COMP47700
Speech and Audio

Dr Alessandro Ragano

Office: 3rd Floor, Science East, Insight Centre
School of Computer Science
University College Dublin

alessandro.ragano@ucd.ie

1.3
Digital Representation of Sound

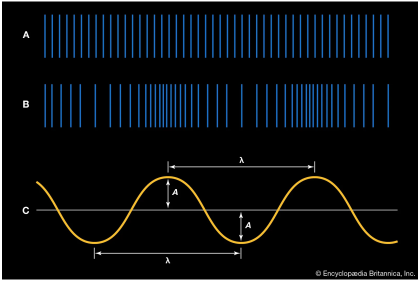
- Sound
  - Amplitude
  - Frequency
- Analogue to Digital Conversion
  - Time Domain Sampling
  - Quantisation

Sound

Sound: A Physical Phenomenon

Sound is a perception of a pressure wave[a] whether spoken, audio or music. The vibrations of molecules of the media cause it. The wave characteristics are determined by the rate of change in vibration (frequency) and the pressure variation (amplitude).

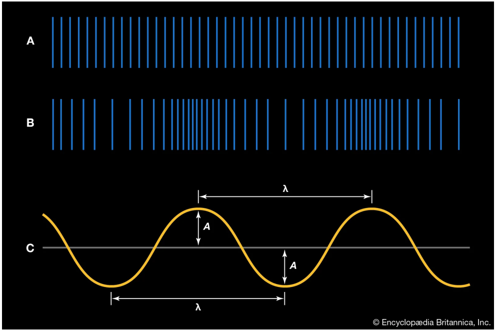
[a] longitudinal in air and transverse some media



Sound Properties

Amplitude

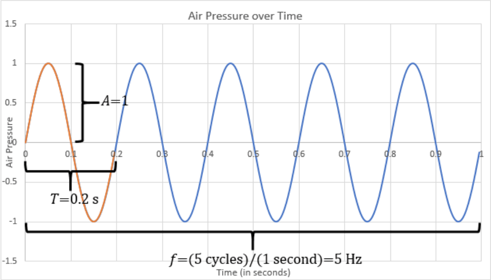
Amplitude describes the magnitude of change in air pressure within a sound wave. It represents the displacement from the equilibrium point, which is the normal atmospheric pressure. Amplitude is related to the perceived loudness: a larger amplitude typically results in a louder sound.



Sound Properties

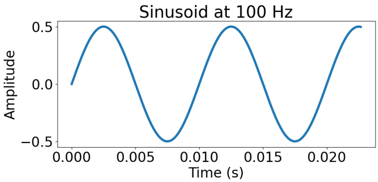
Frequency

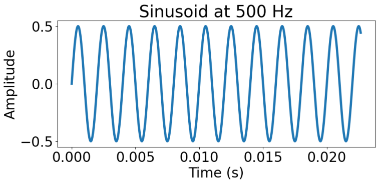
Frequency is the rate at which something occurs over a particular period of time. Frequency is measured in Hertz (Hz). It indicates how many oscillations/cycles occur in one second.

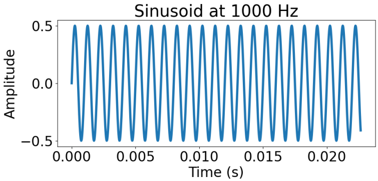


Source: vru.vibrationresearch.com

Sine Waves


Sinusoid at 100 Hz
Amplitude
0.5
0.0
−0.5
0.000
0.005
0.010
0.015
0.020
Time (s)

