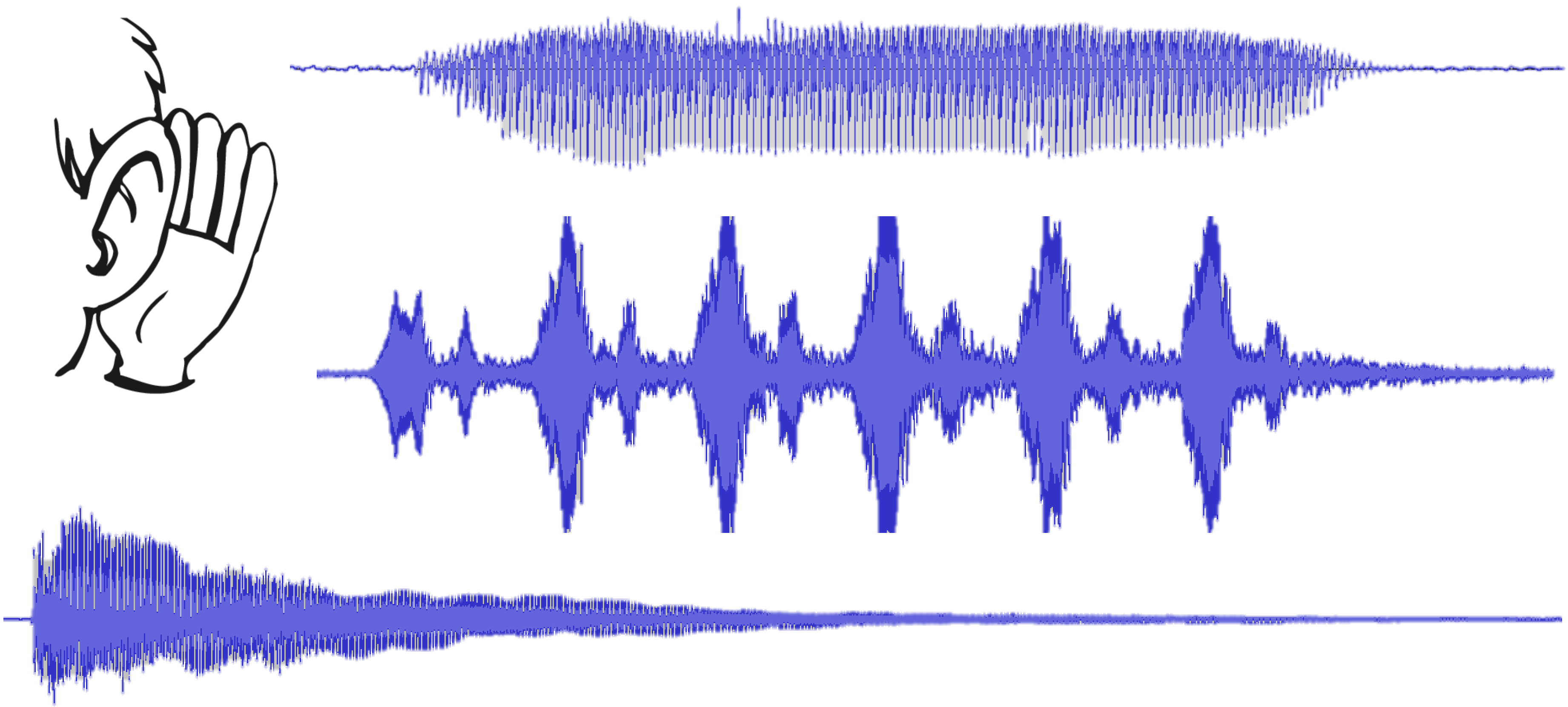
CS1102
Lecture 9
Digital Media: Audio

Semester A, 2020-2021
Department of Computer Science
City University of Hong Kong

Audio Waveform

- Audio can be represented as a waveform



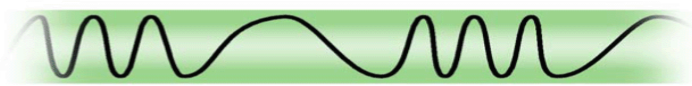
Analog vs Digital

- The real world represents information in analog

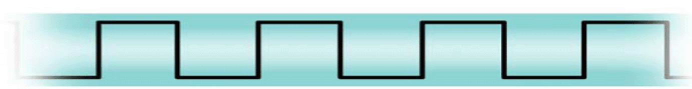
- Computer can only work with digital signals so we need to convert the real world information from analog to digital in a process called analog-to-digital conversion

Digital Audio

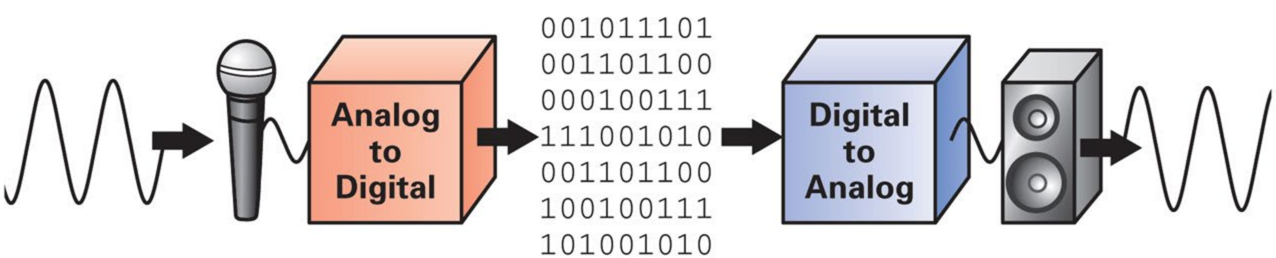
- Audio is originally an analog waveform signal, and it must be converted into discrete digital signal in order to be stored in and processed by computers



