Multimedia Systems

II. Introduction to Sound

2.3. Digital Audio

António Sá Pinto
FEUP

Agenda

- Auditory Specifications
- Auditory Masking
- Pulse Code Modulation
- Exploring the Auditory System
- Audio Coding and Formats
- Assessment of Audio Quality
- MIDI and 3DAudio

Auditory Specifications

Frequency range

20Hz to 20,000Hz (textbook); 50Hz to 13,000 Hz (effective); most psychophysics is done between 100-5000Hz (range where one obtains interpretable data).

Intensity range

Extends over many orders of magnitude (depending on frequency); at the sweet spot (~1000Hz-3000Hz) about a 120dB dynamic range.

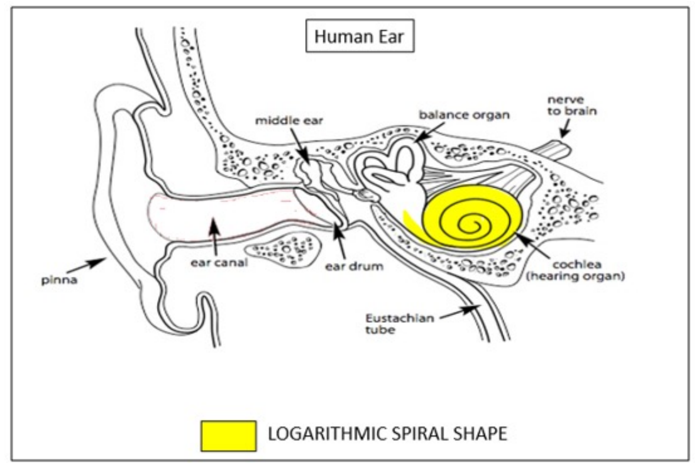
Sensitivity

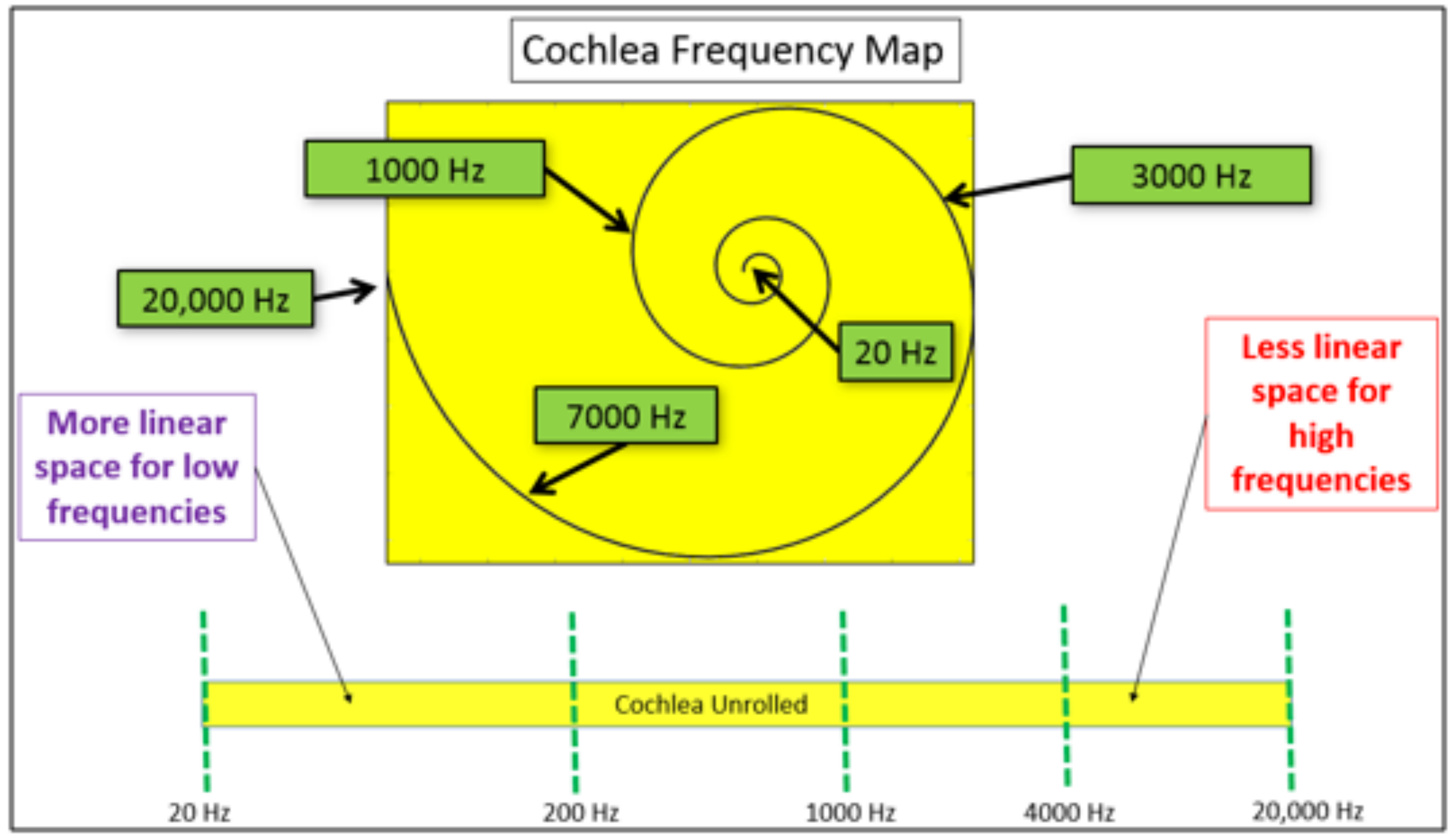
Just Noticeable Difference (JND) for frequency: ~0.2%
JNDs for loudness discrimination: ~1dB
JNDs for Timing differences:

