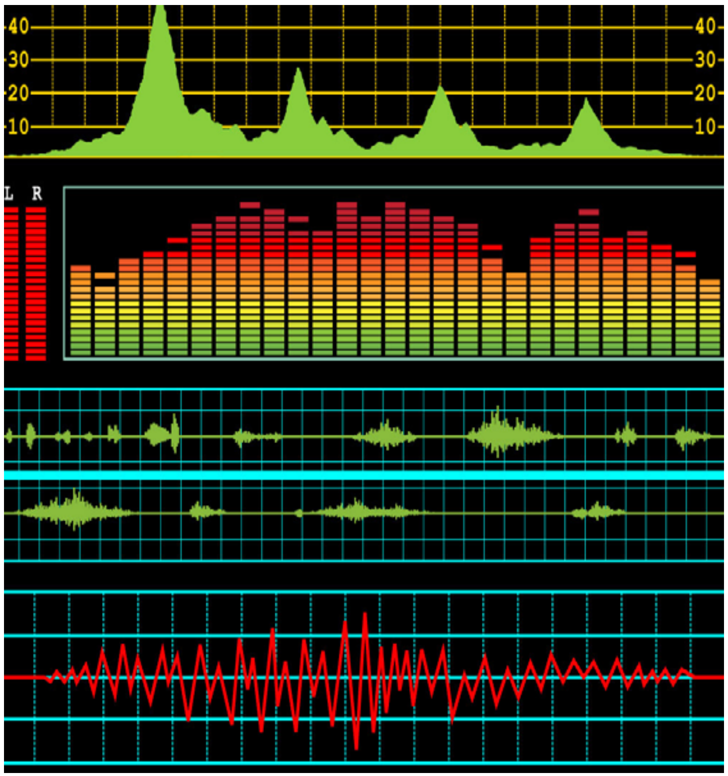
Basics of Digital Audio

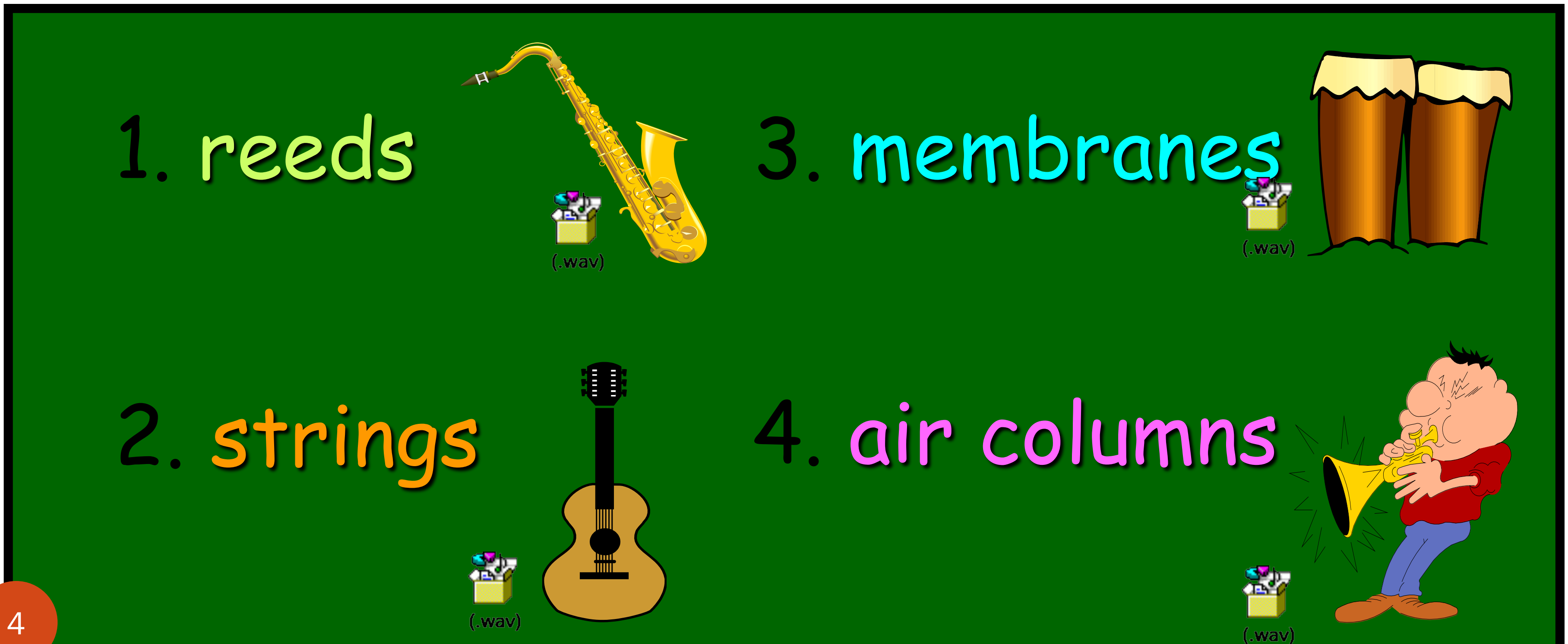
- The nature of sound
- Sound Digitalization
  - Sampling
  - Quantization
  - Compression
- Sound Processing
- Audio conversion
- File formats
- Further exploration

Sound



Sound waves

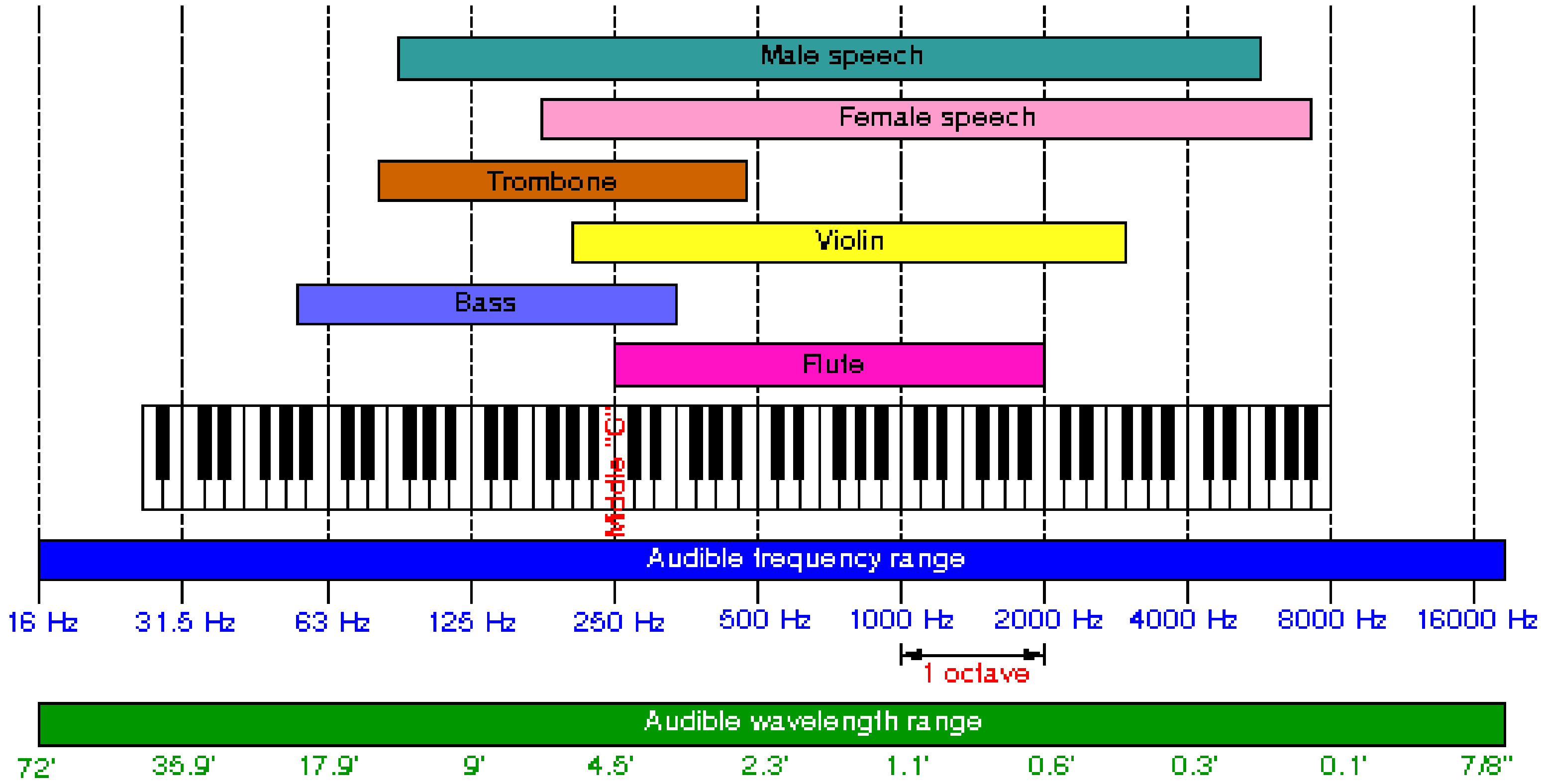
- Sound is a range of wave frequencies to which the human ear is sensitive, it will not travel through a vacuum.
- Sounds are produced by vibrating matter
- The audio spectrum extends from approximately *20 Hz to 20,000 Hz*.



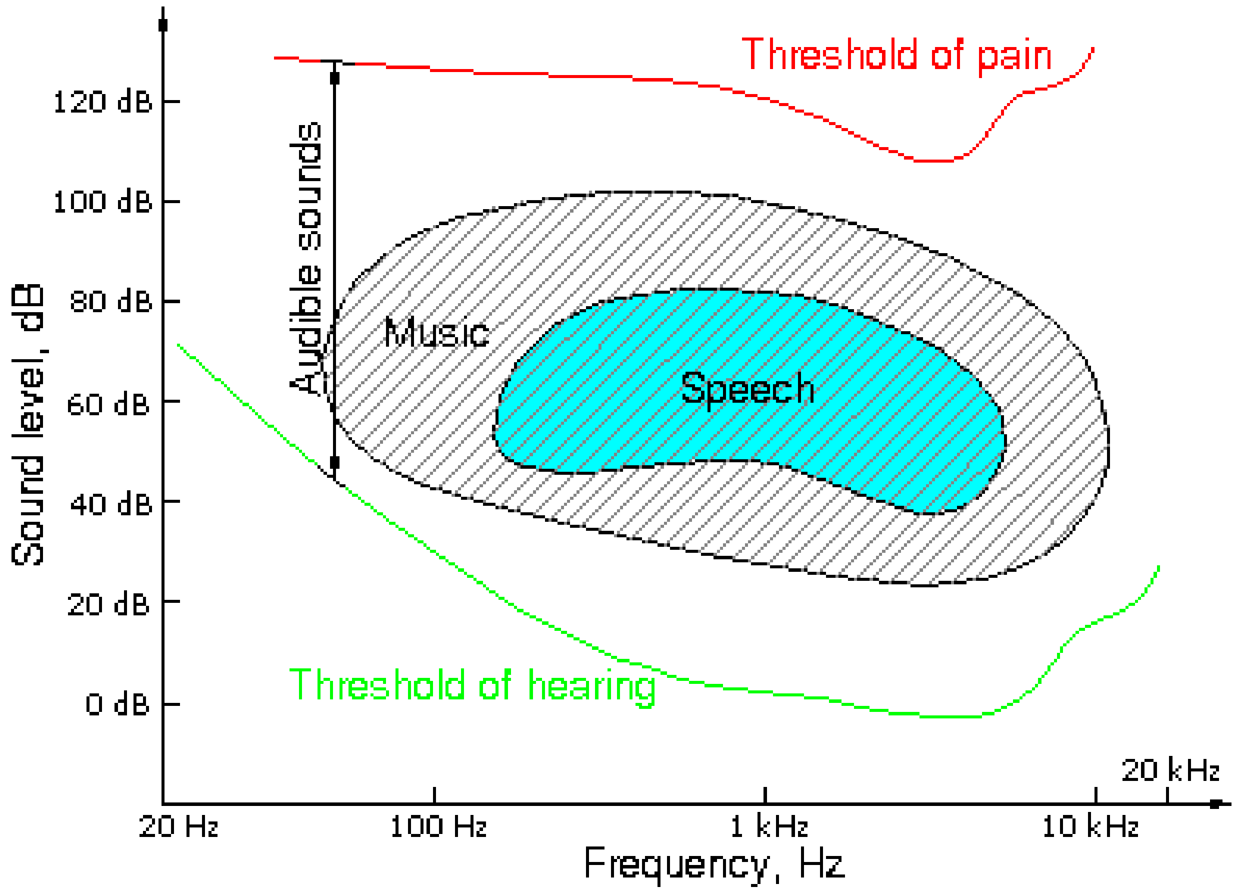
Sound waves

- Sounds are produced by vibrating matter.
  - eg. Speaker vibrates back and forth and produces longitudinal pressure wave.
- Sounds are transmitted via elastic medium like air.
  - No sound in space.
- Sound is a mechanical wave (longitudinal).
  - Sounds possess the characteristics and properties that are common to all waves
  - Sound waves posses a velocity, frequency, wavelength, phase, period and amplitude.
  - Sound waves reflect(bouncing), refract(changing angle) and diffract(bending around).