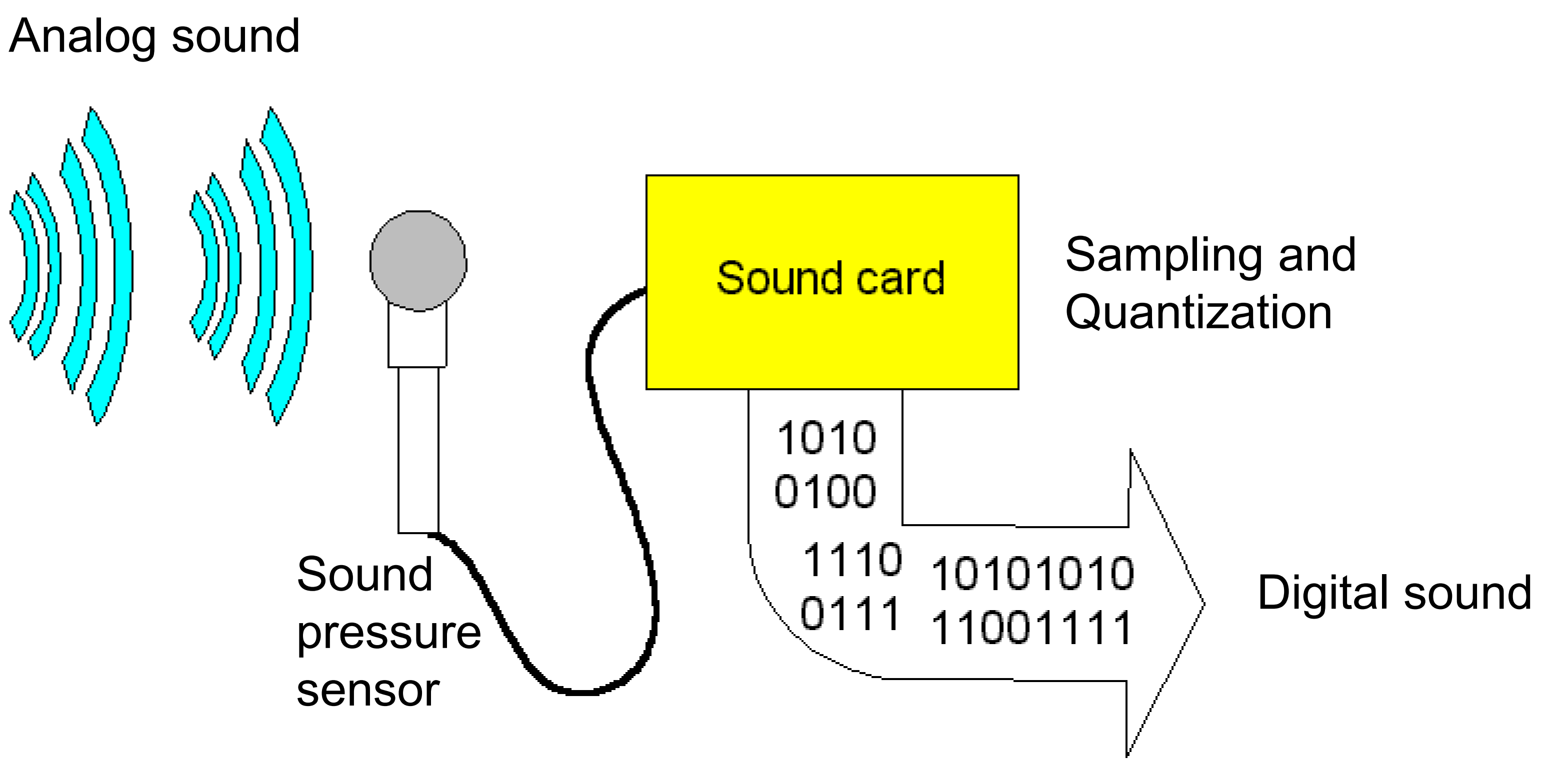
Analog (waveform) data signal



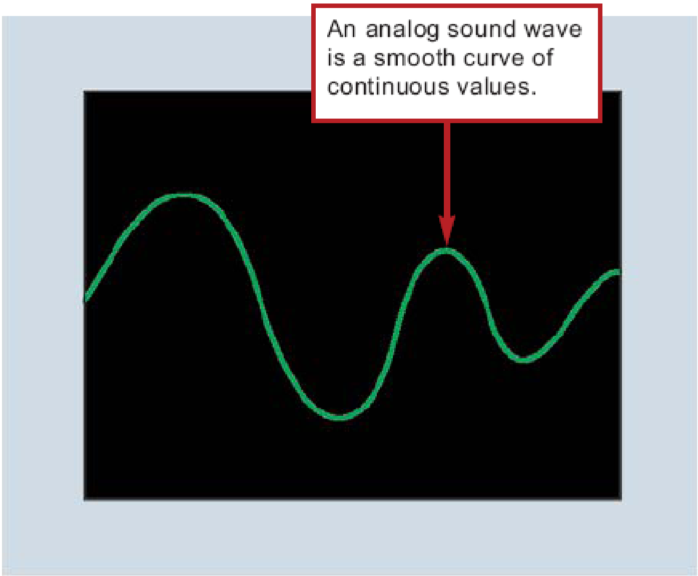
Digital (discrete) data signal

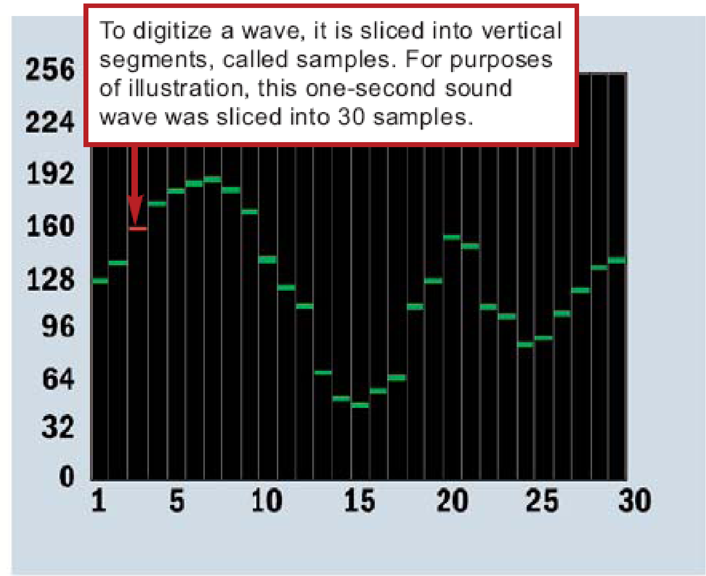


Analog-to-Digital Conversion



Analog-to-Digital Conversion (2)





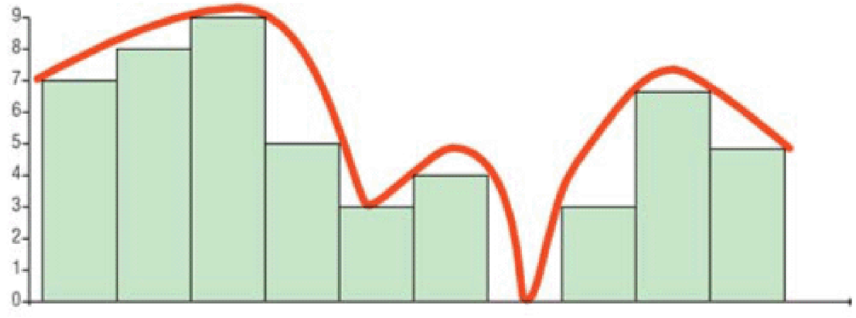
Sample	Sample Height (Decimal)	Sample Height (Binary)
1	130	10000010
2	140	10001100
3	160	10100000
4	175	10101111
5	185	10111001

The height of each sample is converted into a binary number and stored. The height of sample 3 is 160 (decimal), so it is stored as its binary equivalent—10100000.

Analog-to-Digital Conversion (3)

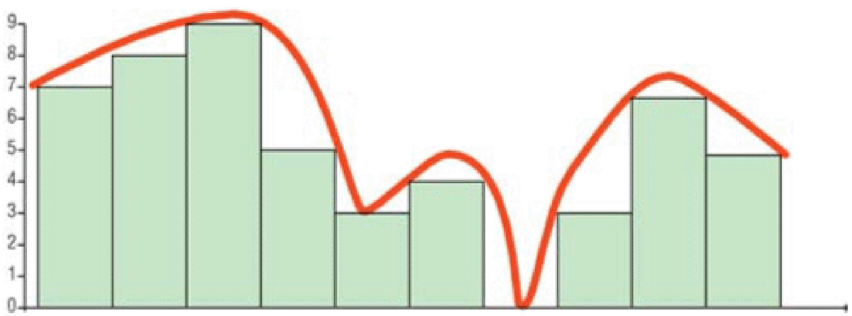
- Two steps in converting analog signals to digital signals

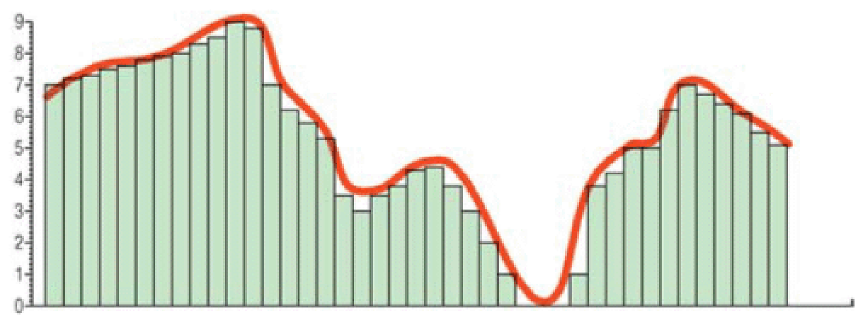
1. Sampling
   - Slice the analog wave into small segments regularly at uniform intervals, called samples (i.e. discrete time)
2. Quantization
   - Convert the height (amplitude) of each sample into a binary number that represents a specific amplitude which is an approximation of the original amplitude



Sampling Rate

- the number of times per second that a sound is measured during the converting (recording) process
- measured in Hertz (Hz)
  - e.g., A sampling rate of 10 Hz means that 10 samples are taken in 1 second





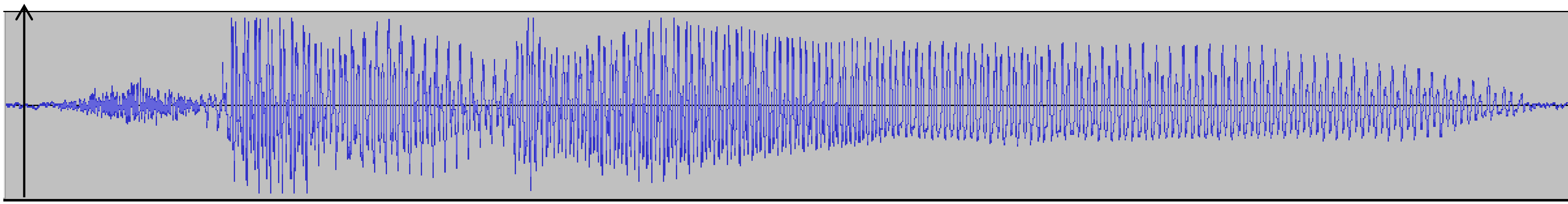
Which resulting digital signal is obtained with a higher sampling rate?