- few microseconds in spatial hearing (JND for azimuthal ~1deg)
- 2 msec (gap thresholds);
- 25 miliseconds (order threshold);

Auditory Specifications

Critical Bands (II)





from: http://hyperphysics.phy-astr.gsu.edu/hbase/Sound

[See the Hudspeth Cochlear Animation]()

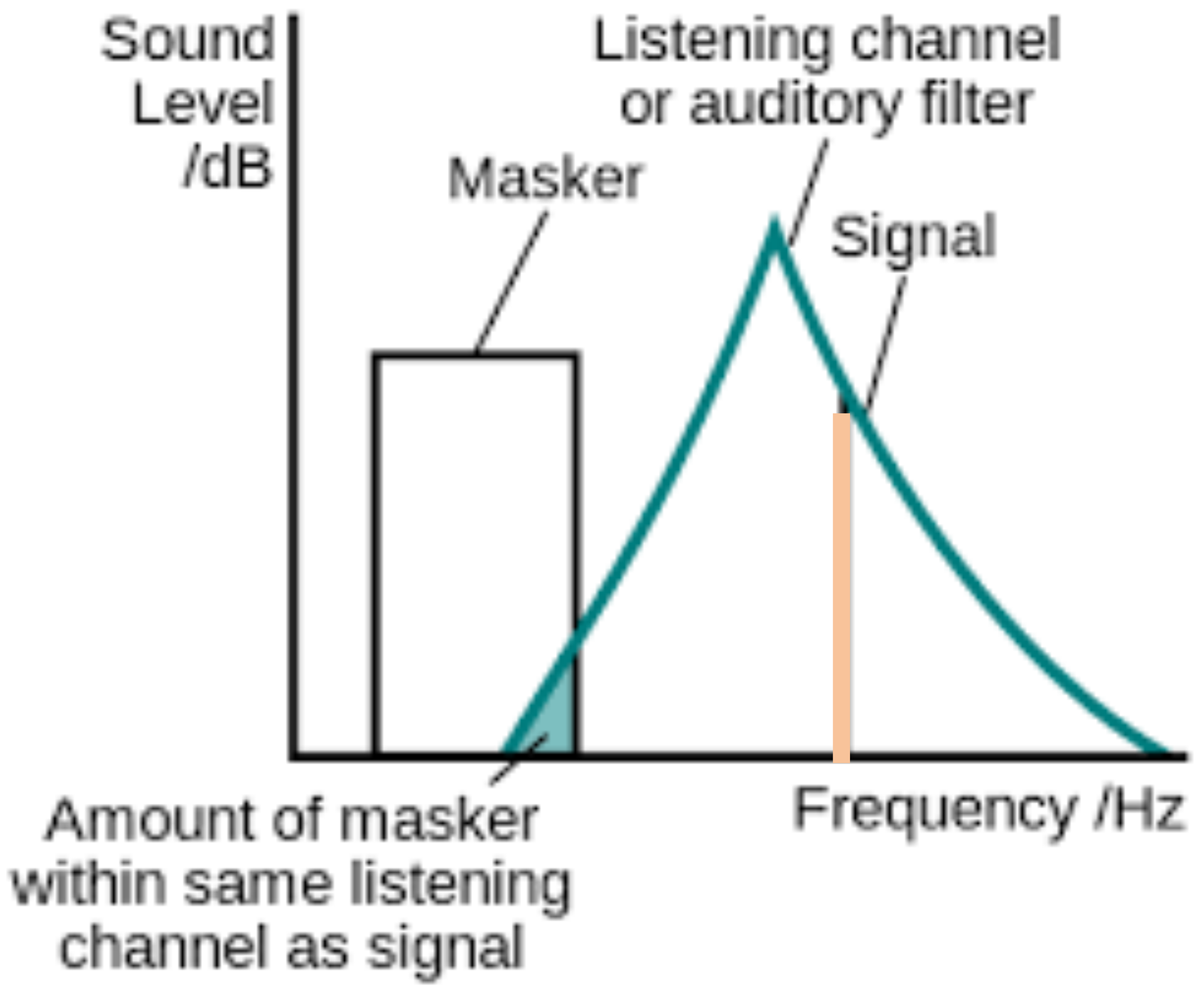
Critical Band (Bark)	Center Frequency (Hz)	Bandwidth (Hz)
1	50	100
2	150	100
3	250	100
4	350	100
5	450	110
6	570	120
7	700	140
8	840	150
9	1000	160
10	1170	190
11	1370	210
12	1600	240
13	1850	280
14	2150	320
15	2500	380
16	2900	450
17	3400	550
18	4000	700
19	4800	900
20	5800	1100
21	7000	1300
22	8500	1800
23	10500	2500
24	13500	3500