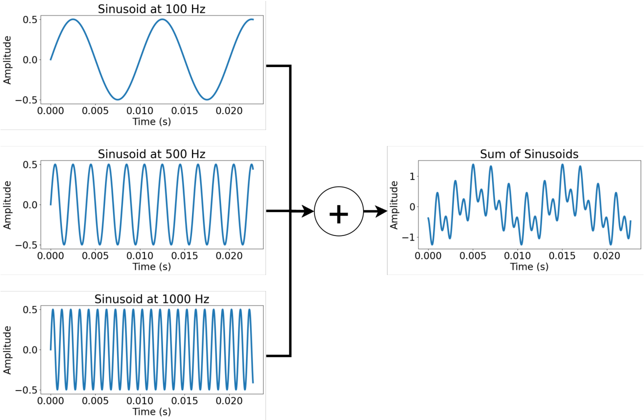

Sinusoid at 500 Hz
Amplitude
0.5
0.0
−0.5
0.000
0.005
0.010
0.015
0.020
Time (s)


Sinusoid at 1000 Hz
Amplitude
0.5
0.0
−0.5
0.000
0.005
0.010
0.015
0.020
Time (s)

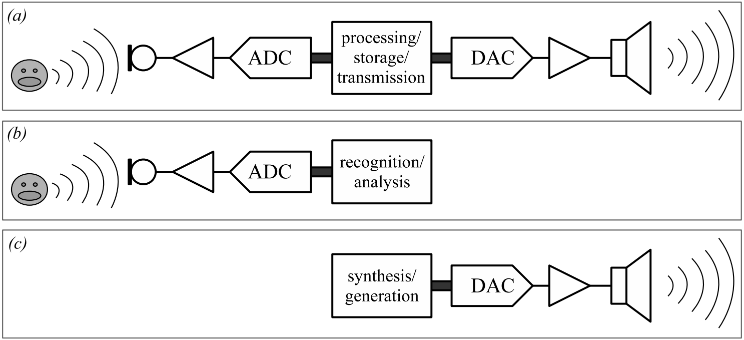
Real-world Sounds

Complex Waves

Real-world sounds e.g., speech, are the sum of sine waves at different amplitudes and frequencies. The average human auditory system can hear frequencies between 20 Hz and 20 kHz.



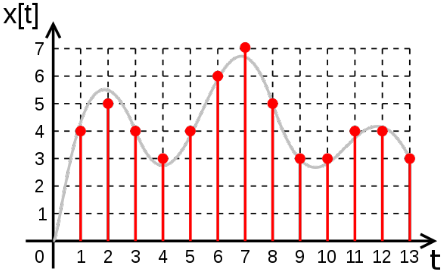
Analogue to Digital Conversion



Source: McLaughlin (2009, p.3)

Digital Signal

Discrete signal mathematical notation: $x(t) \rightarrow x[t]$



- **ADC** transforms a continuous analogue signal (e.g., audio waveform) into a digital representation.
  1. **Sampling:** Capturing the signal's amplitude at regular time intervals.
  2. **Quantisation:** Assigning each sampled signal's amplitude to a discrete level.

Source: https://en.wikipedia.org/wiki/Digital_signal#/media/File:Digital.signal.discret.svg

Time Domain Sampling

Sampling

Extract values at regularly spaced intervals of time.
Sampling period T – duration of time between samples.
Sampling frequency f_s – how often samples are taken.
$f_s = \frac{1}{T} \Rightarrow f_s$ is measured in Hertz indicating the number of samples extracted in 1 second.

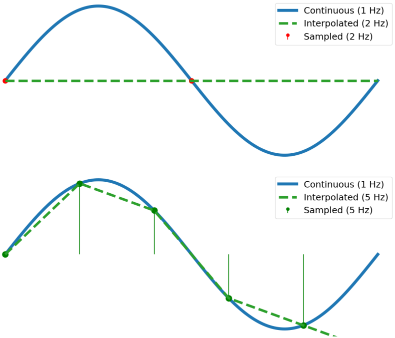
Example with $T = 3$

$t = 0T = 0,$
$t = 1T = 3,$
$t = 2T = 6,$
...
Continuous signal $\rightarrow [10, 0, 5, 20, 10, 0, 15, 80]$
Sampled signal $\rightarrow [10, 0, 5, 20, 10, 0, 15, 80] \rightarrow [10, 20, 15]$

Time Domain Sampling

Why Sample? Why Not Use All Data Points?