Sound Quality

- Sampling – how many times per second?
  - Human ear can hear sound from 20Hz-22,000Hz
  - Audio CD uses a sampling rate of 44,100Hz
  - Some very high quality (studio quality) audio uses a sampling rate of 48,000Hz – 96,000Hz
  - For voice/speech signal, a smaller sampling rate is good enough (e.g. 11,000Hz)
- Quantization – how many bits are needed for a single sample?
  - CD uses 16 bits per sample (65536 values, high quality)
  - Lower quality music uses 8 bits (256 values) per sample
- Number of channels
  - Mono – only one channel
  - Stereo – 2 channels (left, right)
  - Surround - 5.1 channels (left, right, center, left surround, right surround, and a base enhancement channel)

Amplitude

- The amplitude is obtained from the sound pressure



- The loudness level of sound can be calculated from the amplitude and the unit is in decibels (dB). This measure is relative to the quietest sound that humans can hear

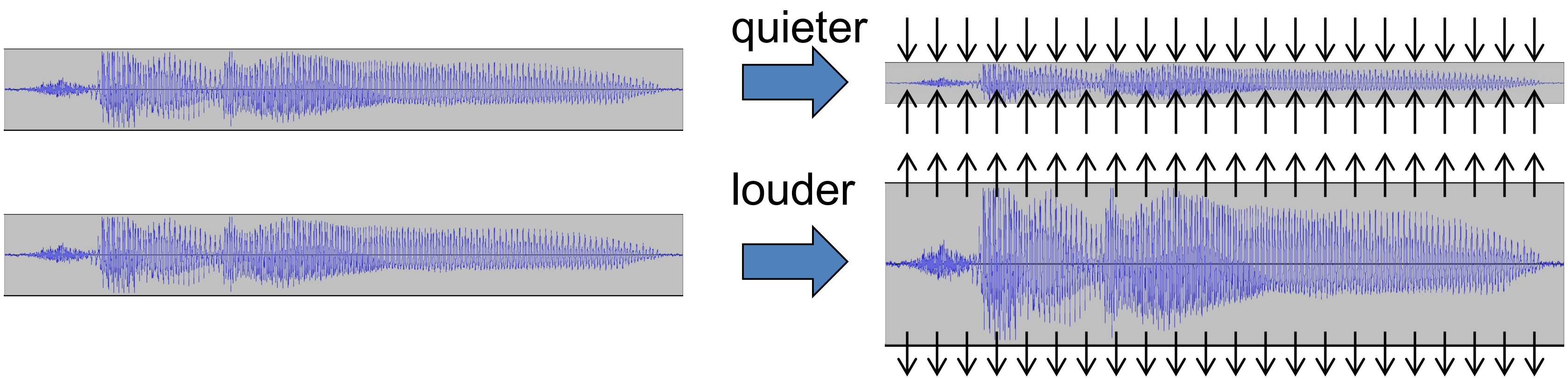
Common sounds	dB
Threshold of hearing	0
Rustle of leaves	10
Very quiet room	20
Average room	40
Conversation	60
Busy street	70
Loud radio	80
Train through station	90
Riveter	100
Threshold of discomfort	120
Threshold of pain	140
Damage to ear drum	160

Audio Editing

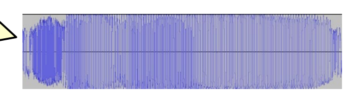
- Audio editing refers to the manipulation of the audio by changing its properties. Here we show some simple audio editing methods by changing its waveform
  - Change loudness
  - Reverse
  - Fade In/Out
  - Mix 2 sounds
  - Echo

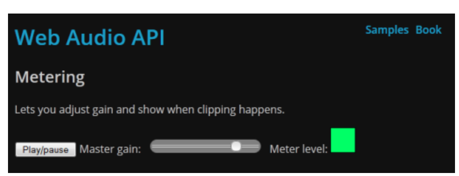
Changing Loudness

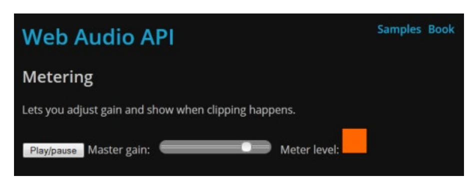
- You can change the loudness of an audio file by changing the amplitude of the samples in that audio file (without the need to tune up the volume of your loudspeaker)



If you try to change the amplitude to go over the maximum/minimum values, the sample will stay at the maximum/minimum value leading to some distortion known as clipping



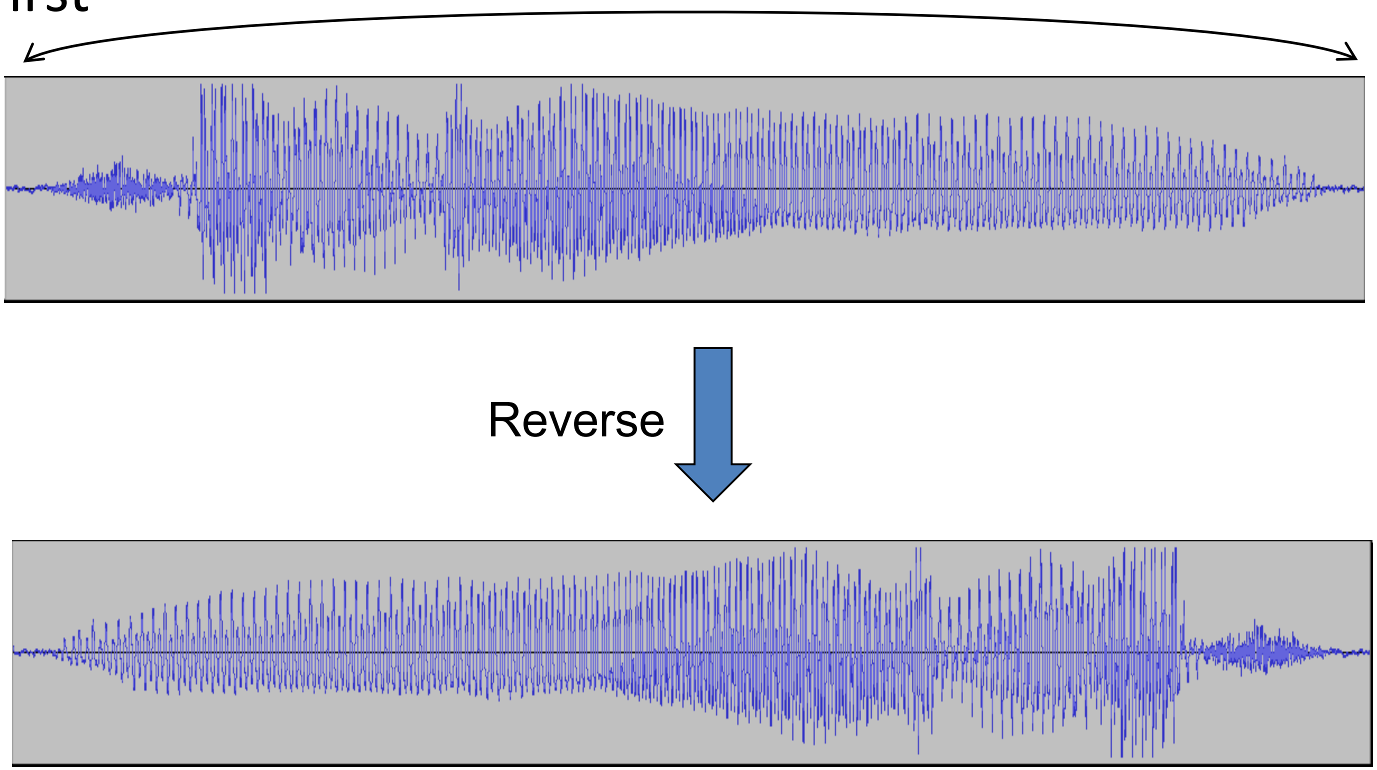




Demo: http://webaudioapi.com/samples/metering/ (Work with Chrome)

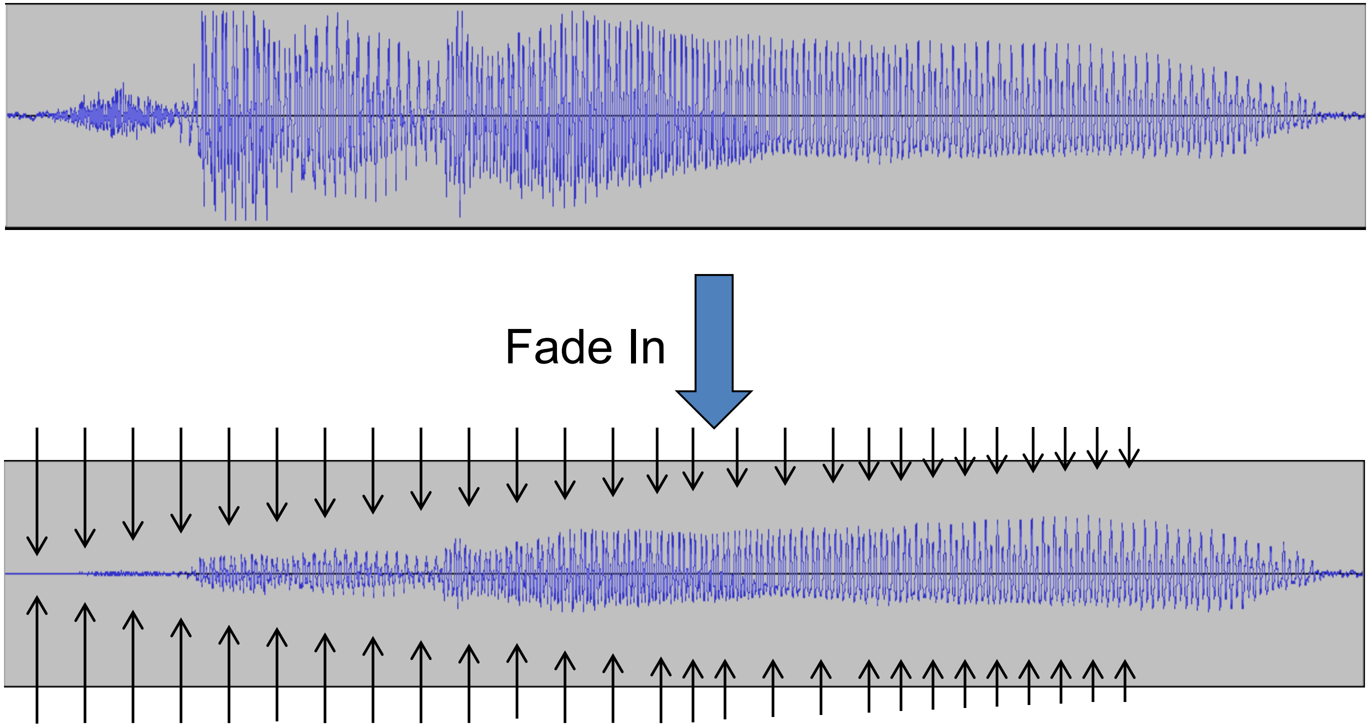
Reverse

- Reversing a sound is obtained by flipping it horizontally so that the first sample becomes the last and the last sample becomes the first