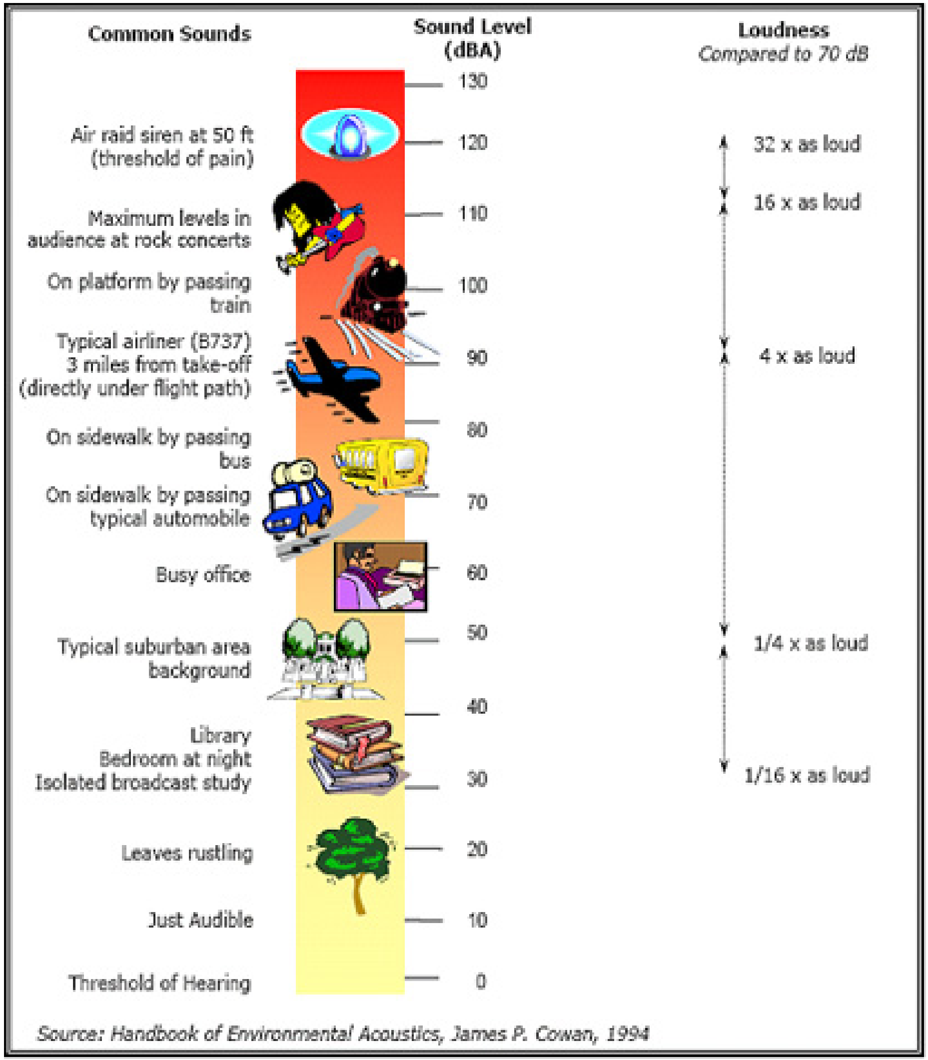
Range of some common sounds



Intensity Range for Some Common Sounds



Intensity Range for Some Common Sounds



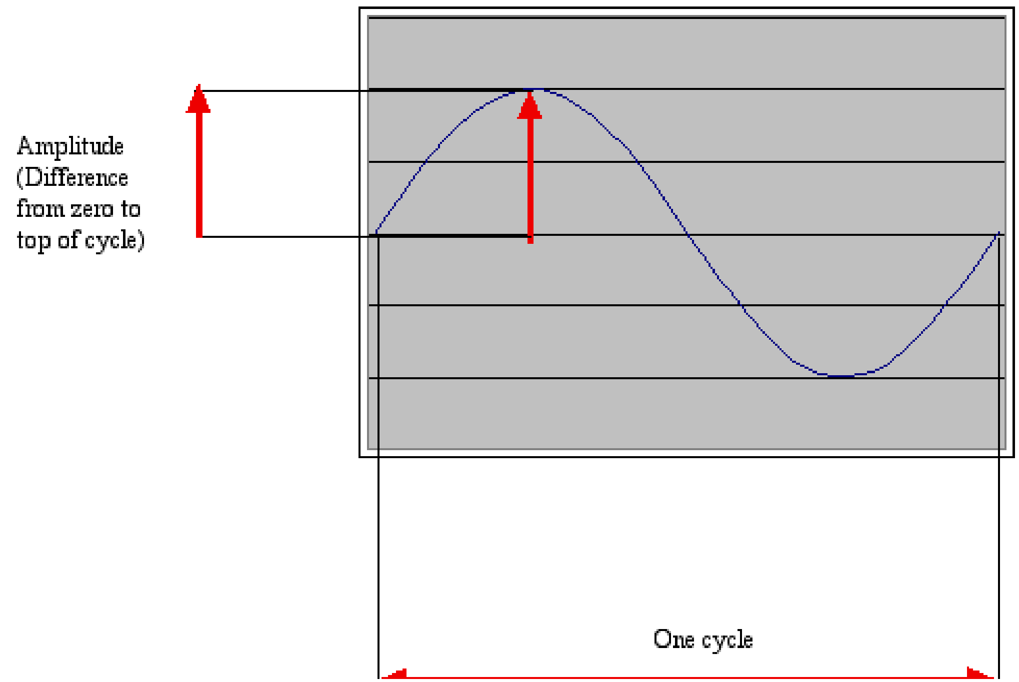
Sound waves

- The velocity of sound in air depends on the air temperature.
  - eg. 311.5 m/s at 0 °C in dry air.
  - Velocity increases 0.6m/s for every 1°C increased in temperature.
- Sound generally travels fastest in solids and slowest in gases but there are exceptions.

Medium	Velocity (m/s)	Medium	Velocity (m/s)
Air	330	Carbon dioxide	260
Helium	930	Hydrogen	1270
Oxygen	320	Water	1460
Sea water	1520	Mercury	1450
Glass	5500	Granite	5950
Lead	1230	Pine wood	3320
Copper	3800	Aluminum	5100

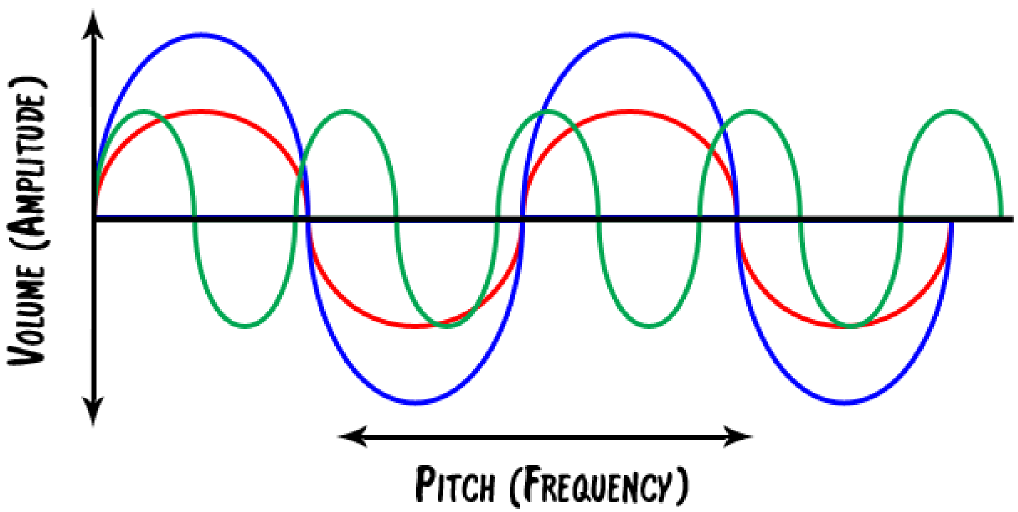
Sound waves

- Sound comes in cycles.
- The *frequency* of a wave is the number of cycles per second (cps), or *Hertz*
  - Complex sounds have more than one frequency in them.
- The amplitude is the maximum height of the wave.



Sound waves

- The human ear relates amplitude to loudness and frequency to pitch.
  - Listen to various sound frequencies here.
  - Listen to sound with frequency from 1 Hz-1000Hz here.



Decibel is a logarithmic measure

- A *decibel* is a ratio between two intensities: $10 * \log_{10}(I_1/I_2)$
  - As an absolute measure, it's in comparison to threshold of audibility
  - 0 dB can't be heard.
  - Normal speech is 60 dB.
  - A shout is about 80 dB.

Sound Recording & Audio Qualities

- We typically use a microphone to collect sounds. In such cases, environmental noises, and many other factors need to be considered. (omitted)

- In our course, we mainly consider the subsequent **sound storage**, **sampling** and **quantization** problems.

- For reproduction quality, we also focus on the effects brought by different sampling and quantization methods.

http://en.wikipedia.org/wiki/Sound_recording_and_reproduction

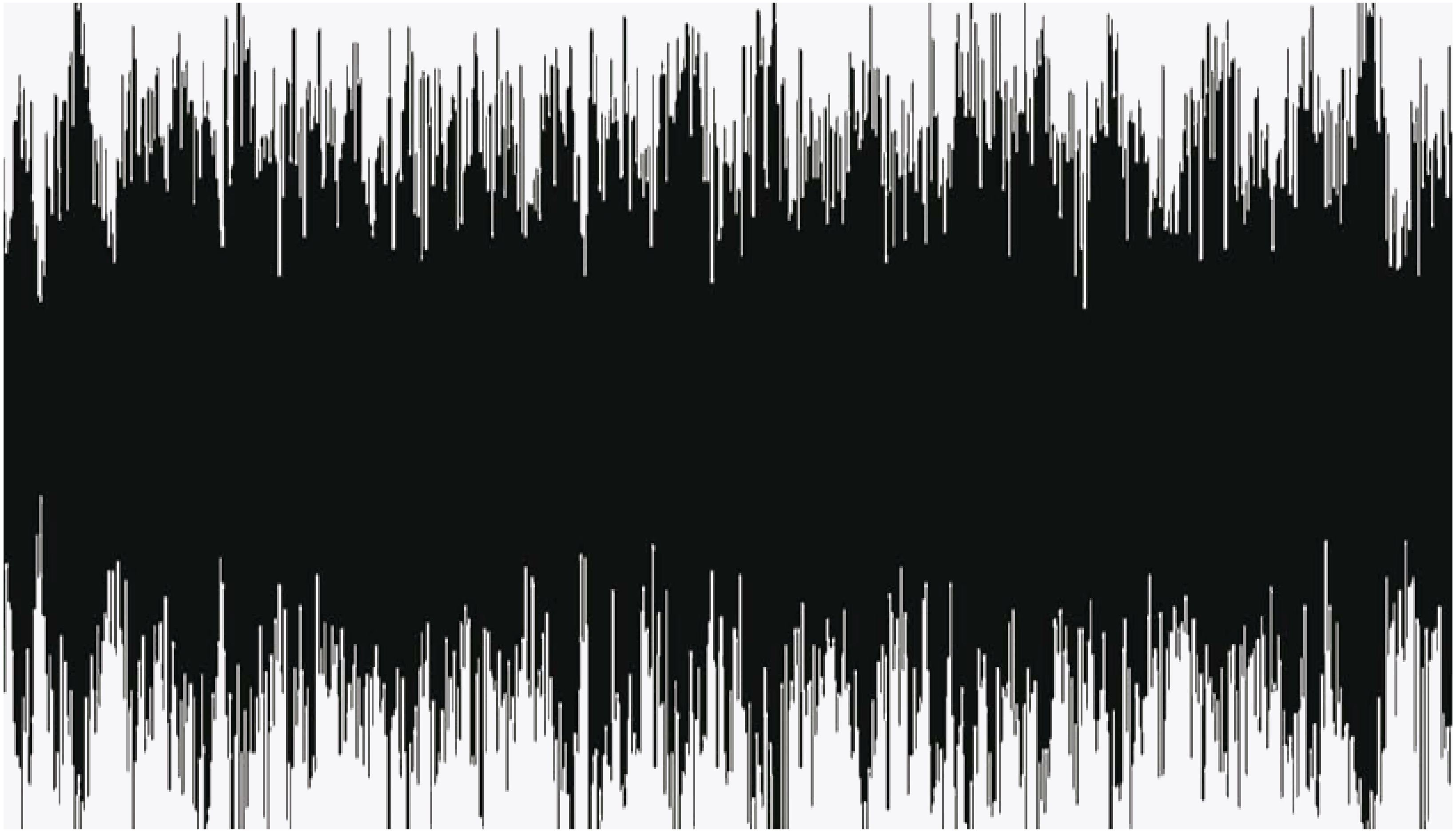
Identifying a waveform?

- A sound's waveform shows how its amplitude varies over time.

