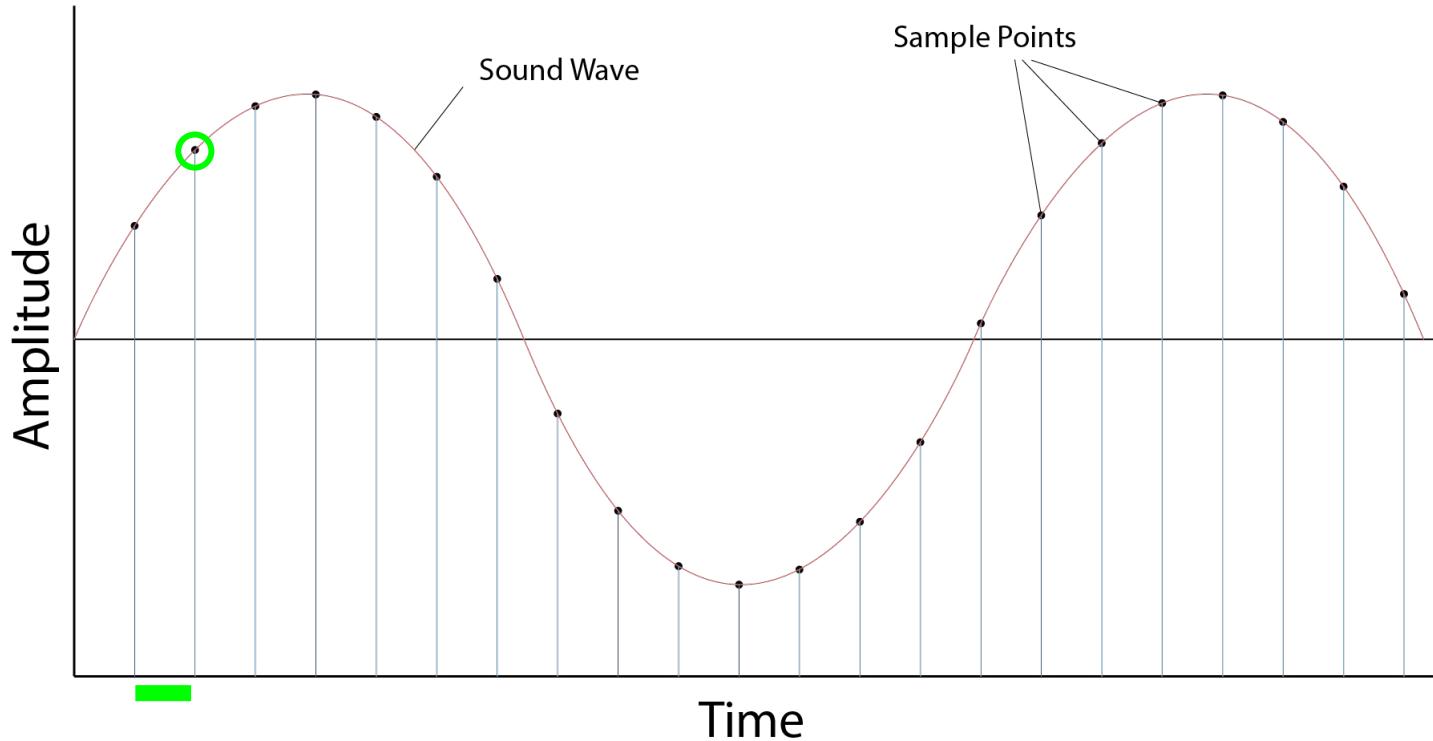
# Sampling period

---



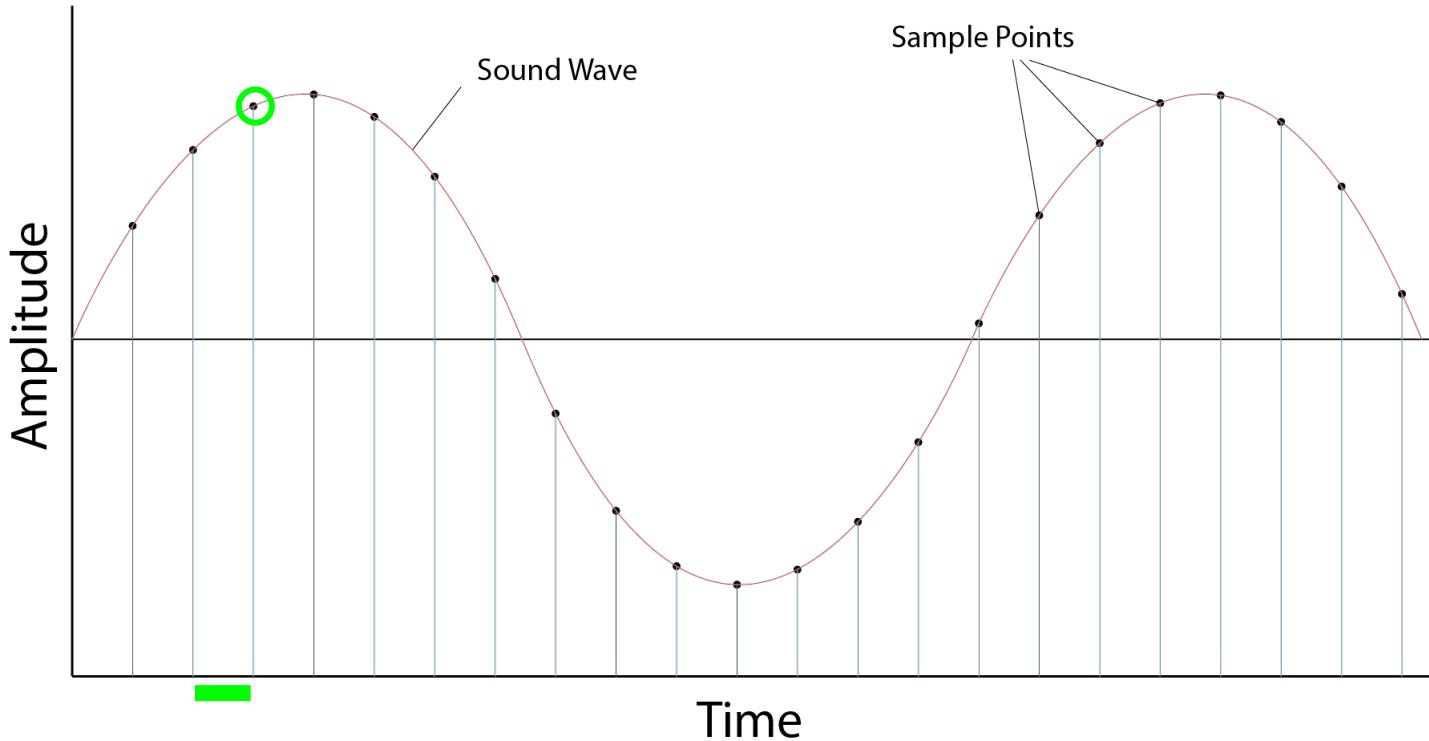
# Sampling period

---



# Sampling period

---



## Locating samples

---

$$t_n = n \cdot T$$

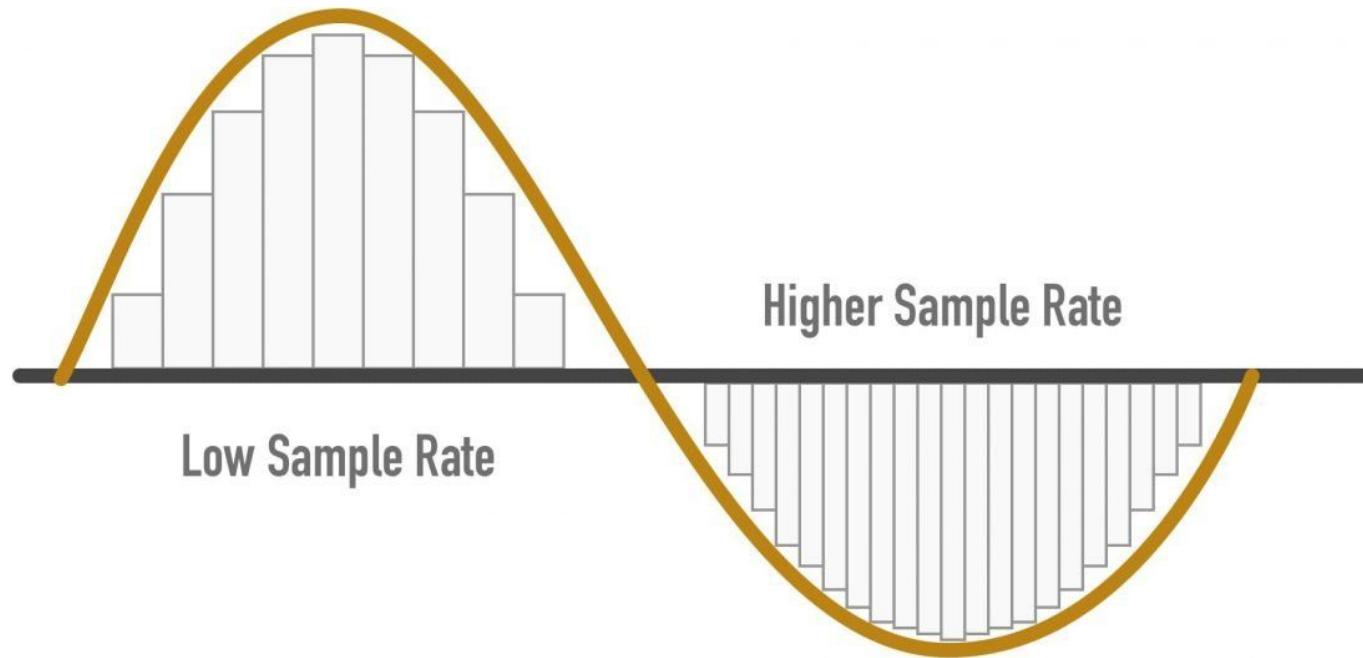
## Sampling rate

---

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

# Sampling rate

---



# Why sampling rate = 44100hz?

---



# Nyquist frequency

---

$$f_N = \frac{s_r}{2}$$

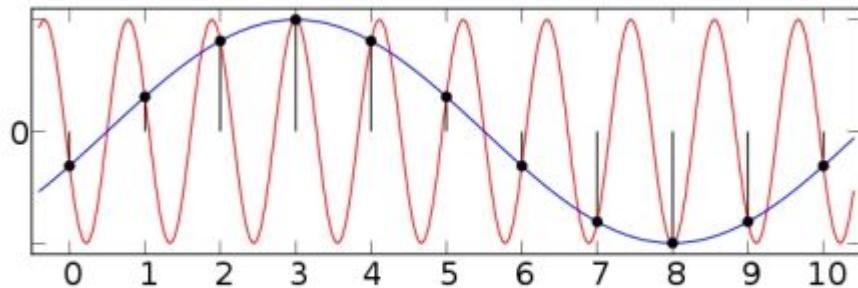
## Nyquist frequency for CD

---

$$f_N = \frac{44100}{2} = 22050$$

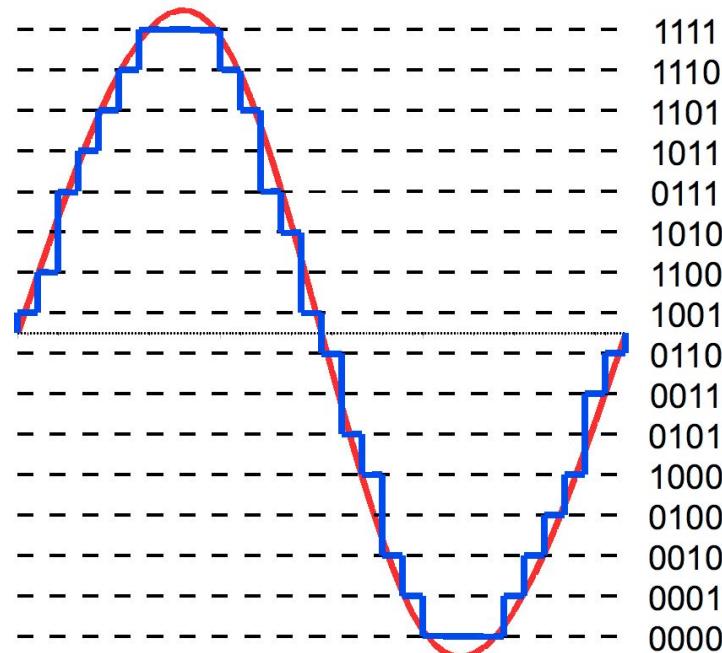
# Aliasing

---



# Quantization

---



# Quantization

---

- Resolution = num. of bits

# Quantization

---

- Resolution = num. of bits
- Bit depth

# Quantization

---

- Resolution = num. of bits