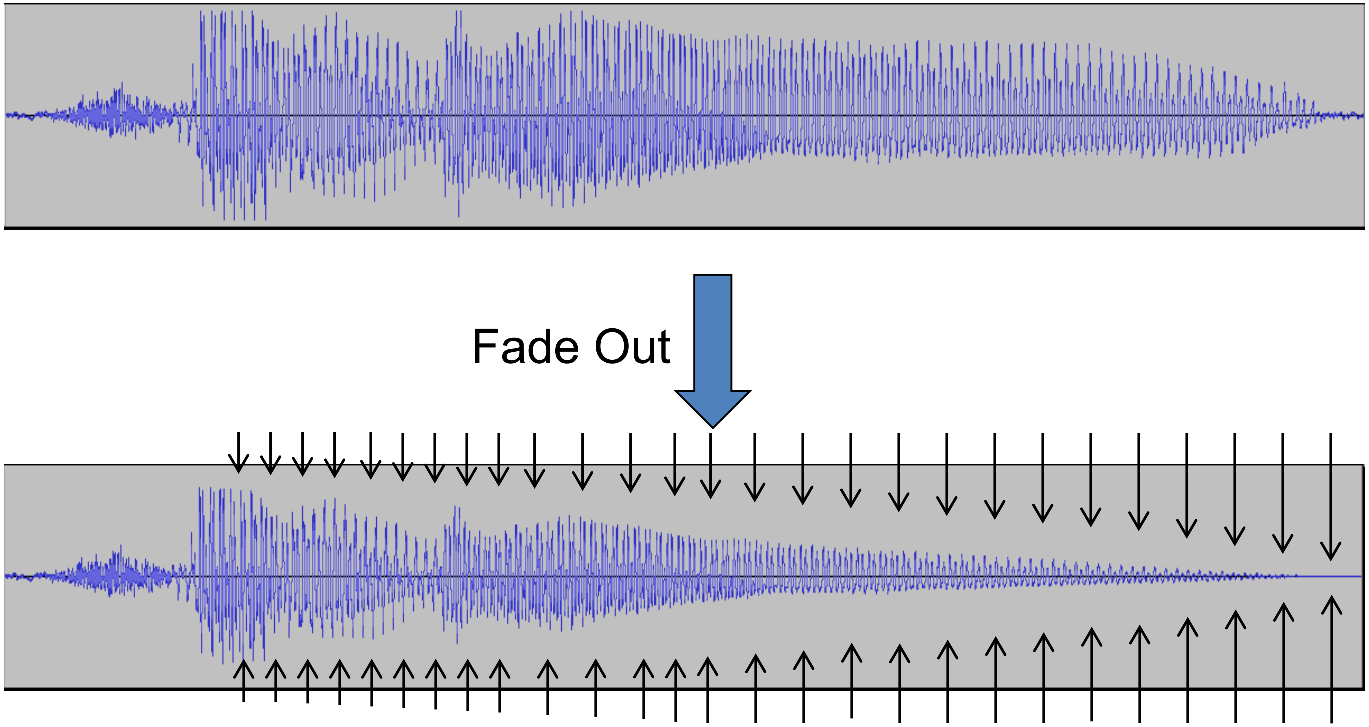
Fade In

- Decrease the amplitude of the beginning part of the audio. The earlier is the sample, the more is the decrease in amplitude



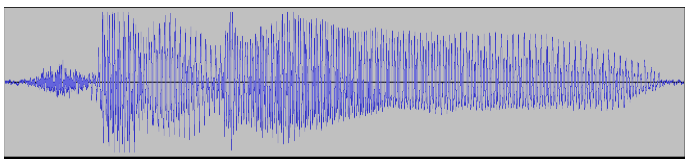
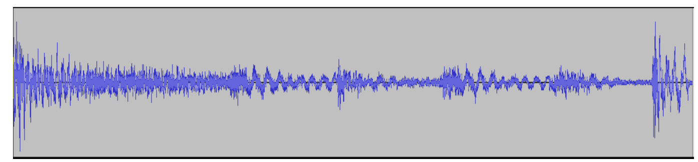
Fade Out

- Decrease the amplitude of the ending part of the audio. The later is the sample, the more is the decrease in amplitude



Combine 2 Sounds

- If you want to play 2 sounds simultaneously, ….