- **Storage and Processing** Continuous signals would require infinite storage and processing power, which is impossible.
- **Redundancy** Many signals contain redundant information; not all data points are necessary for accurate reconstruction.



Time Domain Sampling

The Nyquist-Shannon Theorem

- **Key Idea:** We can perfectly reconstruct a signal from its samples if we sample fast enough.
- For a signal with maximum frequency (bandwidth) B, the minimum sampling rate:

$$f_s > 2B \quad \text{(Nyquist Rate)}$$

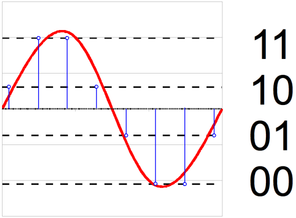
- **Example: Human Hearing**
  - Audible range: up to 20 kHz ($B = 20\,\text{kHz}$)
  - Minimum sampling rate: $f_s > 40\,\text{kHz}$
- **Common Sampling Rates:**
  - Music, CDs, streaming, video calling: 44.1 kHz, 48 kHz (covers human hearing and provides some margin)
  - Older telephone systems: 8 kHz, 16 kHz for speech intelligibility

Quantisation

Why Quantise?

- Computers use discrete values (bits).
- Audio is continuous (analogue) $\rightarrow$ Must be digitised.
- **Quantisation:** Converting continuous amplitudes to discrete levels.
- Each sample is rounded to the nearest level based on bit-depth. Example: 16-bit audio = 65,536 levels.

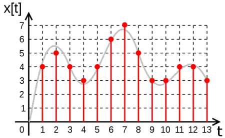
Example with *bit depth* $= 2$

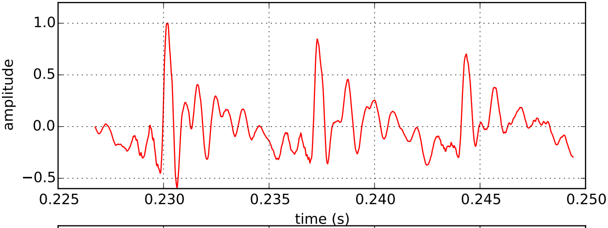


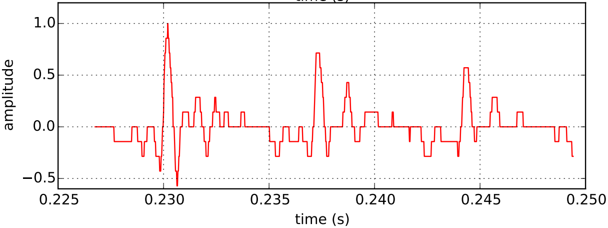
source: Quantization (signal processing). (2025, January 16). In Wikipedia. https://en.wikipedia.org/wiki/Quantization($_{signal_p rocessing}$)

Quantisation Example

From a bit depth of 16 ($2^{16} = 65,536$) to a bit depth of 4.







Quantisation Error

Dynamic Range (DR) can describe the ratio of a signal's maximum amplitude to the noise amplitude of a playback device (e.g. loudspeaker). It is also used to evaluate quantisation noise, i.e. the largest to smallest signal that can be represented.

Quantisation Error/Quantisation Noise

Quantisation clearly introduces an error into the signal.

The Signal-to-Quantisation-Noise is $\approx 96\ dB$ for 16 bit audio calculated as: $\mathrm{SQNR} = 20\log_{10}(2^Q) \approx 6.02 \cdot Q\ \mathrm{dB}$

Bit depth and dynamic range

Every bit adds $6dB$.