- Bit depth
- CD resolution = 16 bits



# Quantization

---

- Resolution = num. of bits
- Bit depth
- CD resolution = 16 bits

$$2^{16} = 65536$$



# Memory for 1' of sound

---

- Sampling rate = 44100 Hz
- Bit depth = 16 bits

# Memory for 1' of sound

---

- Sampling rate = 44100 Hz
- Bit depth = 16 bits

$$(((16 \cdot 44,100)/1,048,576)/8) \cdot 60$$

# Memory for 1' of sound

---

- Sampling rate = 44100 Hz
- Bit depth = 16 bits

$$(((16 \cdot 44,100)/1,048,576)/8) \cdot 60$$

# Memory for 1' of sound

---

- Sampling rate = 44100 Hz
- Bit depth = 16 bits

$$(((16 \cdot 44,100) / 1,048,576) / 8) \cdot 60$$

# Memory for 1' of sound

---

- Sampling rate = 44100 Hz
- Bit depth = 16 bits

$$(((16 \cdot 44,100) / 1,048,576) / 8) \cdot 60$$

# Memory for 1' of sound

---

- Sampling rate = 44100 Hz
- Bit depth = 16 bits

$$(((16 \cdot 44,100)/1,048,576)/8) \cdot 60$$

# Memory for 1' of sound

---

- Sampling rate = 44100 Hz
- Bit depth = 16 bits

$$(((16 \cdot 44,100)/1,048,576)/8) \cdot 60$$

# Memory for 1' of sound

---

- Sampling rate = 44100 Hz
- Bit depth = 16 bits

$$(((16 \cdot 44,100)/1,048,576)/8) \cdot 60 = 5.05MB$$

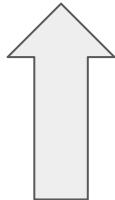
# Dynamic range

---

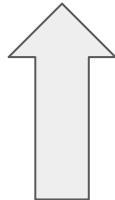
- Difference between largest/smallest signal a system can record

# Dynamic range

---



resolution



dynamic range

# Signal-to-quantization-noise ratio

---

- Relationship between max signal strength and quantization error
- Correlates with dynamic range

# Signal-to-quantization-noise ratio

---

- Relationship between max signal strength and quantization error
- Correlates with dynamic range

$$SQNR \approx 6.02 \cdot Q$$

# Signal-to-quantization-noise ratio

---

- Relationship between max signal strength and quantization error
- Correlates with dynamic range

$$SQNR(16) \approx 96dB$$

# How do we record sound?

---

# How do we record sound?

---



# How do we record sound?

---

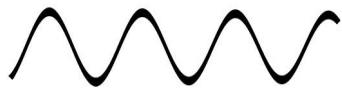


ADC



# How do we record sound?

---



ADC

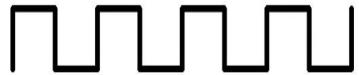


# How do we reproduce sound?

---

# How do we reproduce sound?

---



# How do we reproduce sound?

---



DAC



# How do we reproduce sound?

---



DAC



# What's up next?

---

- Overview of audio features

# Join the community!

---



[thesoundofai.slack.com](https://thesoundofai.slack.com)