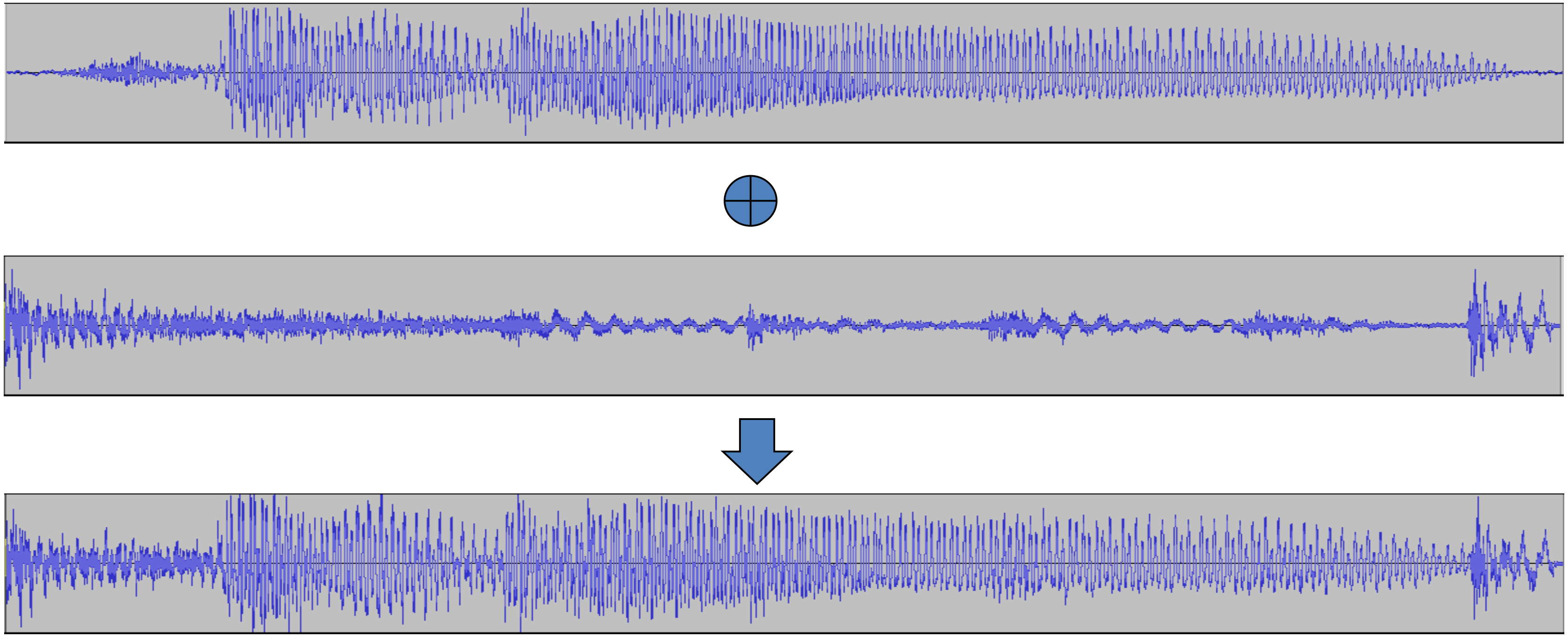
 

- You can use 2 devices with each device playing an individual sound

- Or you can mix the 2 sounds digitally to render a single sound file

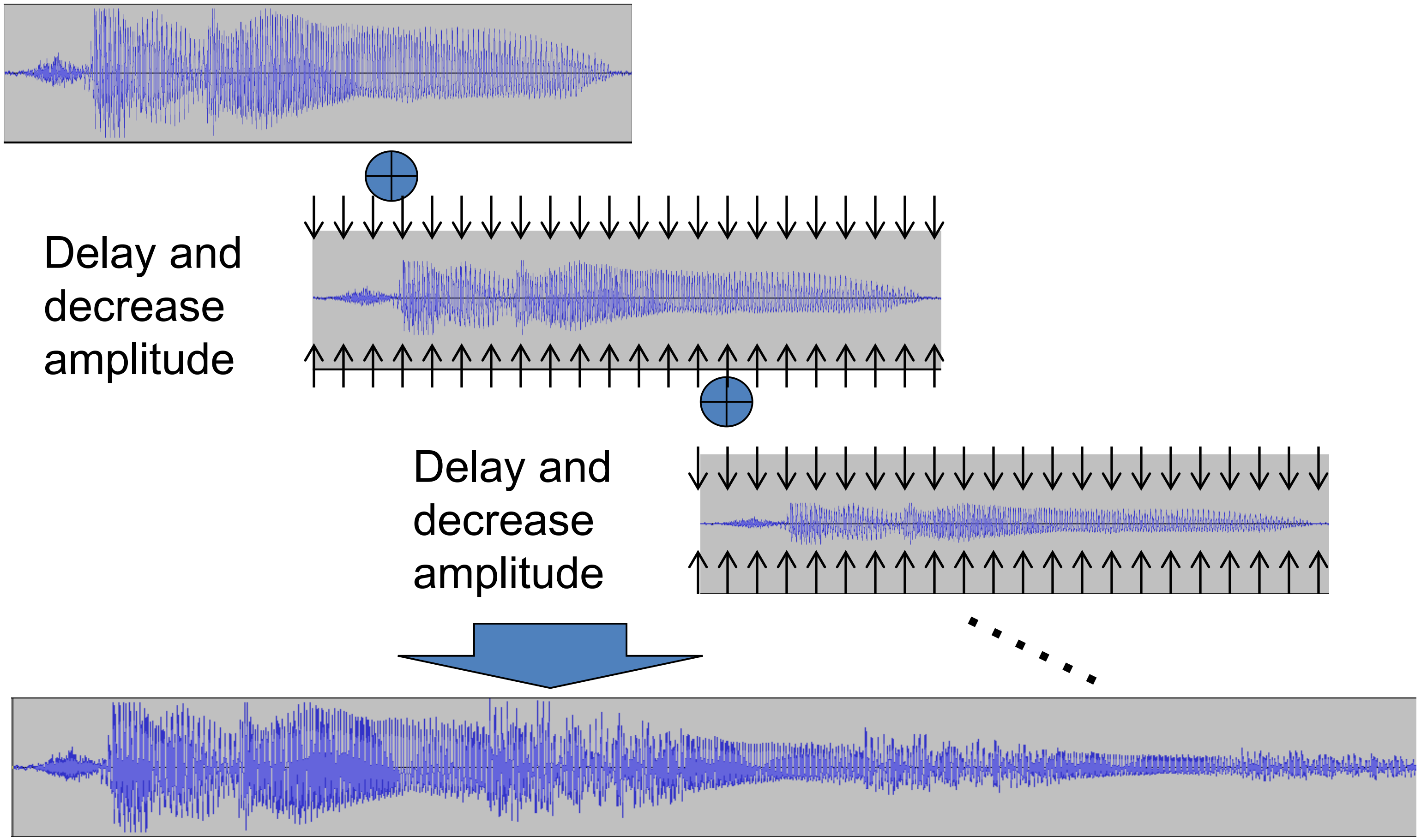
Mix 2 Sounds Digitally

- To mix 2 sounds digitally, you can simply add (or average to avoid clipping) them, i.e., add (or average) the amplitude of each sample from the 2 sound files



Echo

- The echo effect can be synthesized by applying the techniques that you have learnt



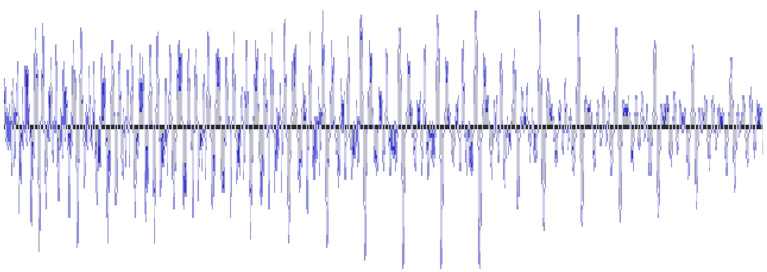
Pitch

- According to the American National Standards definition,

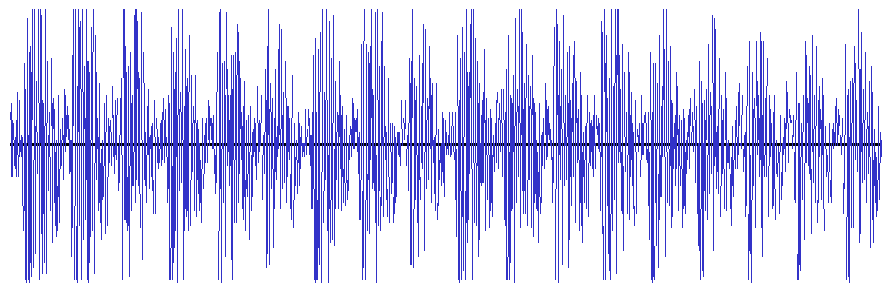
  "Pitch is that attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high. Pitch depends primarily on the frequency content of the sound stimulus, but it also depends on the sound pressure and the waveform of the stimulus"
- The pitch varies with the frequency of a pure tone, but studies suggested that the variation is not linear

Pitch with Real World Examples

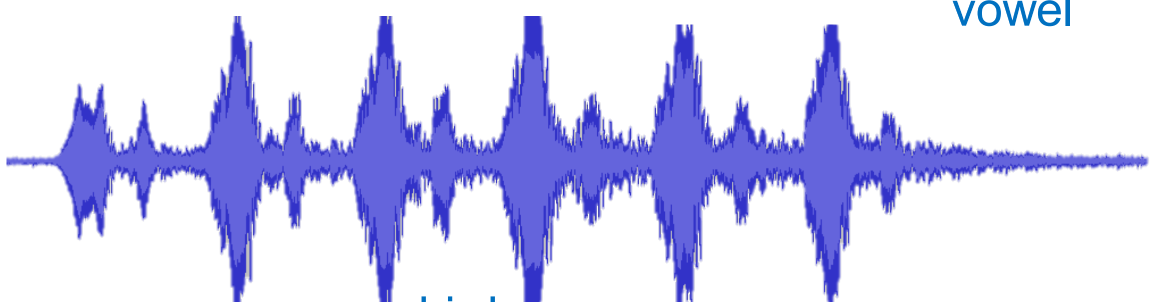
- Often the waveforms of the sounds around us show repetitive patterns. We often perceive these sounds as having a pitch corresponding to the repetition rate.