Speech: "Feisty teenager"

Men grow cold, as girls grow old
And we all lose our charms in the end

by **Marilyn Monroe**



Music: Didgeridoo

Music: Boogie-woogie

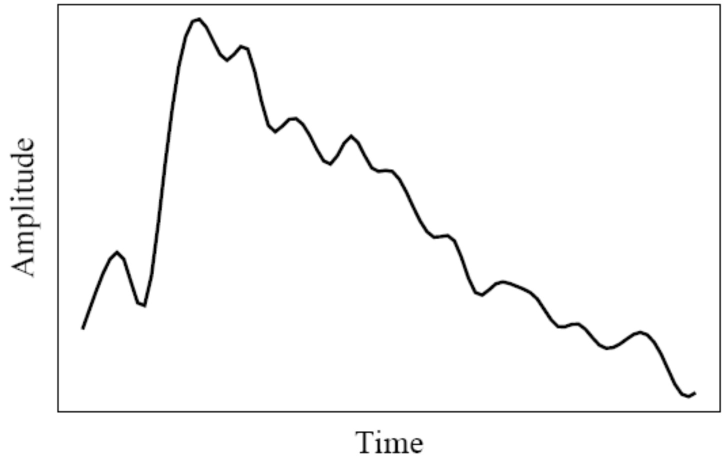
Music: Contemporary classical piece for violin, cello and piano

A trickling stream

The sea

Sound Digitization

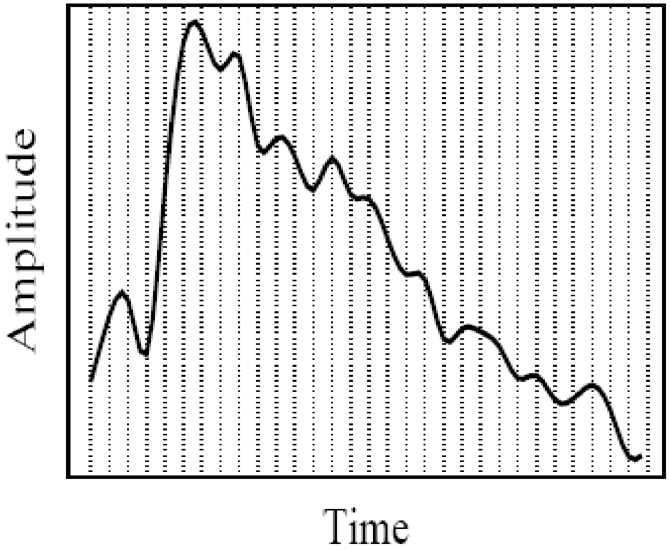
- **Digitization** means conversion to a stream of numbers, and preferably these numbers should be integers for efficiency.
- 1-dimensional nature of sound: **amplitude** values depend on a 1D variable, **time**.



Sound Digitization

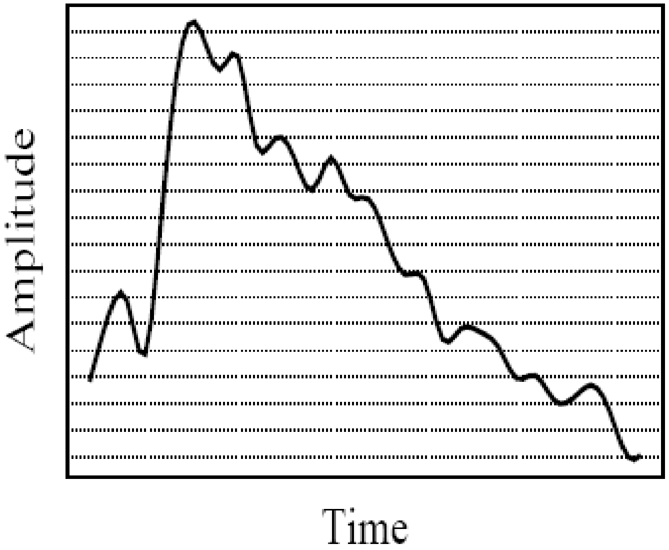
- Digitization must be in both time and amplitude.
  - Measuring the quantity we are interested in, usually at evenly-spaced intervals.
- First, using measurements only at evenly spaced time intervals, is simply called ***sampling***. The rate at which it is performed is called the *sampling frequency*
  - For audio, typical sampling rates are from 8 kHz (8,000 samples per second) to 48 kHz. This range is determined by Nyquist theorem discussed later.
- Second, get samples in the amplitude or voltage dimension is called **quantization**

Sampling and Quantization



(a)

Sampling

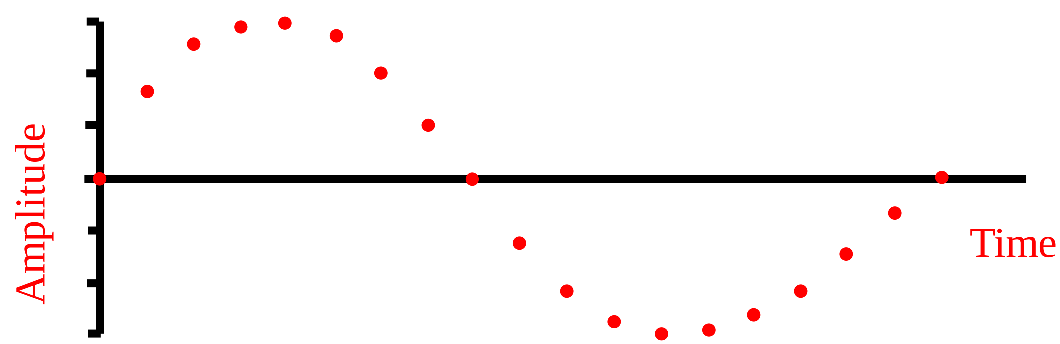


(b)

Quantization

Digitization-Sampling

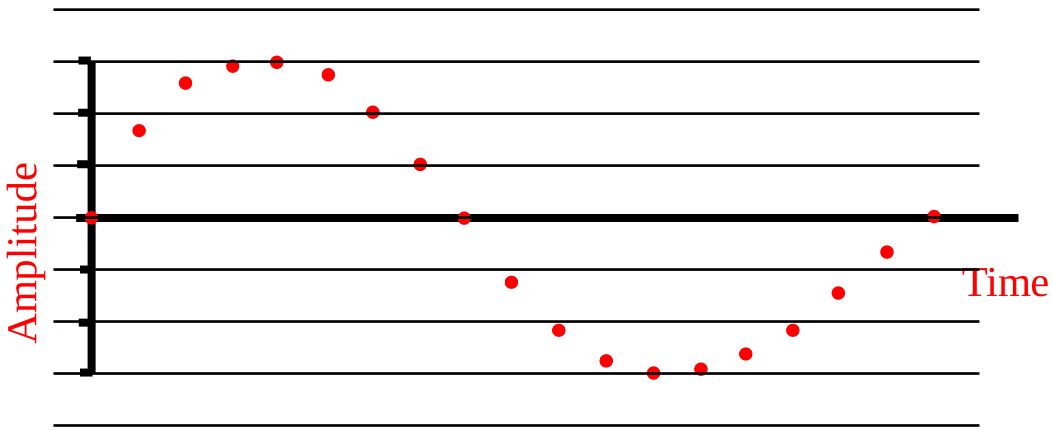
- Sampling
  - Divide the time axis into discrete pieces.
- Sampling rate
  - Number of samples per second (measured in Hz)



Sampling

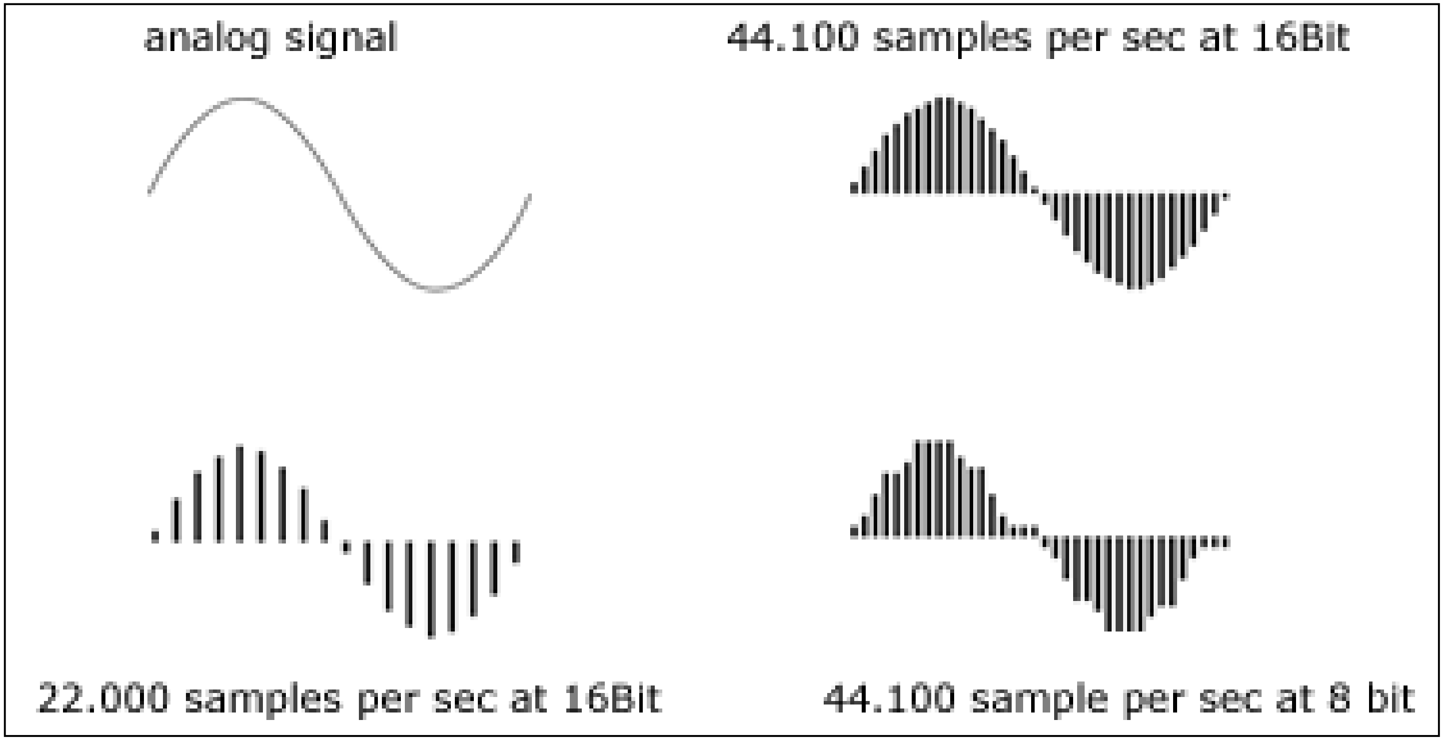
Digitization-Quantization

- Quantization
  - Divide the vertical axis (signal strength - voltage) into pieces
    - 8-bit quantization divides the vertical axis into 256 levels
    - 16-bit → 65536 levels.
  - The lower the quantization→ the lower the quality of the signal
- Example
  - 3-bit quantization →8 possible sample values



Quantization

Audio Digitization (PCM)



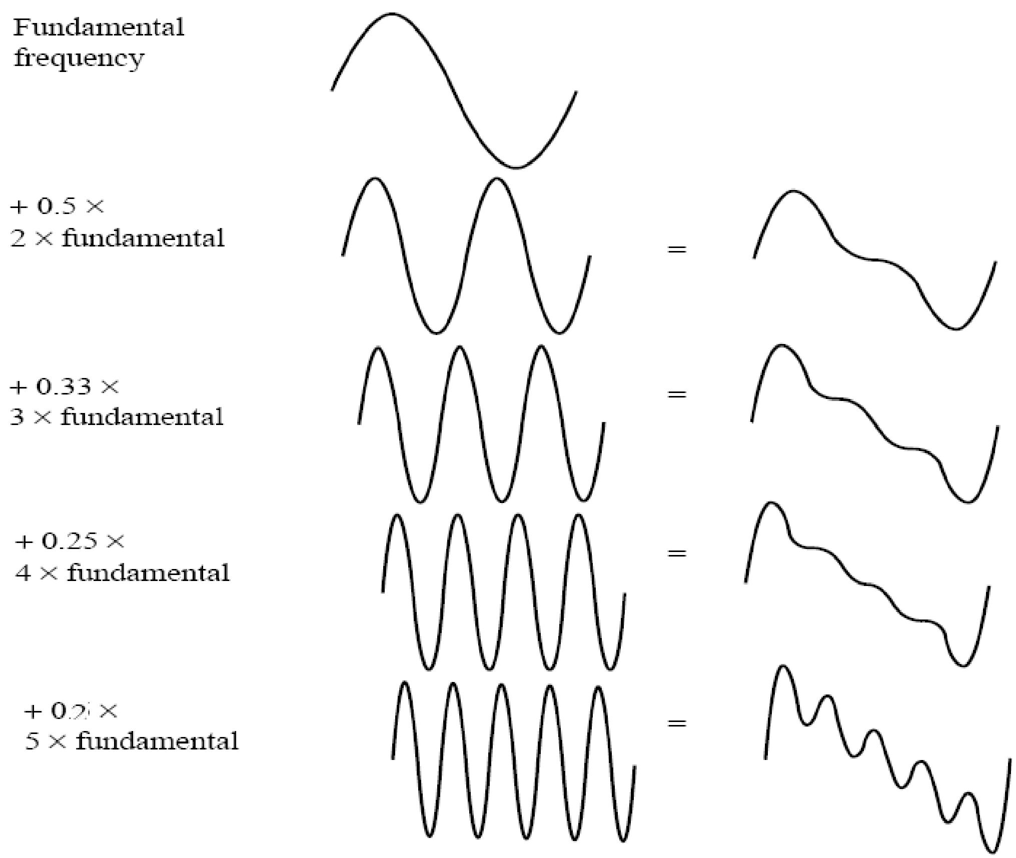
PCM: Pulse coded modulation

Parameters in Digitizing

- To decide how to digitize audio data we need to answer the following questions:
  1. What is the sampling rate?
  2. How finely is the data to be quantized, and is quantization uniform?
  3. How is audio data formatted? (file format)

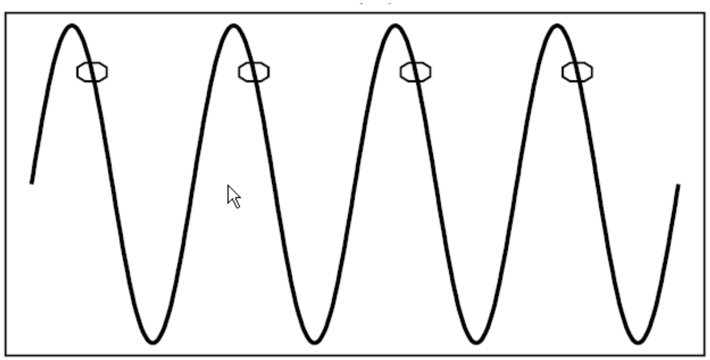
Sampling

- Signals can be decomposed into a sum of sinusoids.
- Weighted sinusoids can build up quite a complex signals.