Critical Bands and the Bark Scale

from: https://community.sw.siemens.com/s/article/critical-bands-in-human-hearing

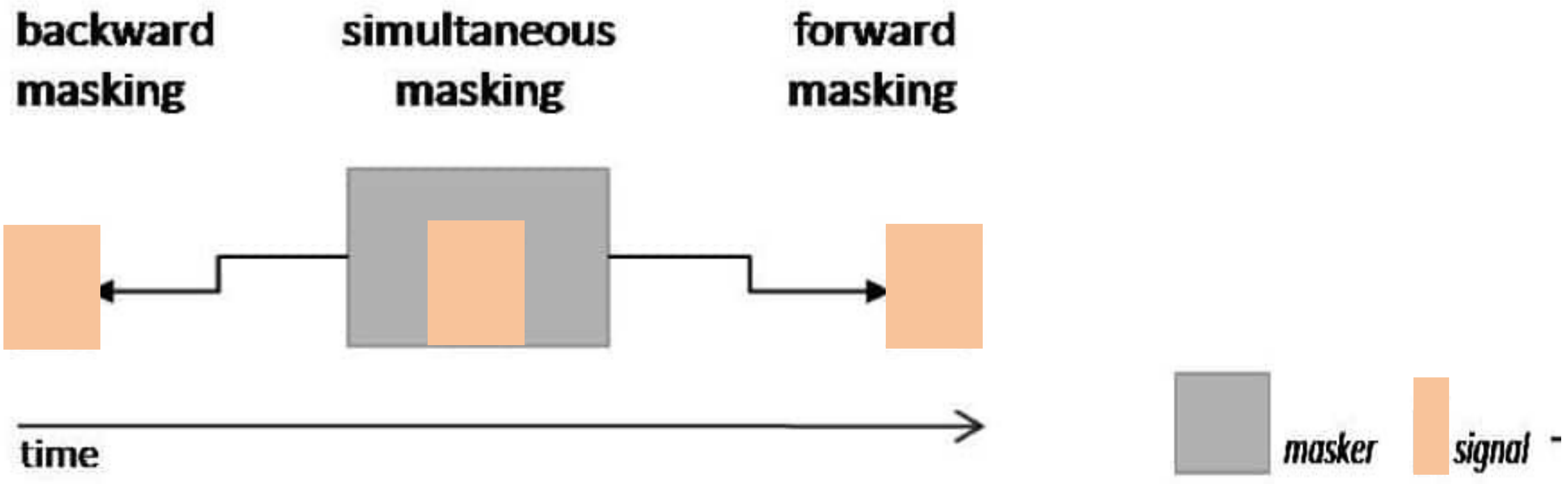
Auditory Masking

Spectral

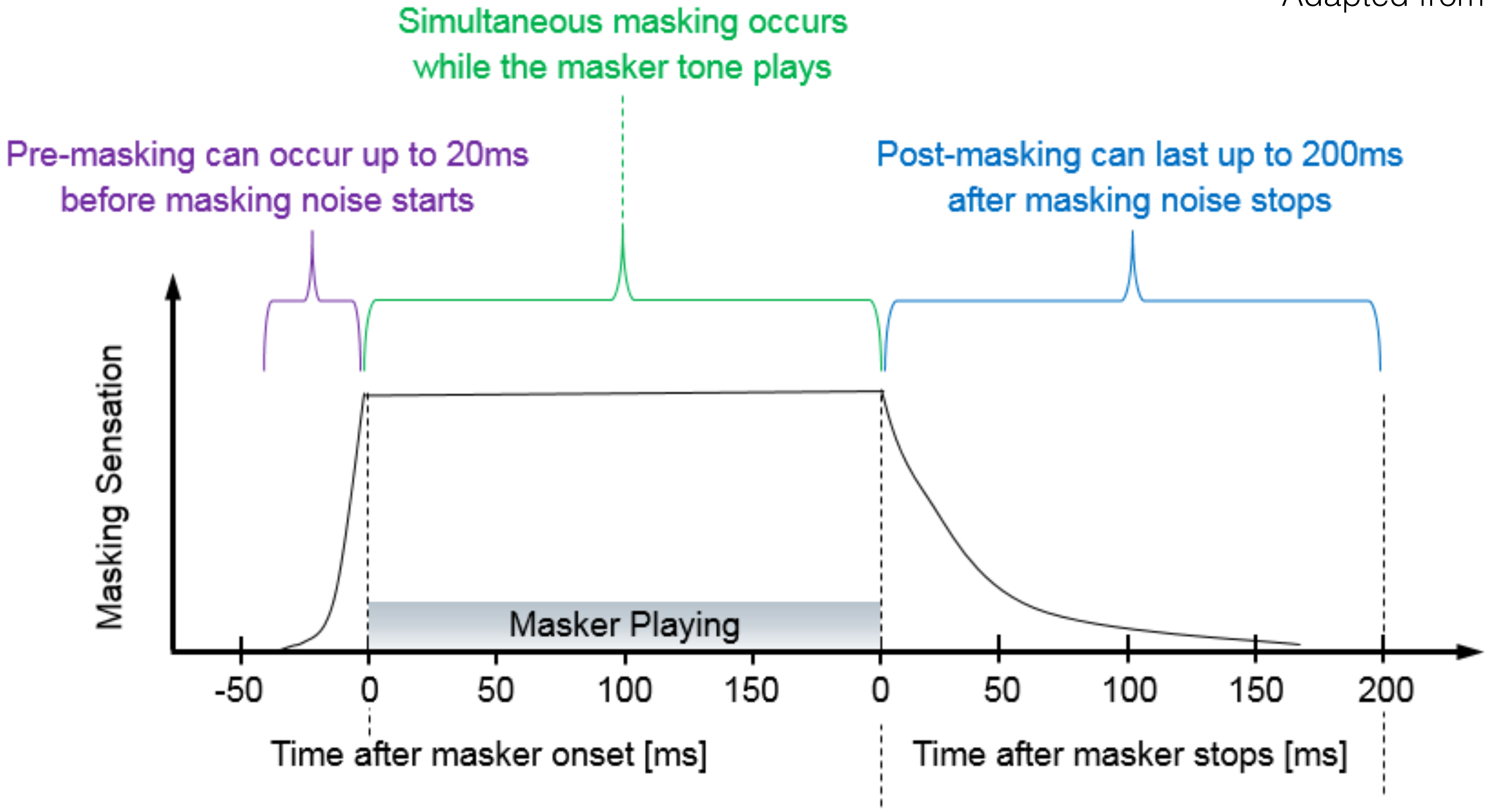


Adapted from Wikipedia

Temporal



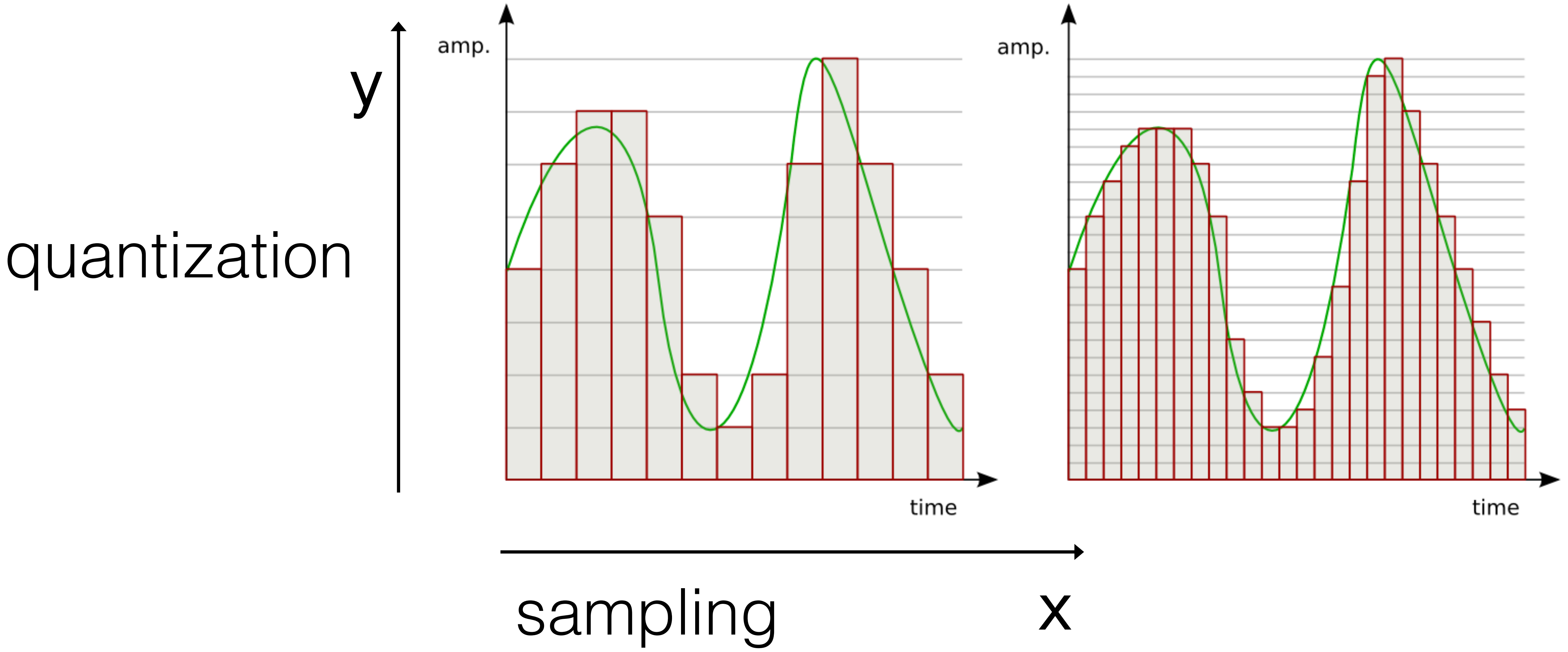
Adapted from Wikipedia