music



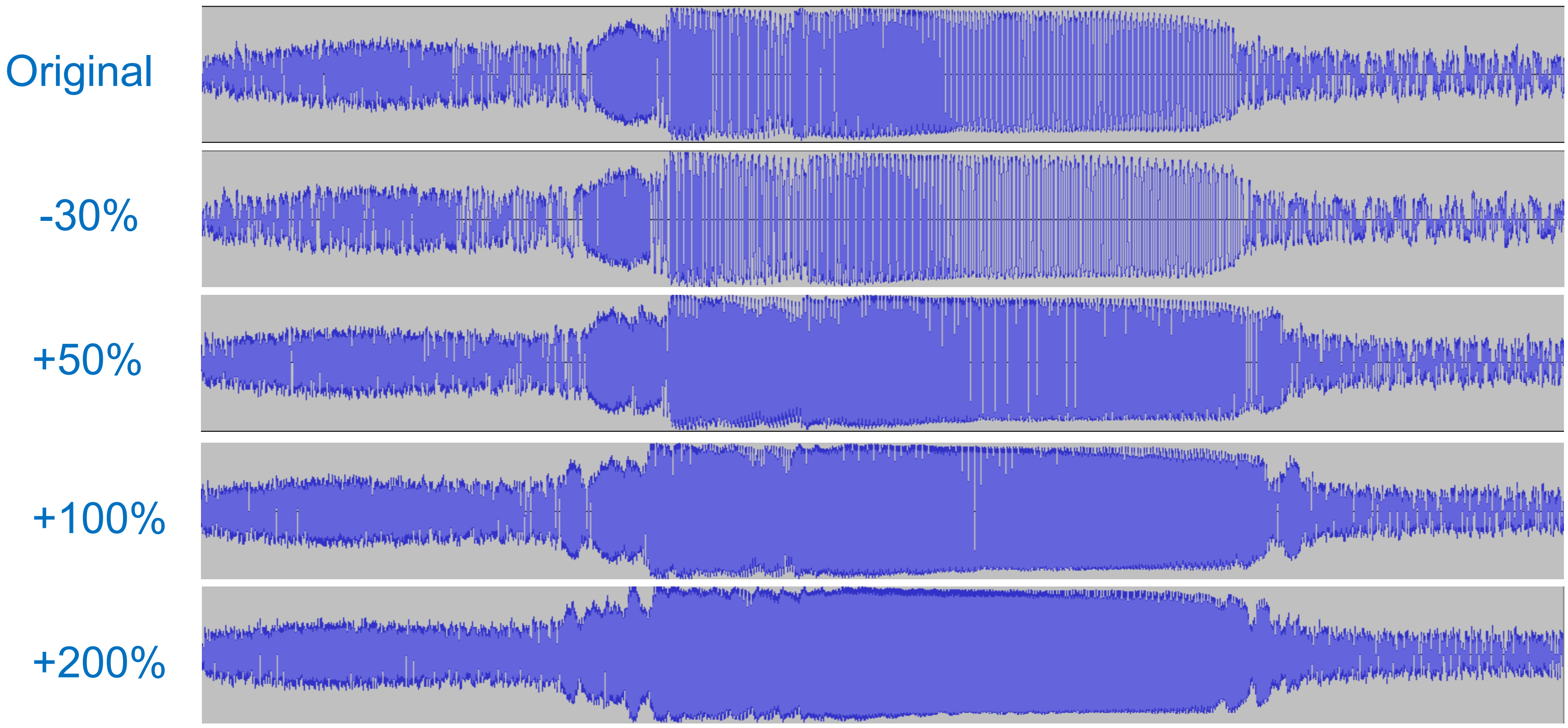
vowel



bird

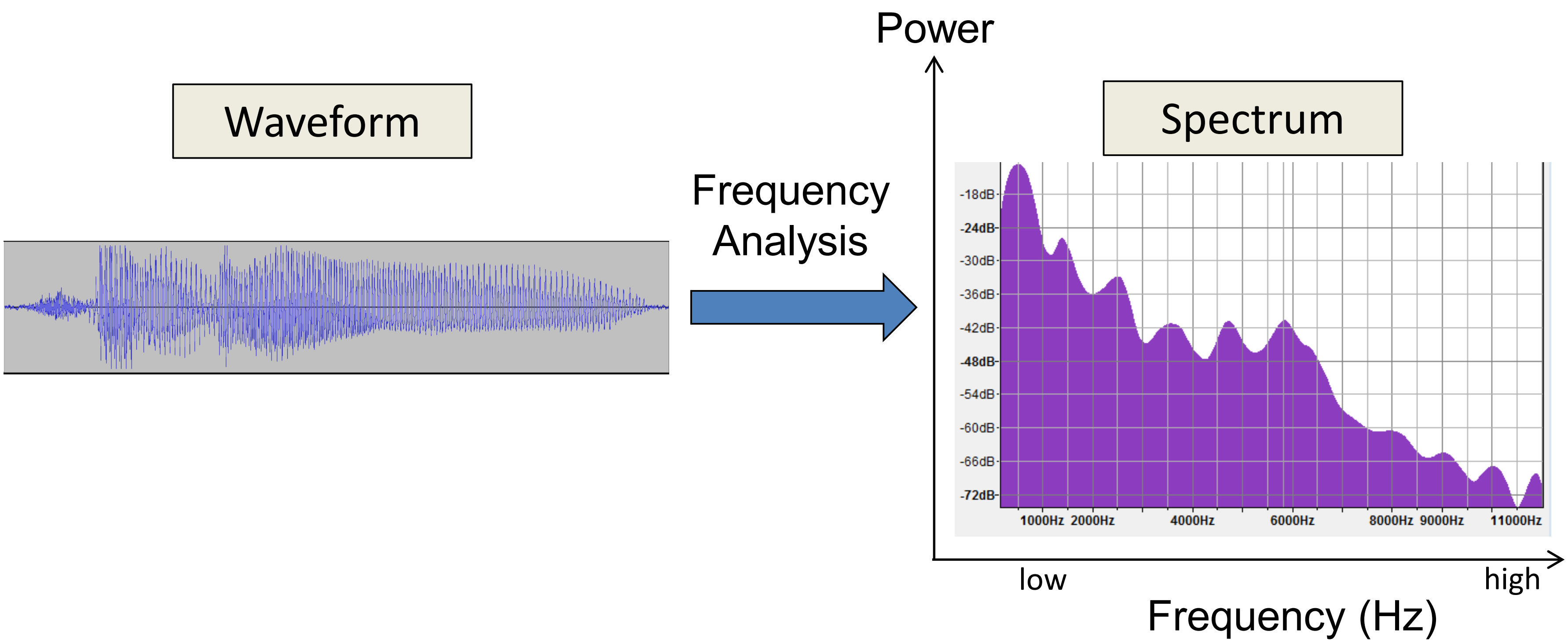
Pitch Shift

- Interesting effect can be created by raising or lowering the pitch of the voice.



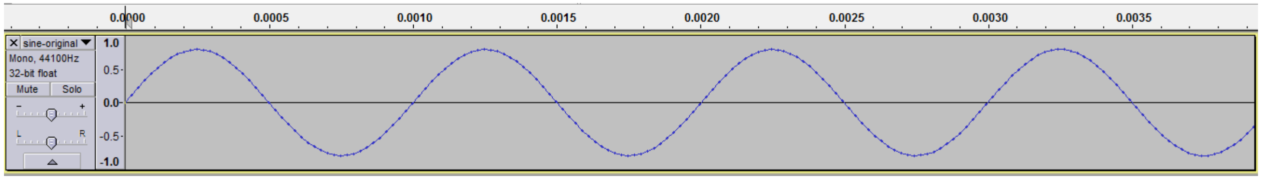
Spectrum

- Spectrum is the distribution of power (or energy) across different frequencies

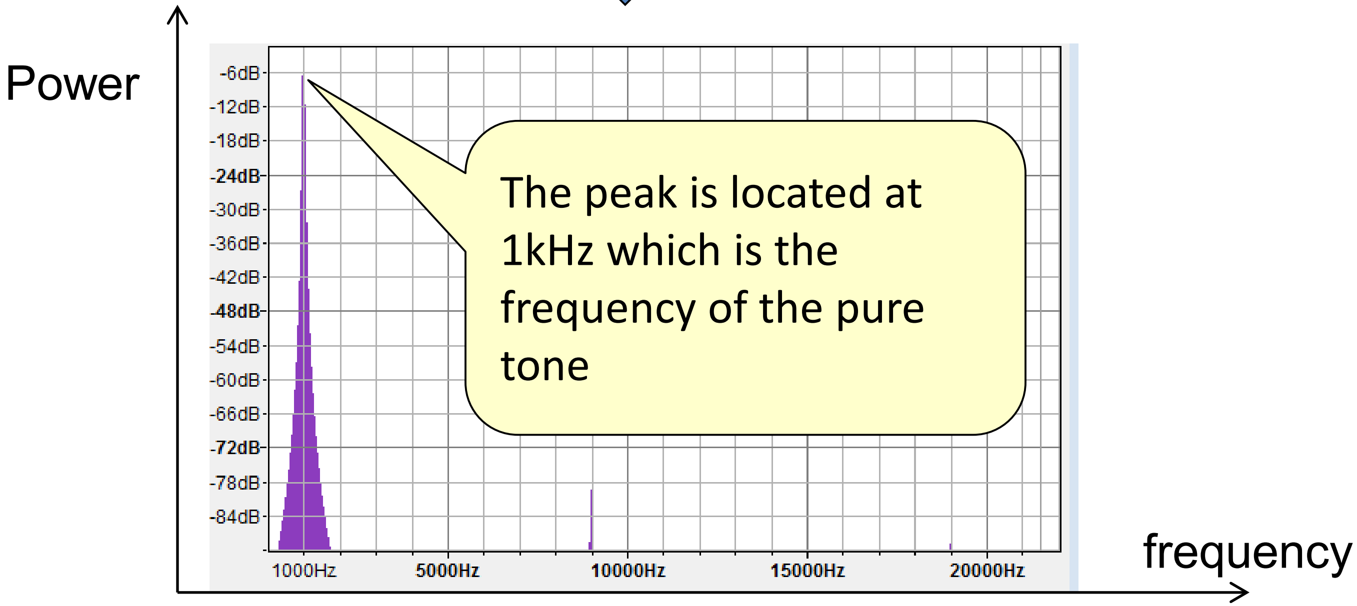


Spectrum Explained (1)