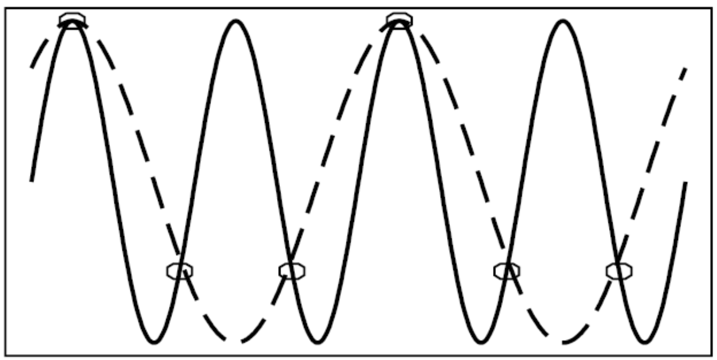


Sampling Rate cont'd

- If sampling rate just equals the actual frequency
    - a false signal (constant) is detected
- If sample at 1.5 times the actual frequency
    - an incorrect (**alias**) frequency that is lower than the correct one
        - it is half the correct one -- the wavelength, from peak to peak, is double that of the actual signal.



Sampling rate=f



Sampling rate=1.5*f

Nyquist Theorem

- For correct sampling we must use a sampling rate equal to **at least *twice the maximum frequency* content** in the signal, in order to recover it in the future. This rate is called the **Nyquist rate**.
- Sampling theory – Nyquist theorem

  If a signal is **band-limited**, i.e., there is a lower limit f_1 and an upper limit f_2 of frequency components in the signal, then the sampling rate should be at least $2(f_2 - f_1)$.

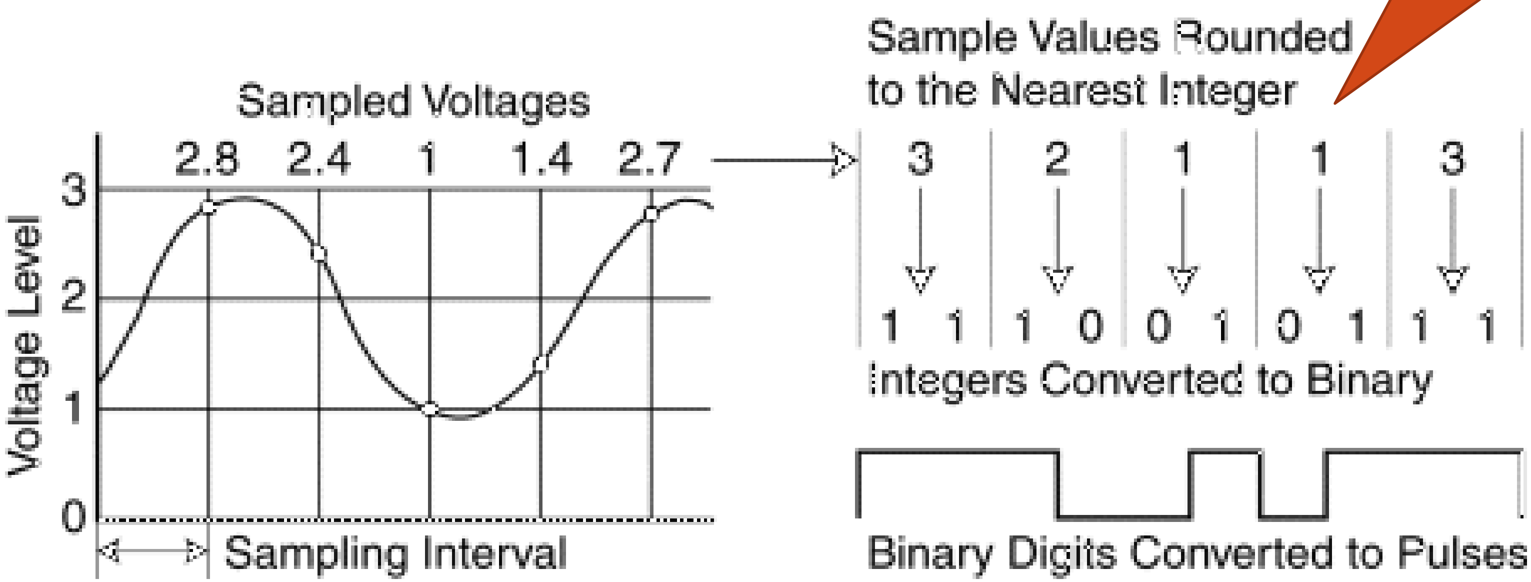
Example: CD

- CD audio is sampled at **44.1kHz**.
- Sampling relies on highly accurate clock pulses to determine the sample intervals.
  - If the clock drifts, Jitter (timing variations) occurs.
  - For CD quality, the jitter < 200 picoseconds.
- Frequencies greater than half the sampling rate are filtered out to avoid aliasing.

Quantization (Pulse Code Modulation)

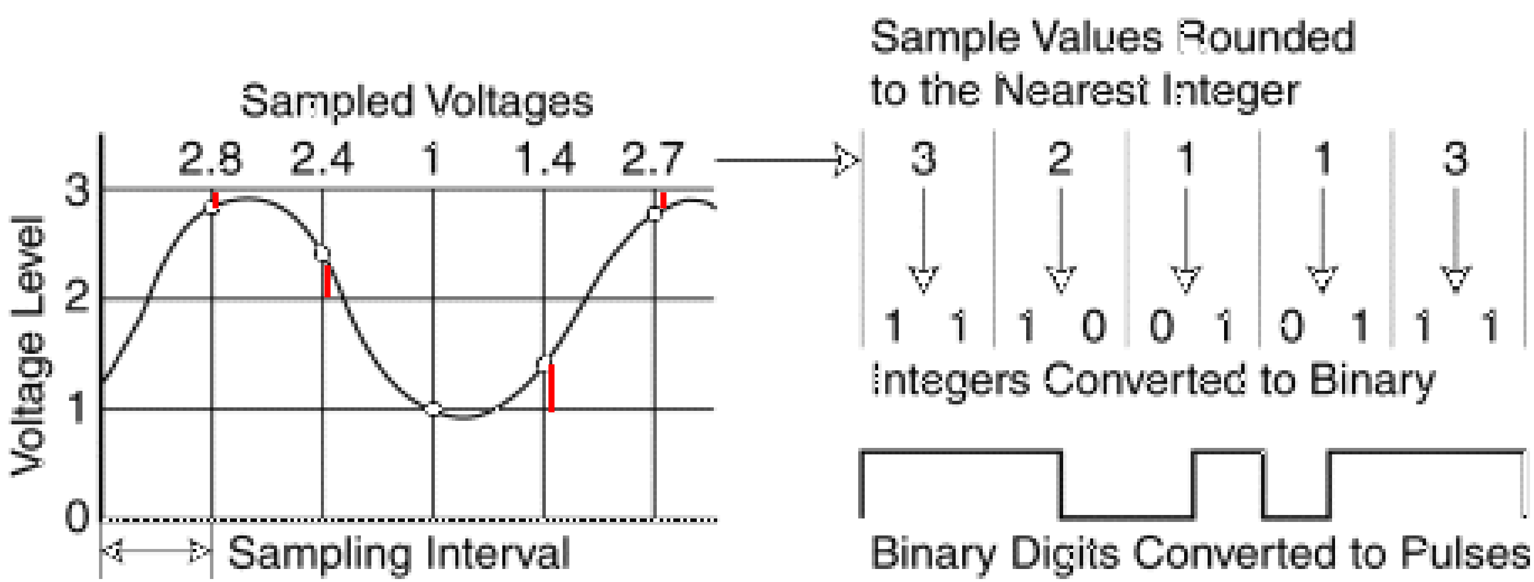
- At every time interval the sound is converted to a digital equivalent.
- Using 2 bits the following sound can be digitized

Quantization brings in Noise!



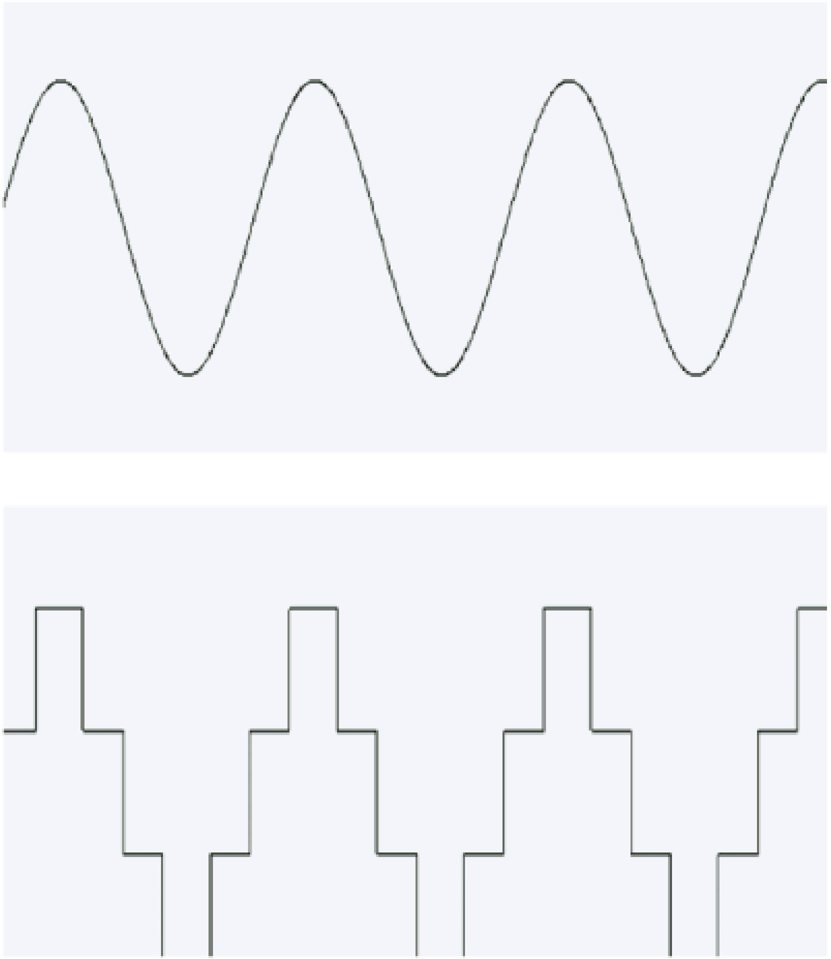
More on Quantization

- Quantization is lossy
- Roundoff errors => **quantization noise/error**

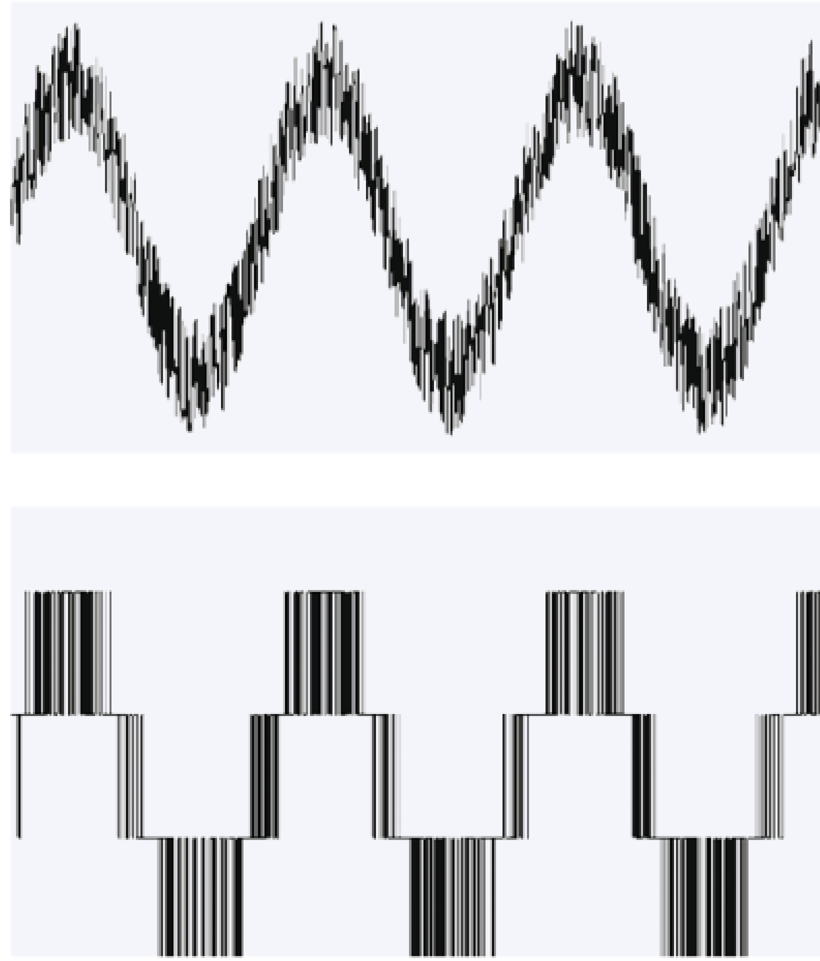


Example: CD

- CD audio uses 16-bit samples to give 65,536 quantization levels.
- Quantization noise can be sufficiently eliminated by dithering.
  - i.e. Dithering here means adding a small amount of random noise which softens the sharp transitions of quantization noise.