A/D Conversion

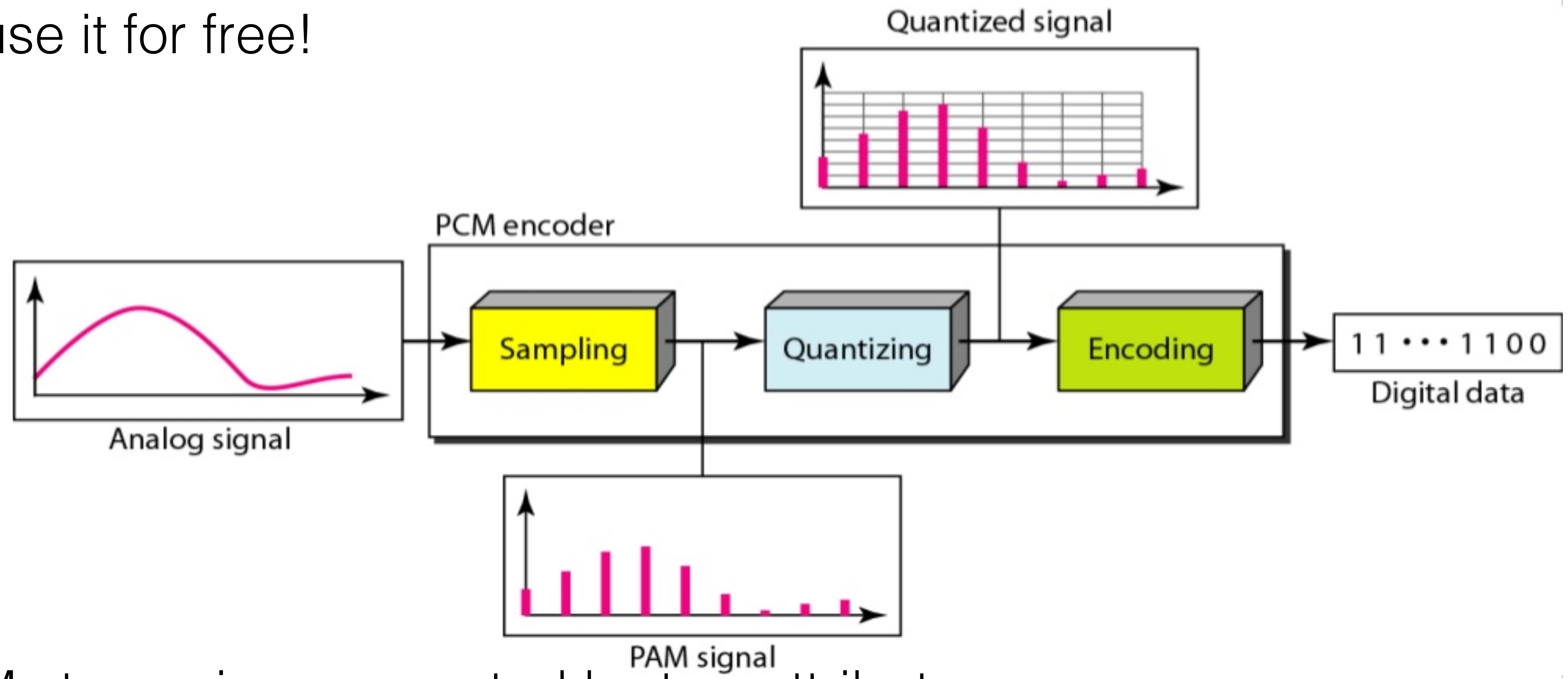
Analog-to-Digital conversion is used to convert the analogue audio signal (a time varying electrical voltage, say, the output of a microphone), into a series of 'samples' which are 'snapshots' of the analogue signal taken at periodic intervals (known as the sampling period).

The two main steps in digitization of sound are sampling and quantization.



Pulse Code Modulation

Pulse Code Modulation (PCM) is a method used to digitally represent sampled analog signals. It is the industry standard for storing analog waves in a digital format. PCM is non-proprietary so anyone can use it for free!