Human Dynamic Range

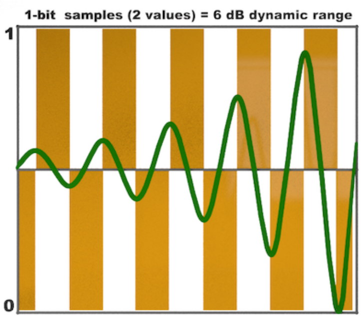
Sound level/ dB SPL	Power ratio	Amplitude ratio	Typical example
140	10^{14}	10^{7}	Gunshot at close range
120	10^{12}	10^{6}	Loud rock group
100	10^{10}	10^{5}	Shouting at close range
80	10^{8}	10^{4}	Busy street
70	10^{7}	3160	Normal conversation
50	10^{5}	316	Quiet conversation
30	10^{3}	31.6	Soft whisper
20	10^{2}	10	Country area at night
6.5	4.5	2.1	Mean absolute threshold @ 1kHz
3	1	1	Reference level

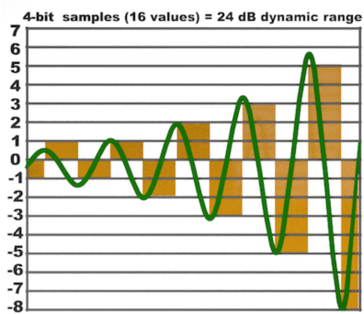
Source: Moore, B.C.J., An Introduction to the Psychology of Hearing, 2ed.

Bit Depth and Dynamic Range

Bit Depth and Dynamic Range

Each bit can represent a difference of 6 dB in dynamic range.



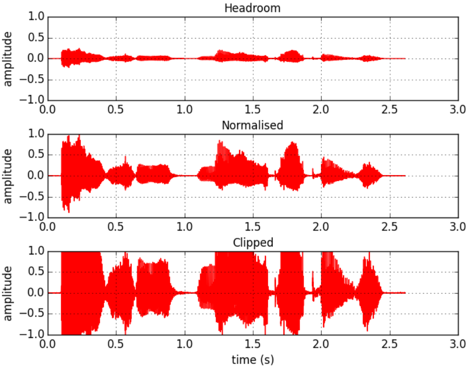


Source: https://cmtext.indiana.edu/digital_audio/chapter5_quantize.php

Dynamic Range: Headroom and Clipping

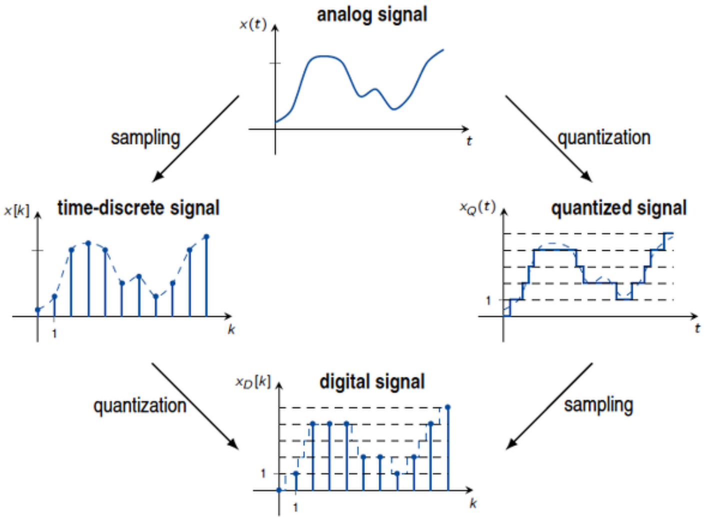
Headroom is the available dynamic range before the signal reaches the maximum limit of the digital system.
Clipping is what occurs when you run out of headroom.

Example with 32-bit floating point representation.



Analog, discrete and digital signals

Connections between different domains



source: https://dsp-nbsphinx.readthedocs.io/en/nbsphinx-experiment/quantization/introduction.html