Undersampling a pure sine wave



Dithering

Audio Quality vs. Data Rate

Quality	Sample Rate (KHz)	Bits per Sample	Mono/ Stereo	Data Rate (uncompressed) (kB/sec)	Frequency Band (KHz)
Telephone	8	8	Mono	8	0.200-3.4
AM Radio	11.025	8	Mono	11.0	0.1-5.5
FM Radio	22.05	16	Stereo	88.2	0.02-11
CD	44.1	16	Stereo	176.4	0.005-20
DAT	48	16	Stereo	192.0	0.005-20
DVD Audio	192 (max)	24 (max)	6 channels	1,200.0 (max)	0-96 (max)

Question

- For a sampling rate of r *Hz and sample size* of s *bits, each second of digitized sound will* occupy $rs/8$ *bytes.*
  - For CD quality, $r = 44.1 \times 10^3$ *and* $s = 16$, *so each* second occupies just over 88 kbytes (for a mono signal).

Question

An analog audio signal has a bandwidth from 10kHz-25kHz. To converting it to digital form:

1. What sampling rate should be used?
2. After sampling, if 8 bit quantization is used, what is the data rate of that digital signal?

- **Answer:**

  According to Nyquist Theorem, Nyquist rate=$2(f_2 - f_1)$=2*(25-10)=30kHz.

  30kHz sampling rate→ 30,000 samples/sec.

  8-bit quantization→8 bits/sample→256 quantization levels.

  The data rate= 30,000*8= 240,000 bps=240kbps.

- **Example rates**
  - CD: 1.411 Mbps
  - MP3: 96, 128, 160, 320 kbps
  - Internet telephony: 5.3 - 13 kbps

Compression

- Sound is difficult to compress using lossless methods, except for special cases.

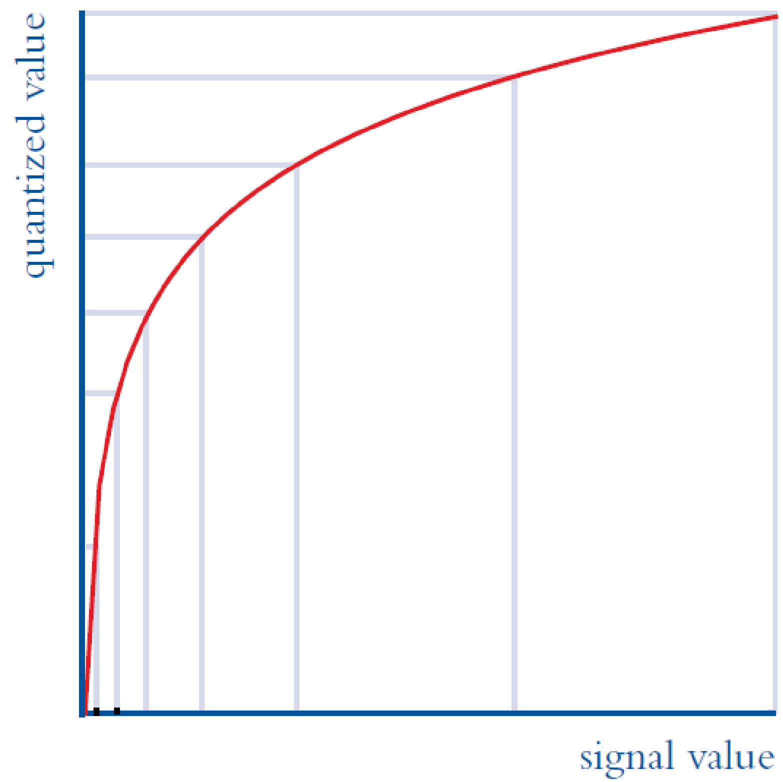
- Some compression of audio can be obtained by **run-length encoding** samples that fall below a threshold that can be considered to represent silence.
  - That is, instead of using 44,100 samples with the value of zero for each second of silence (assuming a 44.1 kHz sampling rate) we record the length of the silence.

Compression

- **Companding** uses non-linear quantization to compress speech.
  - Quiet sounds are represented in greater detail than louder ones.
- μ-law and A-law companding are used for telephony.

ITU recommendations for north America

ITU recommendations for the rest of the world



Non-linear quantization

Compression

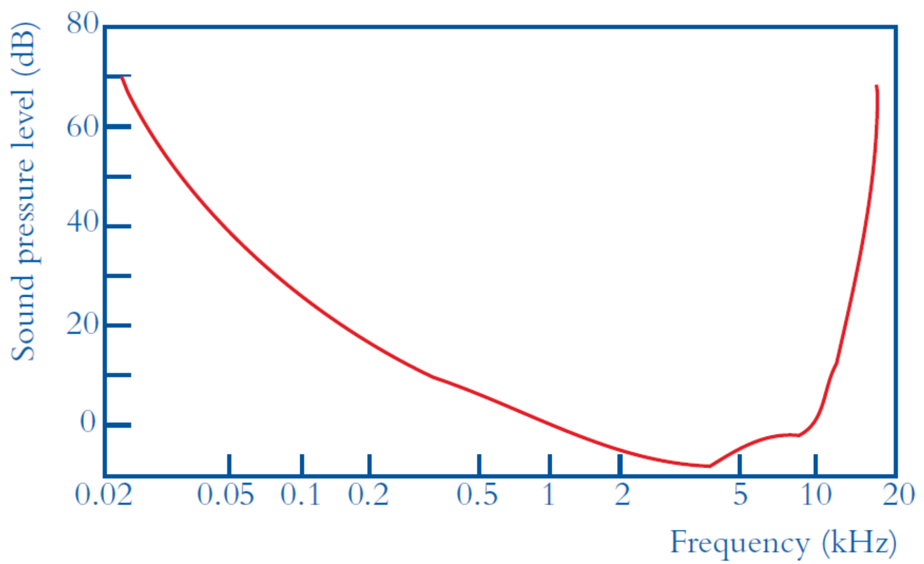
- **Adaptive Differential Pulse Code Modulation (ADPCM)**, which works by storing information about the difference between a sample and a value predicted from the preceding sample, is also used in telephony.

Compression

- **Perceptually-based compression** discards inaudible sounds.
  - Sounds may be too ‘quiet’ to be heard.
  - Sounds obscured by some other sounds.

Compression

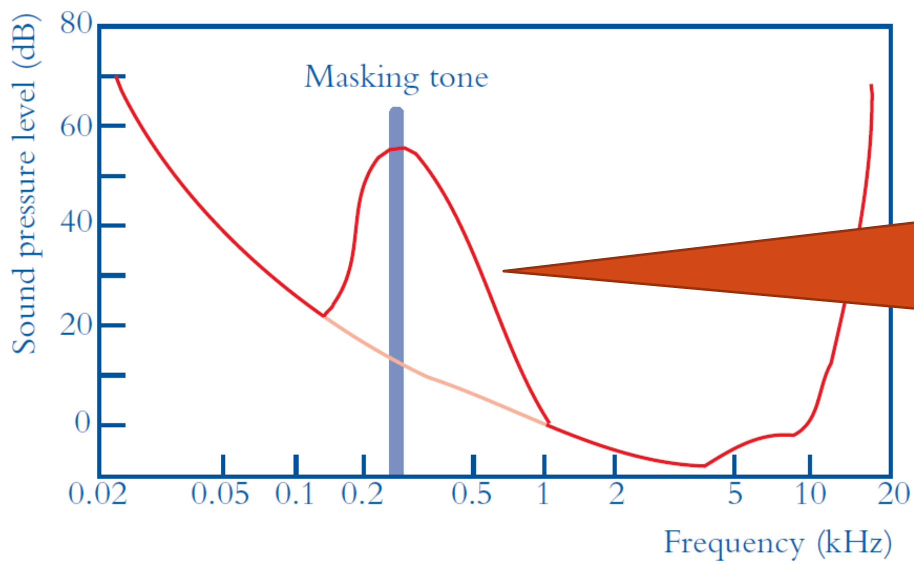
- A psycho-acoustical model describes how the threshold of hearing varies non-linearly with frequency.
  - The threshold of hearing is the minimum level at which a sound can be heard.
  - It varies non-linearly with frequency. A very-low or very high-frequency sound must be much louder than a mid-range tone in order to be heard. We are more sensitive to sounds in the frequency range that corresponds to human speech.



The threshold of hearing

Compression

- Loud tones can obscure softer tones that occur at similar time.
- **Masking** is a modification of the threshold of hearing curve in the region of a loud tone.
- The threshold is raised in the neighborhood of the masking tone.



Masking curve is non-linear and asymmetrical- rising faster than it falls.

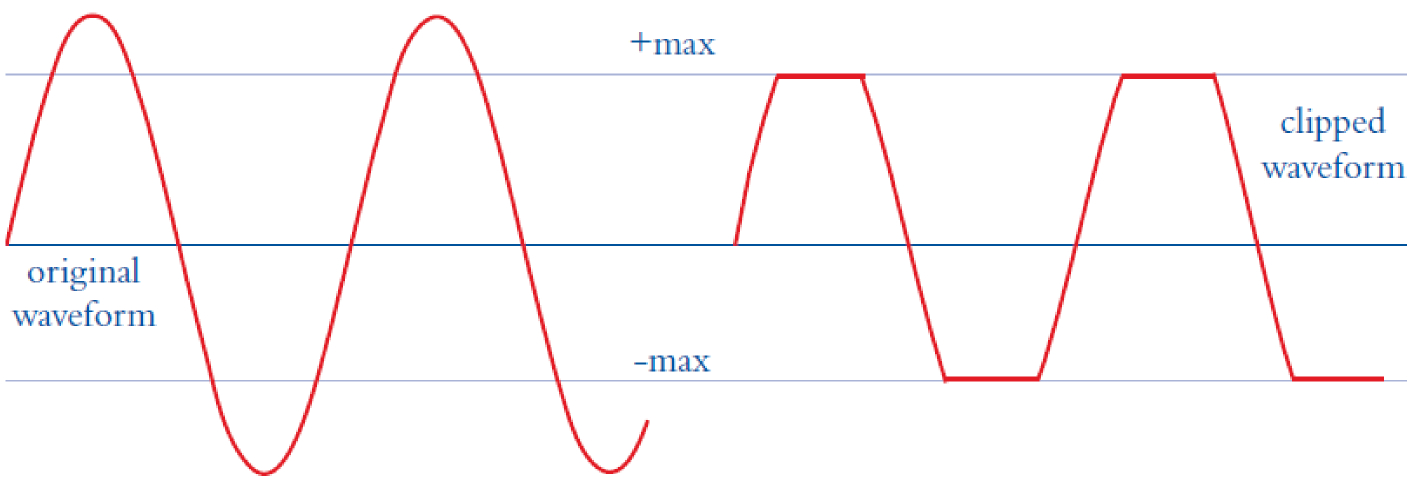
Modification of the threshold of hearing by a masking tone

Compression

- Filters are used to split a signal into 32 bands, and a masking level for each band is computed. Signals that fall below the level can be discarded.
- Practical implementations of perceptually based compression are the basis of *MP3* and *AAC* compression.