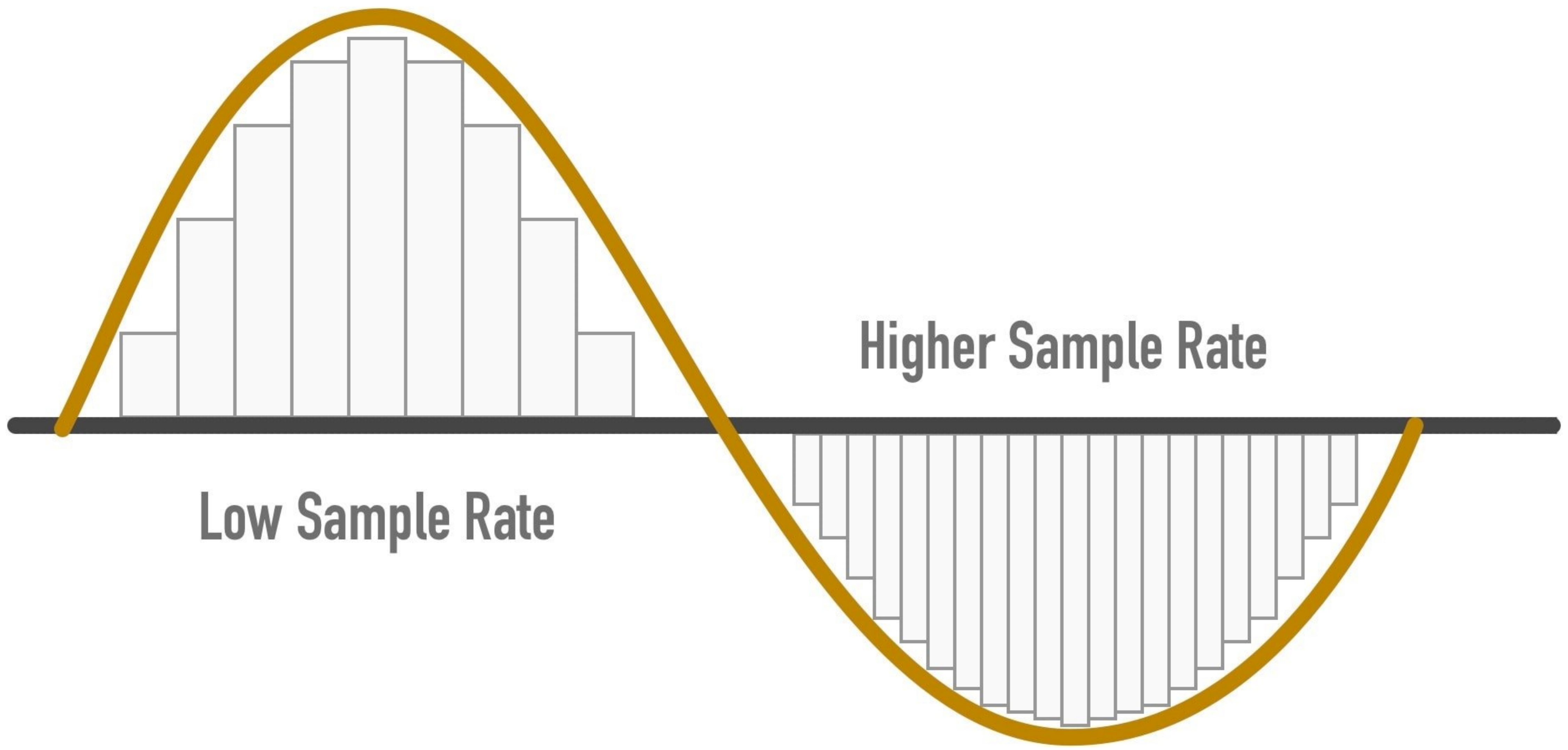


The fidelity/quality of a PCM stream is represented by two attributes:

- Sample Rate
- Bit Depth

Sample Rate

In a typical digital audio CD recording, the sampling rate is 44,100 or 44.1kHz.
Why this sampling rate? (Nyquist-Shannon sampling theorem)
A sampling rate of 44,100 samples per second or 44.1kHz allows for accurate reproduction of frequencies around about 22kHz.



Other examples of common sampling rates are 8KHz in telephones and between 96KHz to 192KHz for Blu-ray audio tracks.

Bit depth

The bit depth determines how much information can be stored, and with what resolution.

A sampling with 24-bit depth can store more nuances and hence, more precise than a sampling with 16-bit depth:

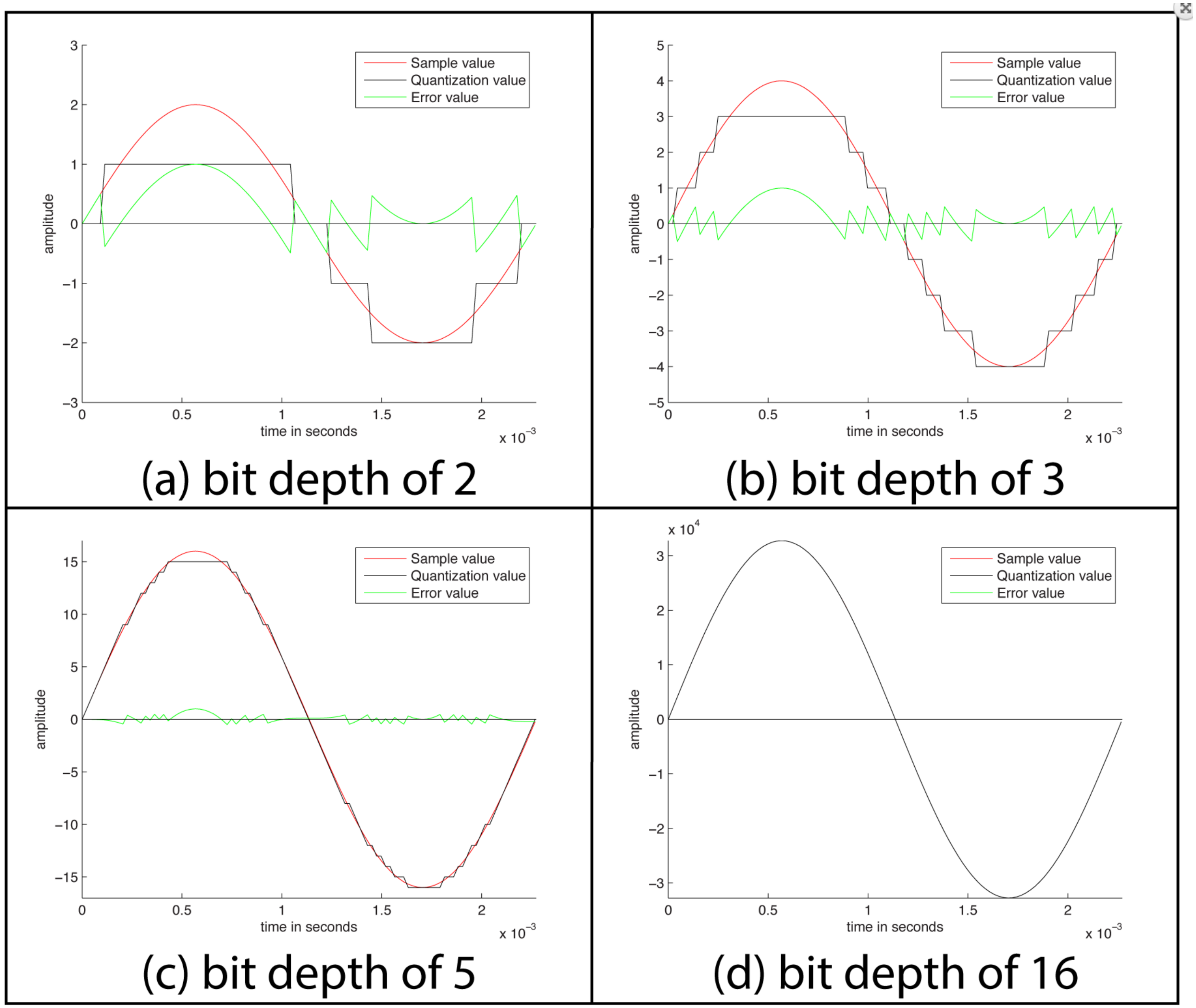
- 16-bit: We are able to store up to 2^{16}=65,536 levels of information, a dynamic range of 96dB
- 24-bit: We are able to store up to 2^{24}=16,777,216 levels of information, a dynamic range of 144dB.