1kHz tone

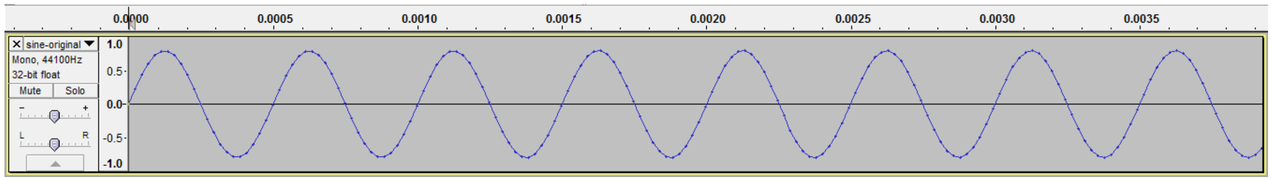


Frequency Analysis

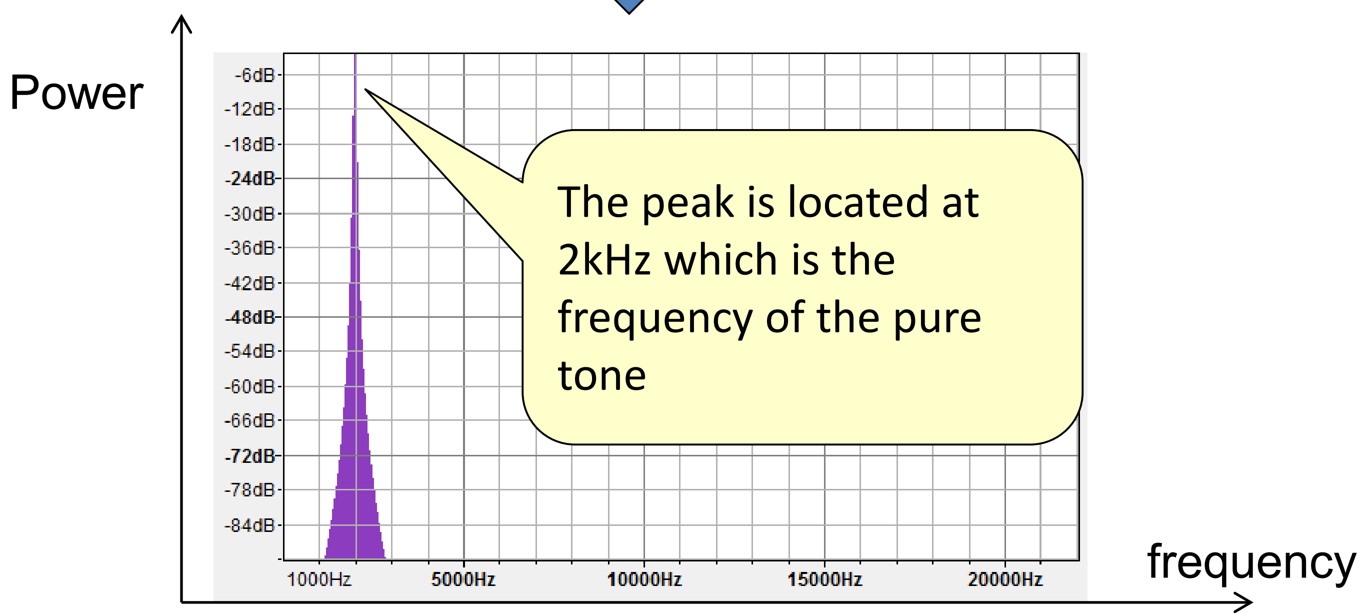


Spectrum Explained (2)

2kHz tone

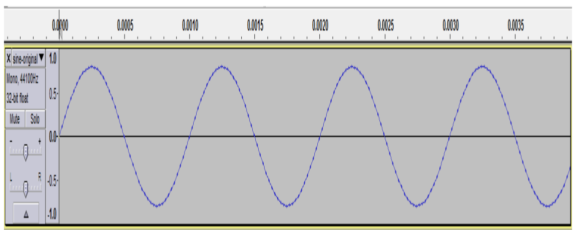


Frequency Analysis

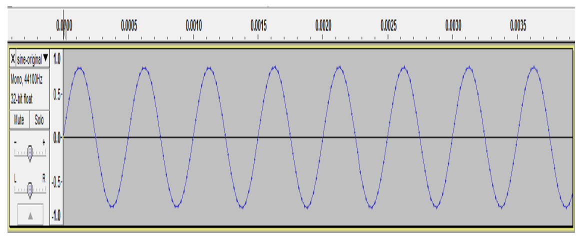


Spectrum Explained (3)