Sound Processing

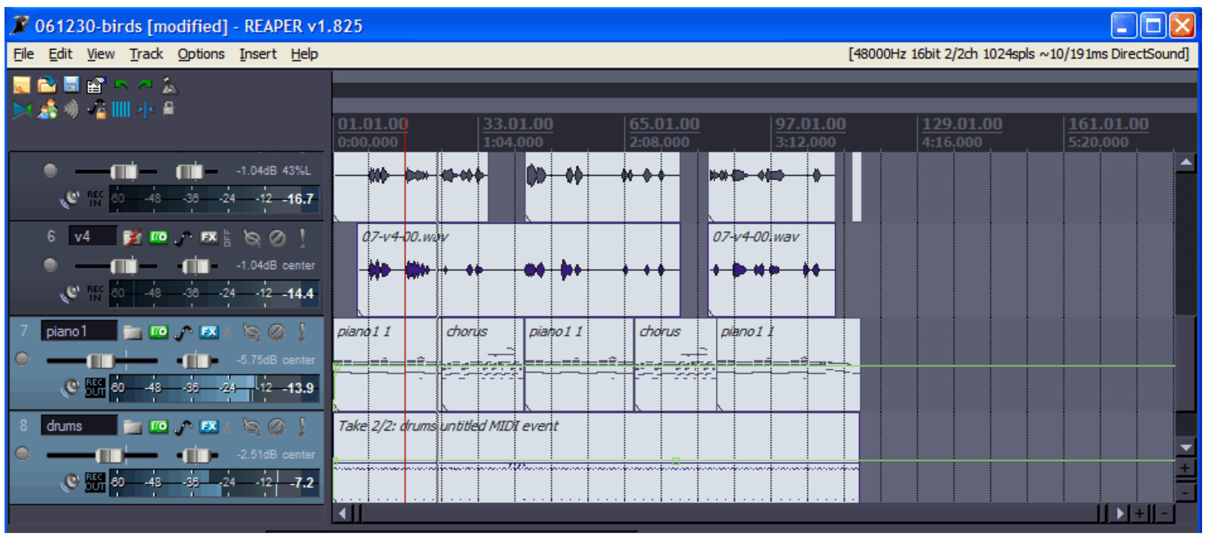
- If the recording level is too high, exceeding the max value that can be recorded, clipping will occur, causing distortion.



Clipping

Sound Processing

- Sound editing programs use a timeline interface, with multiple tracks (usually displayed as waveforms), which are mixed down to produce a stereo or mono output.



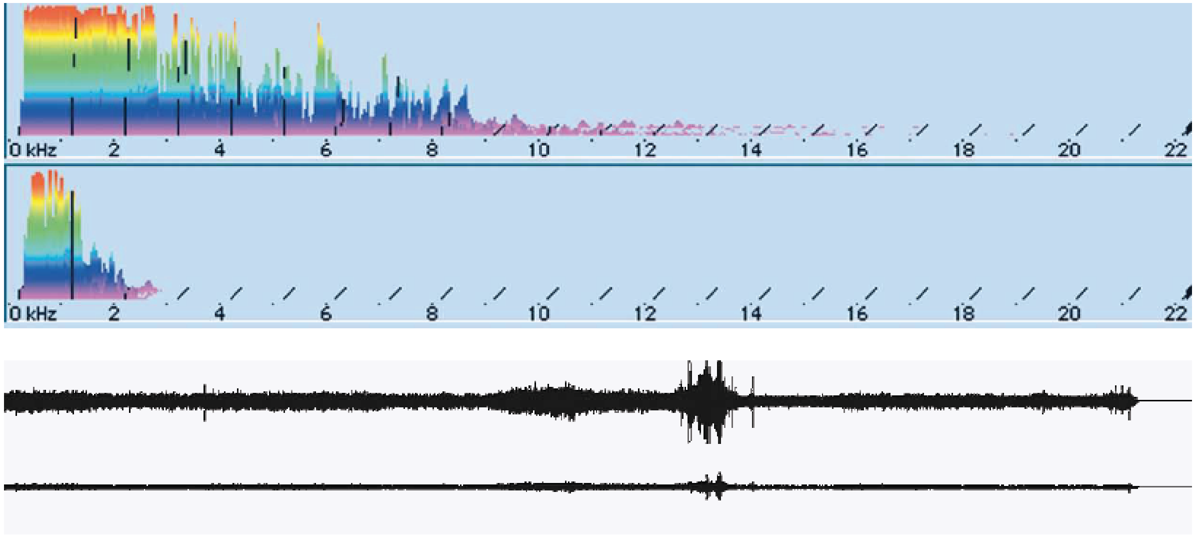
Sound Processing

- Short loops may be used to create voices for samplers; longer loops may be combined (e.g. in GarageBand) to build songs from repeating sections.

- Filters and gates are used to correct defects (e.g. remove noise) or to enhance or modify sounds (e.g. reverb).

Sound Processing

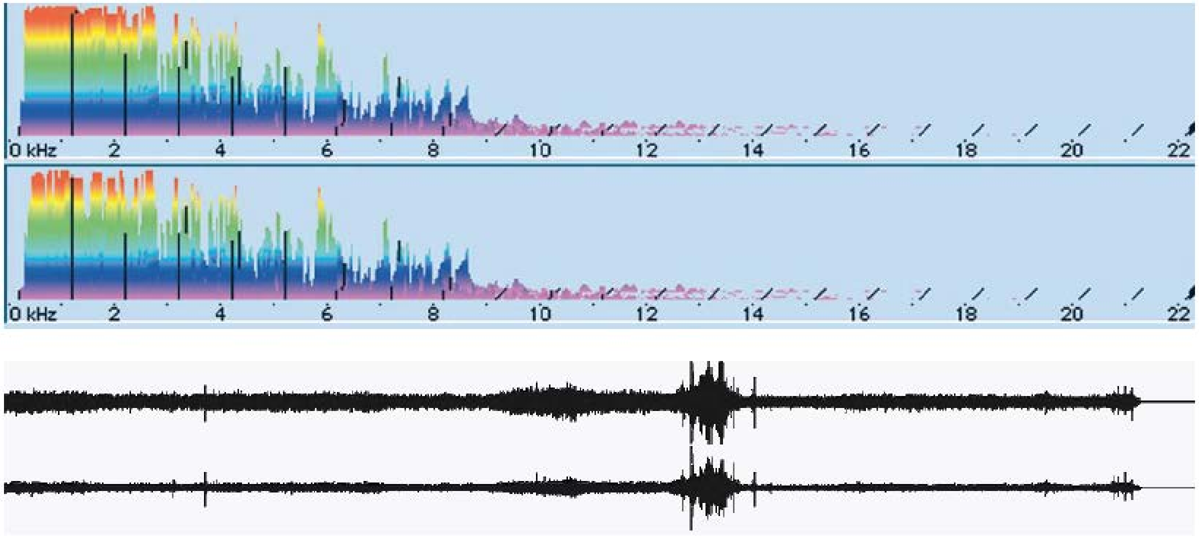
- Low pass filtering:
  - Allow low frequencies to pass through, remove high frequencies
  - Can be used to take out hiss



Low pass filtering

Processing Sound

- High pass filtering:
  - Allow high frequencies to pass through, remove low frequencies.
  - Can be used to take out low frequency 'rumble' noise caused by mechanical vibrations.



High pass filtering

Processing Sound

- Notch filter
  - Remove a single narrow frequency band.
  - Example: remove hum picked up from the mains, typically a frequency 50 or 60Hz.

- Some sophisticated software supports:
  - Redraw a waveform.
  - Rub out the spikes corresponding to clicks, etc.
  - Not always easy, requiring considerable experience.

Processing Sound

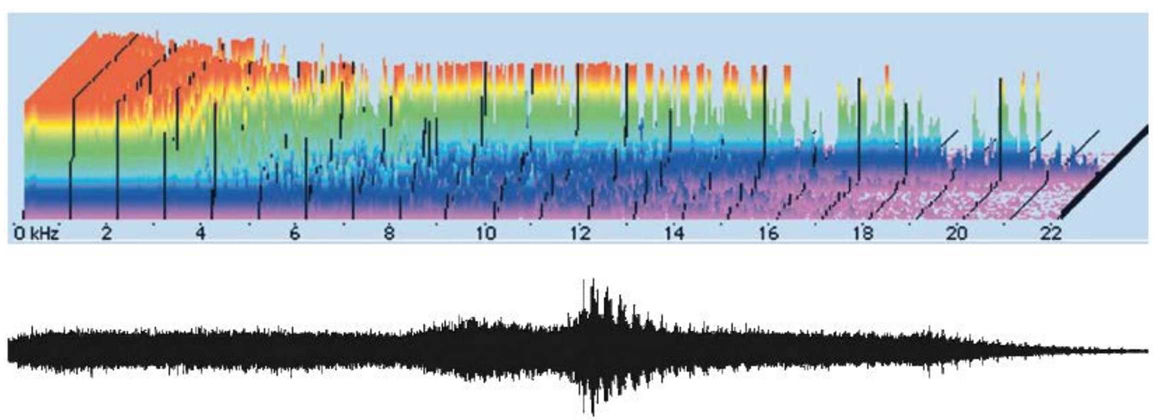
- Specialized filters:
  - De-esser
    - Remove the sibilance that results from speaking or singing into a microphone placed too close to the performer.
  - Click repairers
    - Remove clicks from recordings taken from damaged or dirty vinyl records.

Processing Sound

- All the filters are not infallible.
- The only sure way to get perfect sound is to start with a perfect take.
    - Microphones should be positioned to avoid sibilance
    - Kept away from fans and disk drives
    - Cables should be screened to avoid picking up hum, and so on.

Processing Sound

- Special effects:
  - Reverb effect:
    - Produced digitally by adding copies of a signal, delayed in time and attenuated, to the original.
    - These copies model reflections from surrounding surfaces.
    - The delay corresponding to the size of the enclosing space and the degree of attenuation modeling surfaces with different acoustic reflectivity.



Echo reverb

Processing Sound

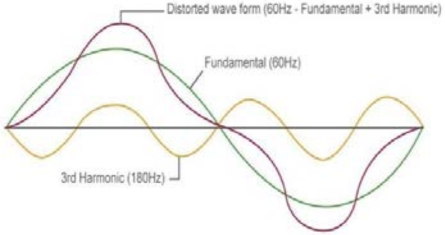
- Faders:
  - A sound's volume gradually decreases or increases.
- Tremolo:
  - The amplitude to oscillate periodically from zero to its maximum
- Time stretching (slowing down and speeding up) and pitch alteration are more easily applied to digital audio than they were to analogue audio.
  - They are used for synchronization and for matching (e.g. when combining separately recorded loops).

Processing Sound

- Sound can be combined with pictures in a video editing program: sound tracks are displayed on the same timeline as video tracks, where they can be synchronized.
- Timecode is just a fiction when working with sound, owing to the high sampling rate, but it is valuable for synchronization.
- If sound and video are physically independent in a movie, synchronization may be lost, especially when it is sent over a network.

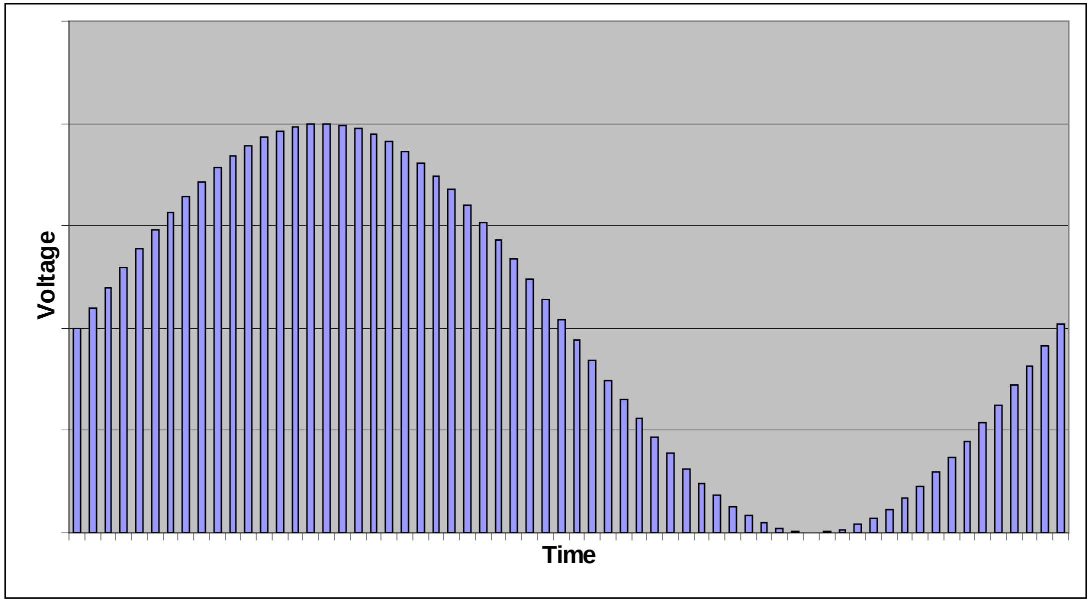
Revision

- When voice or music is captured by a microphone, it generates electrical signal.
- The signal consists of fundamental sine wave with certain **frequency** and **amplitude**.
- The fundamental sine wave is accompanied by harmonics.
- Adding the fundamental to harmonics, forms a composite sinusoidal signal that represents the original sound.
- Analog sinusoidal waveforms are converted to digital format by the **sampling** process.



Revision

- Sampling Process
  - The analog signal is sampled over time at regular intervals to obtain amplitudes of the signal at sampling time.
  - The interval at which sampling occurs is associated with **sampling rate**.



$$\text{Sampling rate} = \frac{1}{\text{sample time}} = \frac{1}{T}$$

T: Time interval between two samples

Revision

- Sampling Process (cont.)
  - The sample amplitude obtained at sampling instants is represented by an 8-bit value (one byte) or 16-bit (two bytes) value.
  - Higher values can also be used for higher resolution systems (high fidelity sound).
  - A composite signal of 11.025 kHz sampled 4 times every cycle will yield 44.1 kHz sampling rate. If you sample at higher rate, you need to store more samples.
  - For CD quality music at 44.1 kHz rate at 16-bit resolution, a one minute recording will require 44.1 x 1000 x 16 x 60 / 8 = 5.292 Mbytes.