Given a bit depth of n, the dynamic range of a digital audio recording is equal to

$$20\log_{10}\left(\frac{2^{n-1}}{1/2}\right)dB$$

Bit depth

Quantisation



Audio Coding

Key-Ideas

In perceptual audio coding, the main concerns in audio signals representation are:

- removal of **Redundancy**;
- removal of **Irrelevancy**;

Adapted from: Marina Bosi Perceptual Audio Coding, IEEE Signal Processing Society, 2015

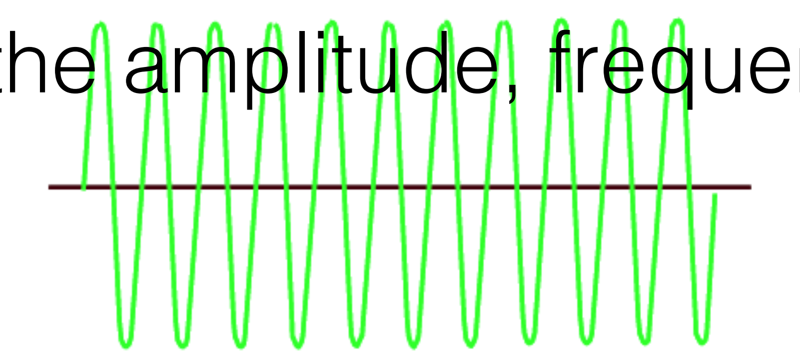
Audio Coding

Redundancy

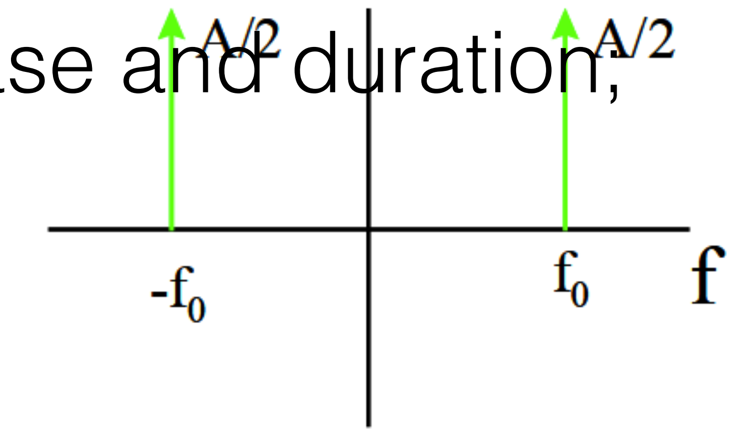
In audio coding, redundant means that the same information can be represented with fewer bits.

Example: represent a sine wave signal in 2 different representations:

- **Redundant**: sample the waveform 44.100 times per second and describe each sample with 16 bits;
- **Concise**: Describe the amplitude, frequency, phase and duration;



x(t)



X(f)

The concise representation of the sine wave is basically equivalent to the information in the Fourier transform;

Since music and many other audio signals are very tonal, most coders work in the frequency domain to reduce redundancy;

Adapted from: Marina Bosi Perceptual Audio Coding, IEEE Signal Processing Society, 2015

Audio Coding

Irrelevancy

In audio coding, irrelevant data means that you can't hear any difference in the audio signal if those data are omitted.

The main causes of irrelevancy:

- Hearing threshold;
- Masking;

Exploiting Irrelevancy:

- Don't code signal components you can't hear;
- Only quantise audible signal components with enough bits to keep quantization noise below the level it can be heard;

Adapted from: Marina Bosi Perceptual Audio Coding, IEEE Signal Processing Society, 2015

Audio Coding

Lossless coding

- Based on statistical relation between symbols within the data
- Entropy coding such as Huffman coding, arithmetic coding etc.
- The original signals can be perfectly reconstructed.

Lossy coding

- Based on the perceptual modelling of audio signals (such as psychoacoustic models of hearing), some redundant information within audio signals can be removed without affecting their perceptual quality.
- The original signals cannot be perfectly reconstructed.

Audio File Formats

Format Types	Pros	Cons	Popular formats
Uncompressed	• No data loss • Compatible with older software	• Occupies excessive space	• WAV & AIFF: Store lossy/lossless formats • LCPM • BWF: British Wave Format (used by Tascam)
Compressed Lossless	• Reduces processing time • Retains data and good compression ratio	• Files are sizable compared to lossy format	• FLAC: Reduces files to 50-60% of original size and decompresses to an identical copy • ALAC: Apple-only .mp4
Compressed Lossy	• Small file size • Most popular consumer audio format	• Uses psychoacoustics to lop off "imperceptible" audio info	• MP3 • AAC • WMA • ATRAC