Revision

- Audio objects generate a large volume of data. This poses two problems.
  - it requires a large storage
  - it takes a long time to transmit the data
- To solve these problems, the data is compressed.
  - Compression helps to shrink the storage and reduce network time.

- Audio industry uses 5.0125kHz, 22.05kHz and 44.1kHz as standard sampling frequencies. These frequencies are supported by most sound cards.

Sound Conversions

- Why considering sound conversions
  - **Tradeoff** between quality and storage

- Audio File Format
  - Uncompressed File Format
  - Format with Lossless Compression
  - Format with Lossy Compression
  - Other formats
  - Recommendations for WWW
- Codecs
- Audio Conversion Softwares

http://en.wikipedia.org/wiki/Audio_file_format

Sound Conversions

- Thinkings in mind…
  - Why are there so many audio formats?
  - What is the mathematical foundation of different audio files?
  - What is the benefit/shortcoming of each file format?
  - Which technique/format should I deploy in my application?

  - How to encode/decode each format?
  - Any non-technical restrictions in audio process

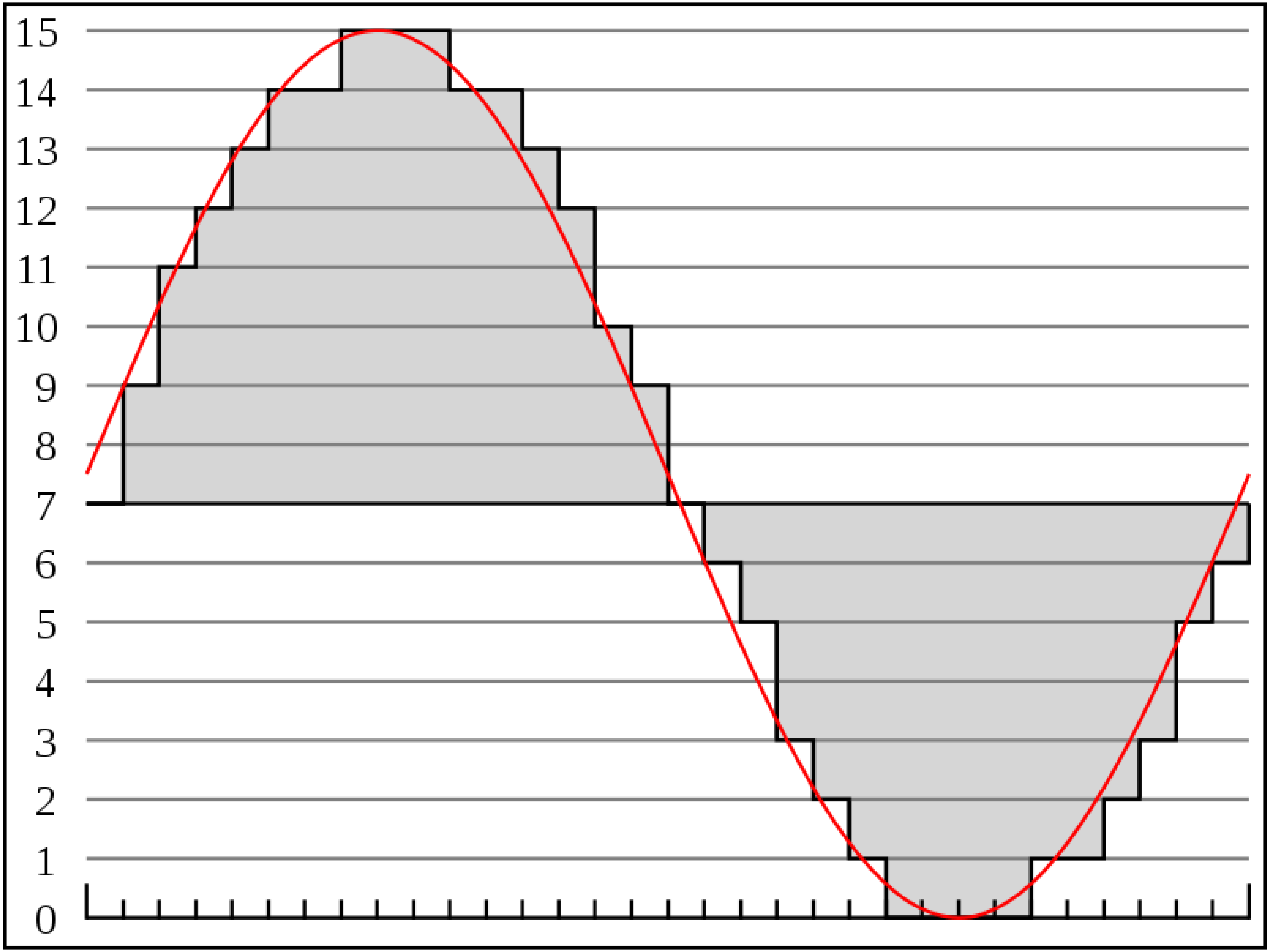
Uncompressed File Formats

- WAV
- AIFF
- AU
- Raw Audio Format, e.g., PCM
- ……

Pulse-Code Modulation

- PCM is a digital representation of an analog signal
  - the magnitude of the signal is sampled **regularly** at uniform intervals
  - then quantized to a series of symbols in a numeric (usually binary) code.

  - It is used in digital telephone systems.
  - The standard form for digital audio in computers and the compact disc format.
  - It is standard in digital video.
  - Uncompressed PCM is **NOT** typically used for video in standard definition consumer applications such as DVD because the bit rate required is far too high.

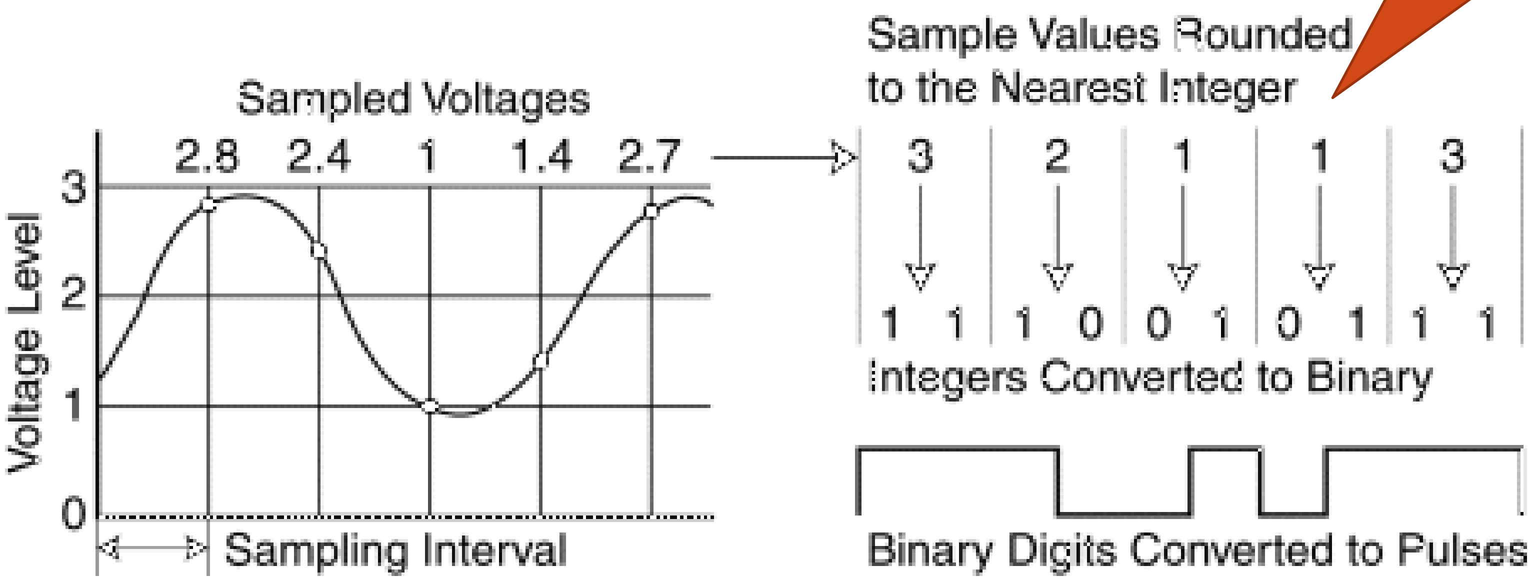


Sampling and quantization of a signal (red) for 4-bit PCM

Recap: Quantization (Pulse Code Modulation)

- At every time interval the sound is converted to a digital equivalent.
- Using 2 bits the following sound can be digitized

Quantization brings in Noise!

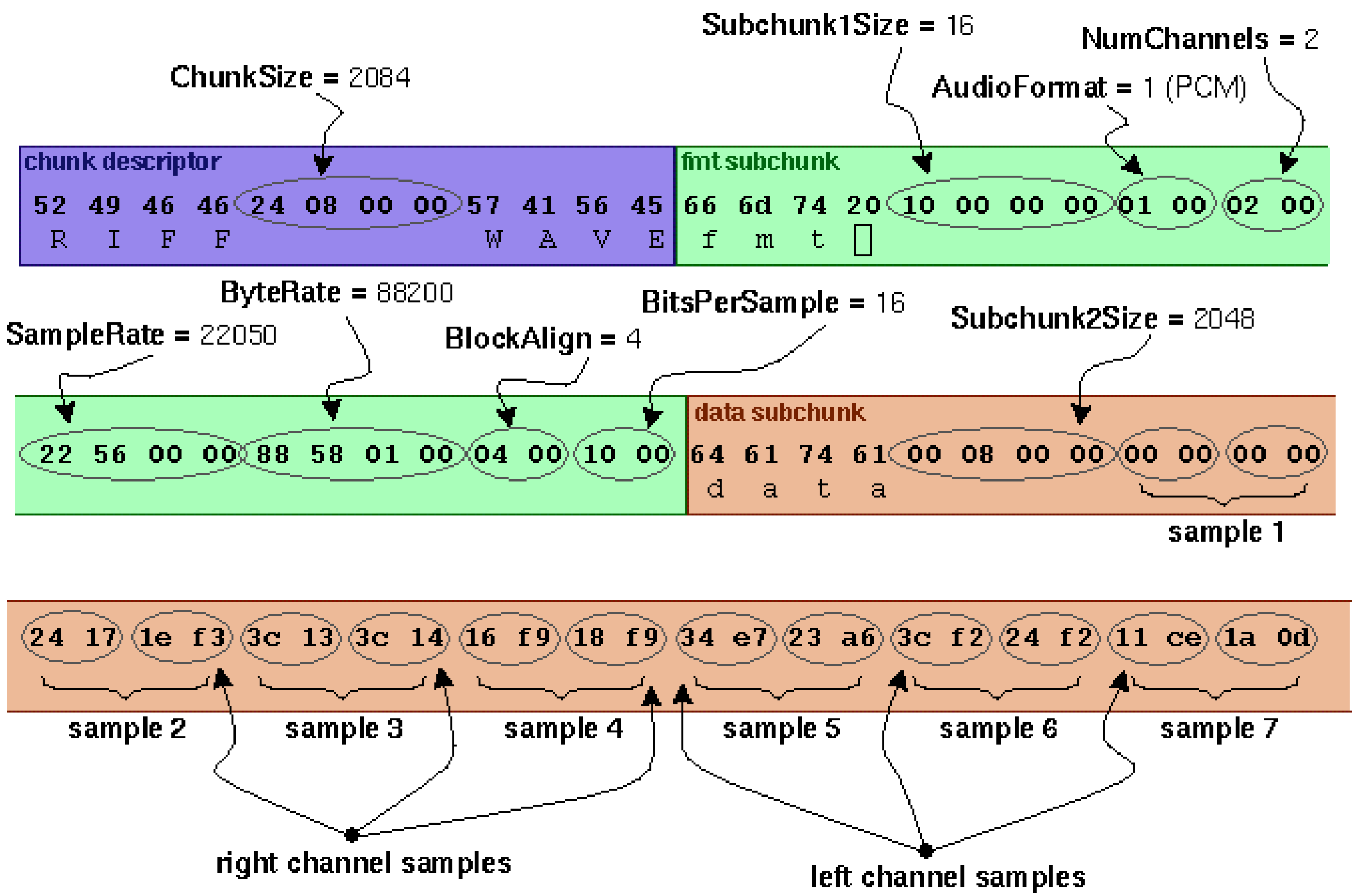


PCM

- PCM is usually stored as a .wav on Windows or as .aiff on Mac OS.
  - WAV and AIFF are flexible file formats designed to store more or less any combination of sampling rates or bitrates.
  - Suitable for storing and archiving an **original recording**.
  - AIFF is based on IFF (Interchange File Format by *Electric Arts*) format.
  - WAV is based on RIFF (Resource Interchange File Format) format, which is similar to IFF.