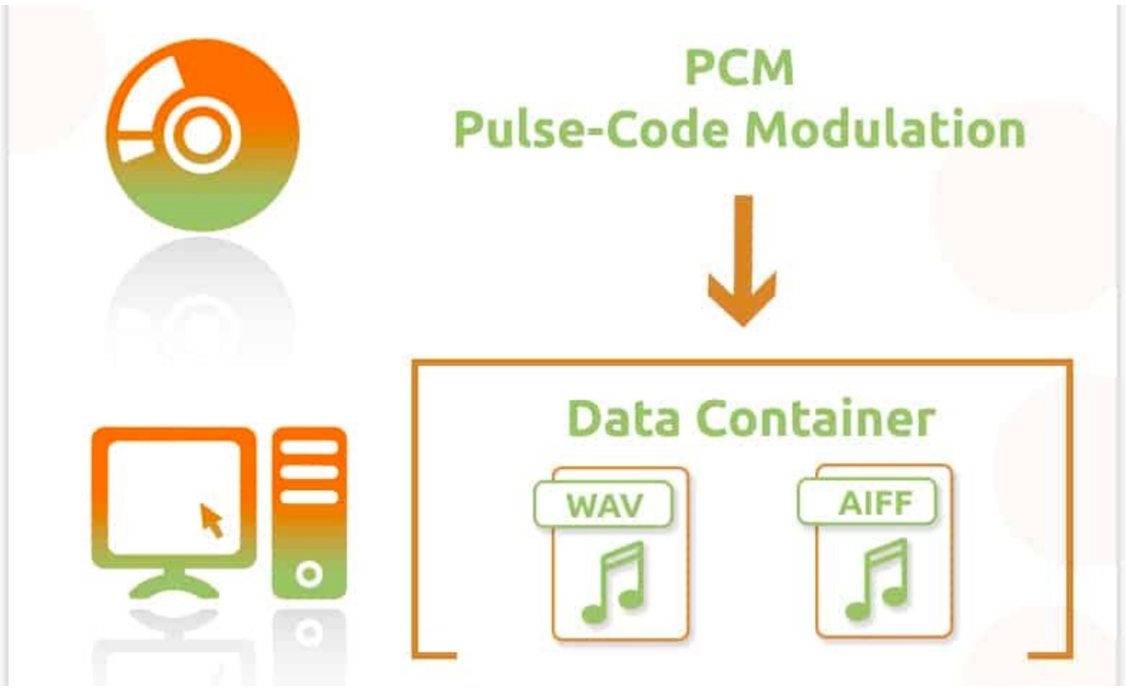
from: soundguys.com

Audio Coding (Codecs and File Formats)

- A Codec is an algorithm that performs the encoding and decoding of the raw audio data.
- Audio data itself is usually stored in a file with a specific audio file format.

- Most audio file formats support only one type of audio data (created with an audio coder), however there are multimedia digital container formats (as AVI) that may contain multiple types of audio and video data.

- A **digital container** format is a meta-file format where different types of data elements and metadata exist together in a computer file.
  - Formats exclusive to audio include, e.g., wav, xmf.
  - Formats that contain multiple types of data include, e.g. Ogg, MP4.



Audio Coding

History of Standards

1979 - the „Critical Band Coder“
1982 - „classic ATC“ for Music
1985 - MSC
1987 - OCF
1987 - MASCAM
1987 - PXFM
1990 - ASPEC, MUSICAM
1992 - MPEG 1
1996 - ePAC
1997 - MPEG 2 AAC
1999 - MPEG 4 AAC
2002 - HE AAC
Current standardization: MPEG surround, ALS, SLS

Timeline for near-CD-quality

1990 256 kbit/s ASPEC, MUSICAM
would fail today’s listening tests

1992 192 kbit/s MPEG-1 Layer-3

1994 128 kbit/s MPEG-1 Layer-3 (".mp3")
including combined joint stereo coding
bad quality for some signals

1997 96 kbit/s MPEG-2 Advanced Audio Coding
better than MP3 at 128, not fully transparent

2000 64 kbit/s AAC-based MPEG-4

2003 48 kbit/s MPEG-4 HeAAC (AAC+ in 2000)
e.g. used for XM Radio

From: Brandenburg Audio Coding Lecture @ Ilmenau University of Technology

Assessment of quality of coded audio

Objective Quality

- **Traditional objective measure**: The quality of audio is measured, ignoring psychoacoustic effects, through the use of objective performance indexes, such as *Signal to noise ratio* (SNR) and/or *Total block distortion* (TBD).
- **Perceptual objective measure**: The quality of audio is predicted based on a specific model of hearing.

Subjective Quality

- **Human listening tests**: When a highly accurate assessment is needed, formal listening tests will be required to judge the perceptual quality of audio.

Quality measurement of coded audio

Experiment of “13dB miracle” (Johnston and Brandenburg, 1991)

The original signal was injected with noise (with SNR signal-to-noise ratio of 13 dB) that was either:

a) White;
b) Shaped according to psychoacoustic masking models;

Demo

Original

a) White noise

b) Shaped noise

Sounds from: https://homes.esat.kuleuven.be/~compi/demos/

MIDI

(Protocol + Format)

Audio Coding

Human Auditory System

3DAudio

GA1 – MIDI and 3DAudio

- **Make a video presentation with 3 (maximum) slides explaining what is MIDI and another 3(max) slides for 3D Audio.**
- One of the presentations must be narrated by one student and the other presentation narrated by the other.
- The recommended time should be <u>5 mins maximum</u> for each presentation;
- The submission deadline is **22-October, 23:59**;

- Moodle: https://moodle.up.pt/mod/forum/view.php?id=78963