http://en.wikipedia.org/wiki/Interchange_File_Format
http://en.wikipedia.org/wiki/RIFF_%28File_format%29

Audio File Format: .WAV

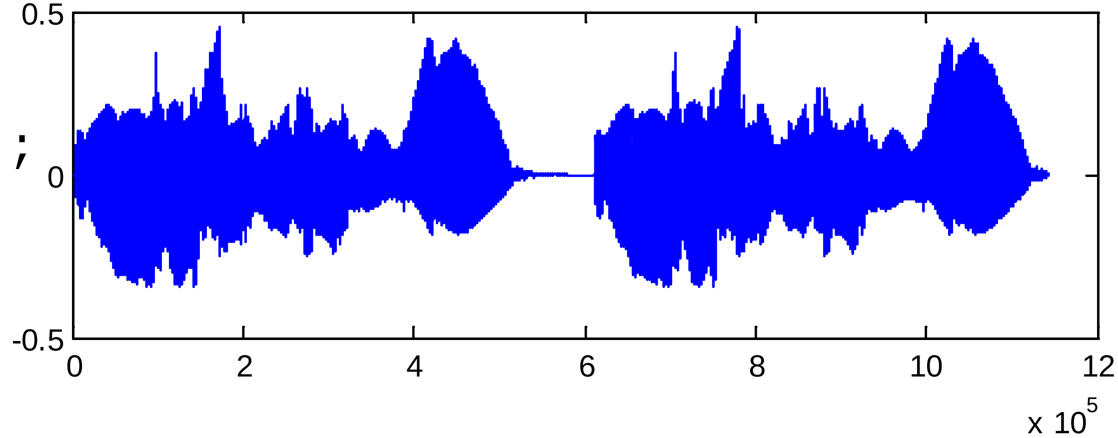


http://ccrma.stanford.edu/courses/422/projects/WaveFormat/

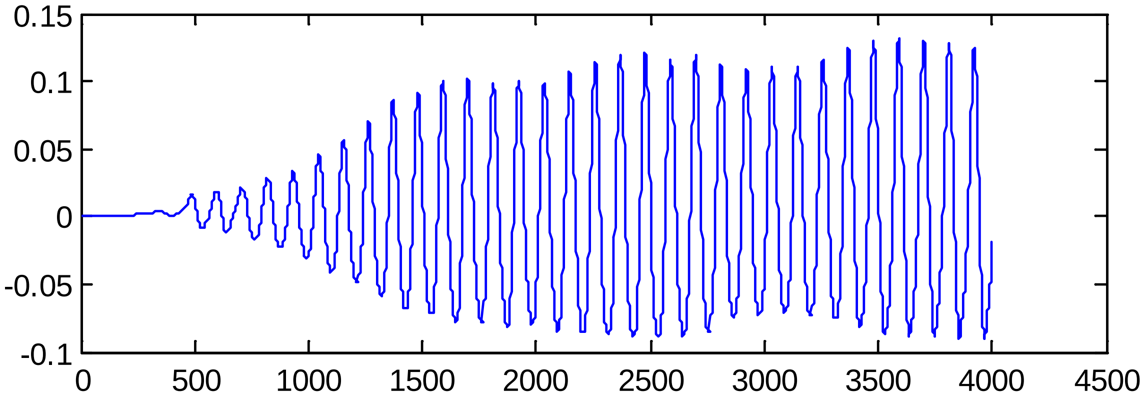
Example

Create this figure in Matlab:

```
x = wavread('horn.wav');
plot(x(:, 1));
plot(x(4000:10000, 1));
```



Note:
Wavread() normalizes the Samples to the range of [-1, 1].



Lossless Compressed Format

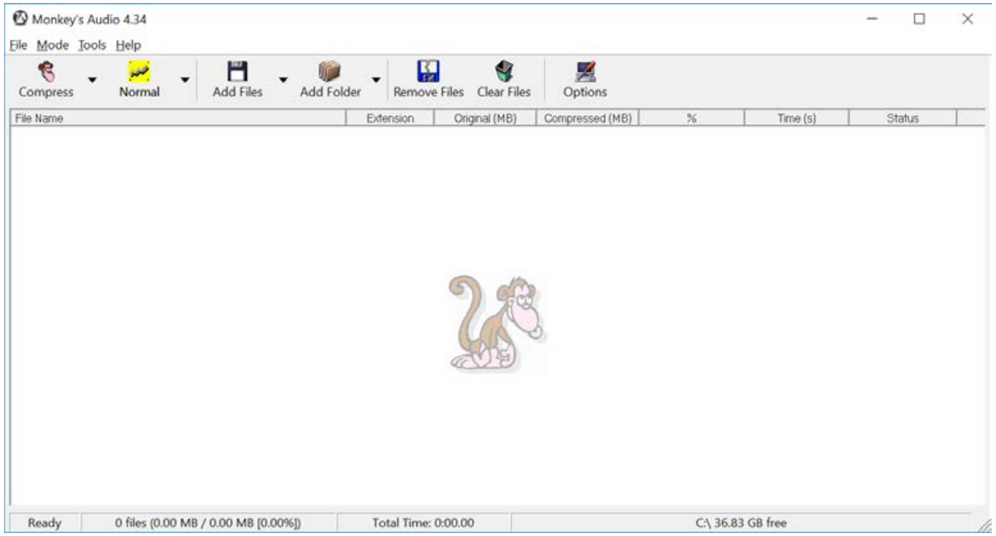
- Lossless compressed formats require more processing time than uncompressed formats but are more efficient in storage.
  - Uncompressed audio formats encode both sound and silence with the same number of bits per unit of time.
  - In a lossless compressed format, the music would occupy a smaller file and the silence take up almost no space at all.

- Lossless compression has a compression ratio of about 2:1
  - FLAC,
  - WavPack,
  - Monkey's Audio (APE),
  - ALAC/Apple Lossless
  - Shorten (SHN) …

Monkey's Audio

- Monkey's Audio is a lossless compression format for audio. It does not discard data during encoding.
- A digital recording (such as a CD) encoded to Monkey's Audio format can be decompressed into an identical copy of the original.
- Drawback: It is **proprietary** software, and has limited support on software platforms other than Windows. Alternatives such as FLAC may offer more options for some users.
- Filename extension: .ape for audio, and .apl for track metadata.

Monkey's Audio



Monkey’s Audio

- Storage:
  - Monkey's Audio compresses a little better than FLAC and a lot better than Shorten.
  - Encoding and decoding times are longer than FLAC and Shorten.
  - Like any lossless compression scheme, Monkey's Audio format takes up several times as much space as lossy compression formats like AAC, and MP3. A Monkey's Audio file is 3–5 times larger than a 192kbps bitrate MP3.
- Development:
  - The latest version of Monkey's Audio, ver. 4.34, was released on 2018-05-02.
  - The Shorten format is no longer in development.
  - FLAC has an active development community that continues to refine the format.
- Platform support:
  - Although Monkey's Audio is distributed as freeware, the license terms prevent most Linux distributions and other free software projects from including it.
  - Monkey's Audio is also supported on Linux and OS X using JRiver Media Center
  - FLAC has only open source licenses, so it comes pre-installed with most Linux distributions, is preferred by Linux users, and enjoys broad support in applications.

Lossy Compressed Formats

- MP3
- Vorbis
- Musepack
- AAC
- ATRAC
- RA
- lossy Windows Media Audio (WMA)
- …

Lossy Compressed Formats

- MP3
  - Designed by Moving Picture Experts Group as part of MPEG-1 standard.
  - The most popular format for downloading and storing music.

  - Eliminating portions of the audio file that are essentially inaudible.
  - A lossy compression algorithm to greatly reduce the amount of data required to represent the audio recording and still sound like a faithful reproduction of the original uncompressed audio for most listeners.
  - MP3 also has higher or lower bit rates, with higher or lower resulting quality.
  - Available sampling frequencies: 32, 44.1 and 48 kHz

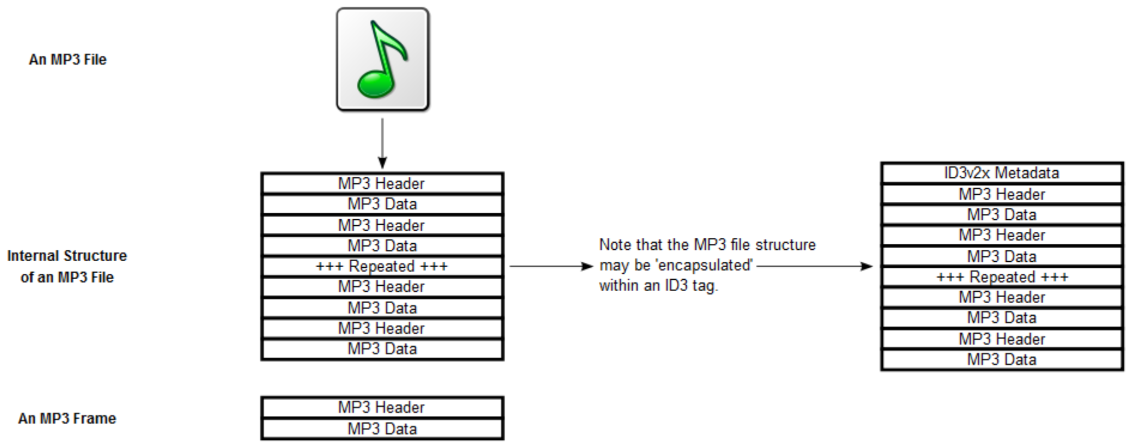
Lossy Compressed Formats

- Bitrates specified in MPEG-1 Audio Layer III standard:
  - 32, 40, 48, 56, 64, 80, 96, 112, 128, 160, 192, 224, 256 and 320 kbps
  - The bitrates 128, 160 and 192 kbps represent compression ratios of approximately 11:1, 9:1 and 7:1 respectively, comparing with CD.

- Constant Bit Rate Encoding (CBR)
  - one bit rate for the entire file
  - makes encoding simpler and faster

- Variable Bit Rate Encoding (VBR)
  - bit rate changes throughout the file
  - using a lower bit rate for the less complex passages and a higher one for the more complex parts

MP3 File Structure



Example MP3 Header

F F F B A 0 4 0 — Colour-coding shows binary bit mapping to hex values below

Detail of an MP3 Header

Bits	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Binary	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
Hex	F				F				F					B				A					0				4				0	
Meaning	MP3 Sync Word												Version	Layer		Error Protection	Bit Rate				Frequency		Pad. Bit	Priv. Bit	Mode		Mode Extension (Used With Joint Stereo)		Copy	Original	Emphasis	
Value	Sync Word												1 = MPEG	01 = Layer 3		1 = No	1010 = 160				00 = 44100 Hz		0 = Frame is not padded	Unknown	01 = Joint Stereo		0 = Intensity Stereo Off	0 = MS Stereo Off	0 = Not Copy-righted	0 = Copy Of Original Media	00 = None	

Lossy Compressed Formats

- WMA
  - Windows Media Audio format owned by Microsoft.
  - Designed with Digital Rights Management (DRM) abilities for copyright protection.
  - Parts of Windows media framework.
    - Windows media audio
    - Windows media audio professional
    - Windows media audio lossless
    - Windows media audio voice

Lossy Compressed Formats

- ATRAC (Adaptive Transform Acoustic Coding)
  - Developed by Sony.
  - It always has a .aa3 or a .oma file extension.
  - To open these files simply install the ATRAC3 drivers.
  - Used in Walkman and MD

Lossy Compressed Formats

- RA
  - Real Audio format designed for streaming audio over the Internet.
  - It can be played while it is downloading.
  - It is possible to stream RealAudio using HTTP, which works best with pre-recorded files.
  - Alternative protocols may work better for live broadcasts.

Other formats

- **MIDI (Musical Instrument Digital Interface)**
  - A simple scripting language and hardware setup
  - MIDI codes "events" stand for the production of sounds. e.g., a MIDI event might include values for the pitch of a single note, its duration, and its volume.
  - MIDI is a standard adopted by the electronic music industry for controlling devices, such as synthesizers and sound cards, that produce music.
  - Supported by most sound cards

Recommendations for WWW

- Portability between platforms, sound boards and software is determining factor
- Recommendation: audio/basic, 8-bit ISDN
- 8000 Hz sample rate, mono
- Also include MPEG audio version of audio clip for stereo support and CD-quality sound

Comparisons

- A detailed comparison of file formats is available from wiki:
  - http://en.wikipedia.org/wiki/Comparison_of_audio_formats

Question

- WAV → MP3, MP3 → WAV
  - Still the same quality?

File Format and Codec

- Difference between a file format and a codec.
  - A **codec** performs the encoding and decoding of the raw audio data;
  - The data itself is stored in a file with a specific audio file **format**.

Codec

- Codec: a device or computer program capable of encoding and/or decoding a digital data stream or signal.
  - The word codec is a blending of 'compressor-decompressor' or, more commonly, 'coder-decoder'.
- Lossless Codecs
  - archiving data in a compressed form while retaining all of the information present in the original stream
- Lossy Codecs
  - reduce quality by some amount in order to achieve compression

Audio Conversion Software

- Switch Audio Converter
  - http://www.nch.com.au/switch/
- Audio MP3 WAV WMA OGG Converter
  - http://www.audio-converter.com/
- Super Audio Converter
  - http://www.audioconverter.net/
- ……