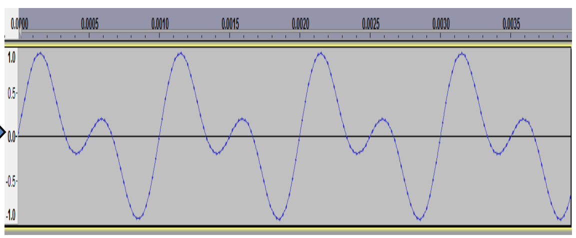
1kHz tone



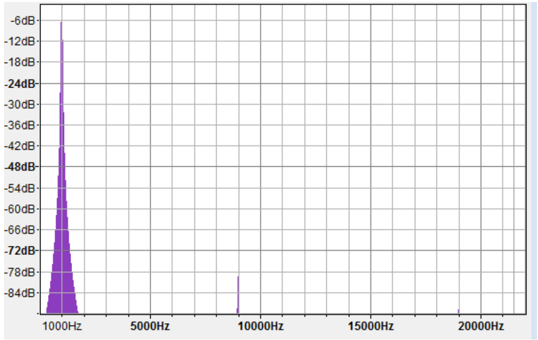
2kHz tone



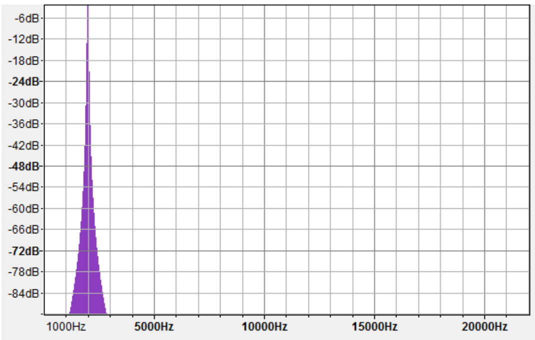
Mix of 2 pure tones



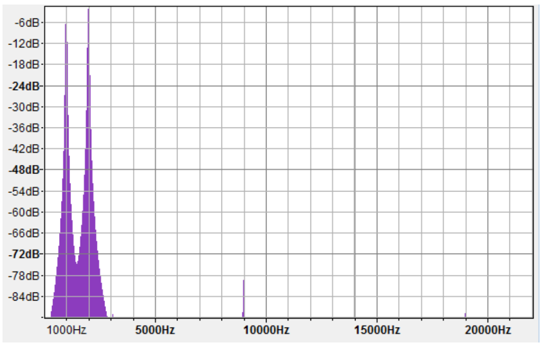
Frequency Analysis



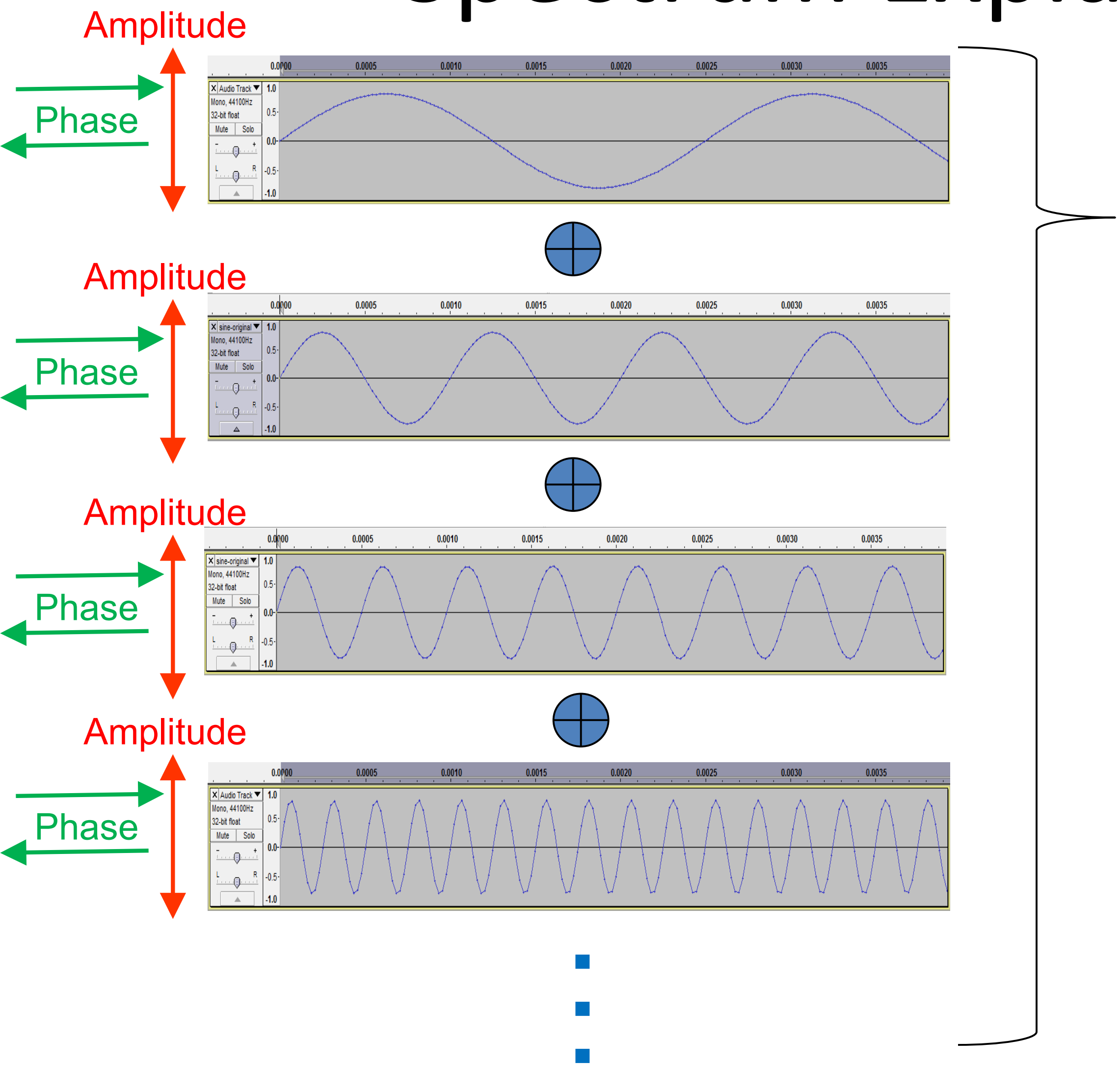
Frequency Analysis

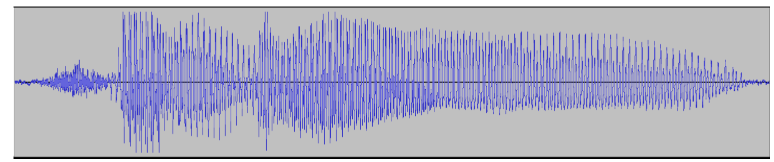


Frequency Analysis

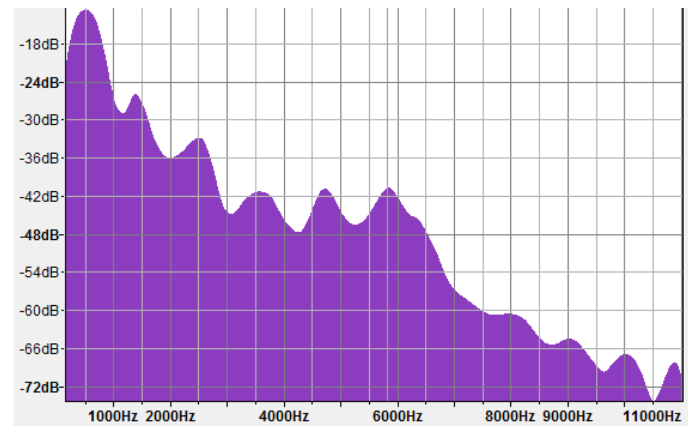


Spectrum Explained (4)





Frequency Analysis

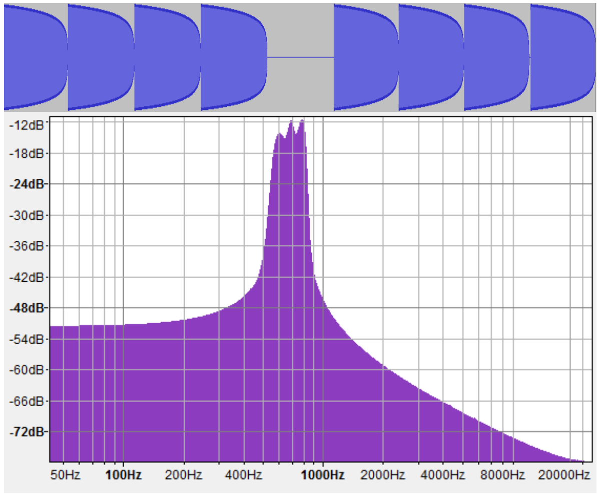


Spectrum is thus the distribution of power (or energy) across different frequencies

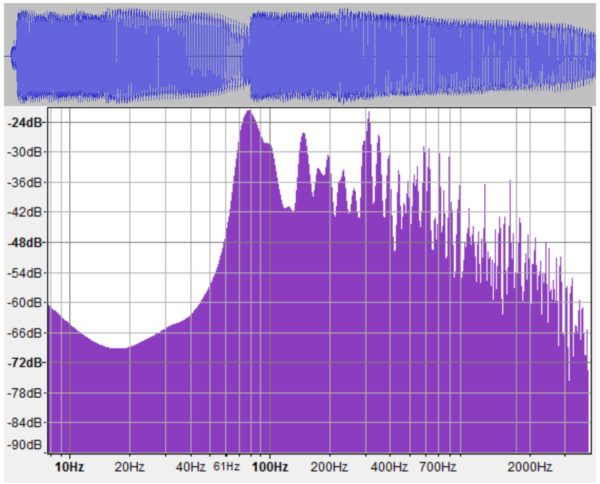
Example with Whistle

A peak can be seen for each tone. The higher the tone, the more right the peak is located, indicating a higher frequency

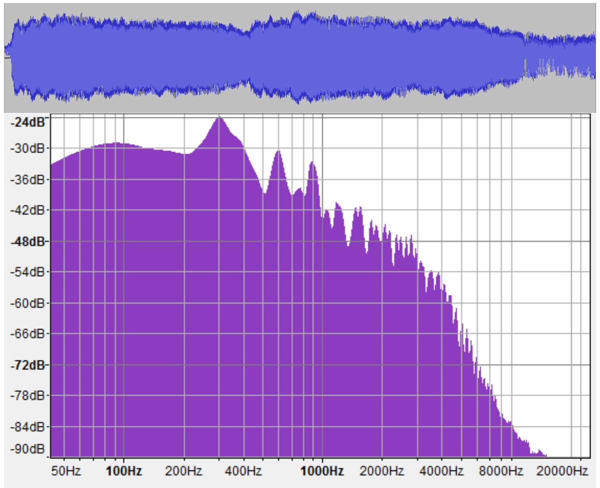
Example with Music



Simple Notes

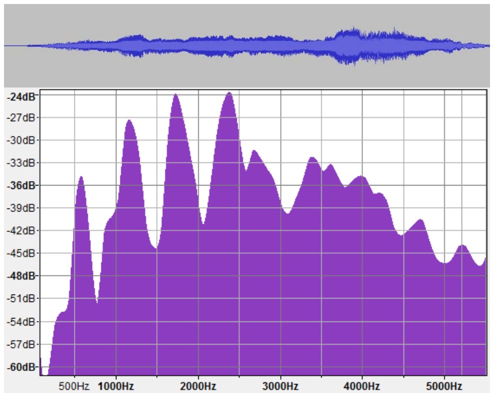


Piano

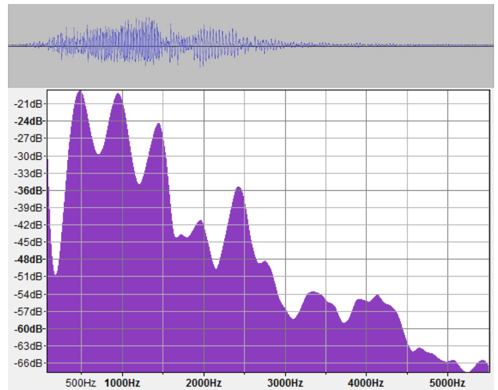


Orchestra

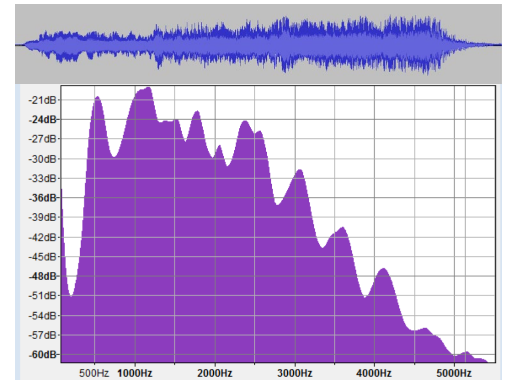
Example with Animals



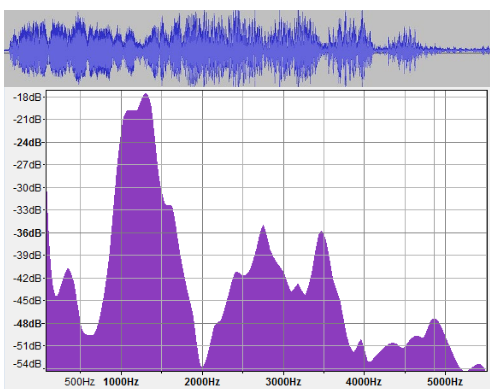
Cat



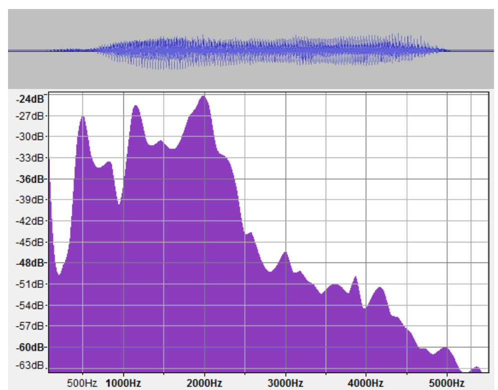
Dog



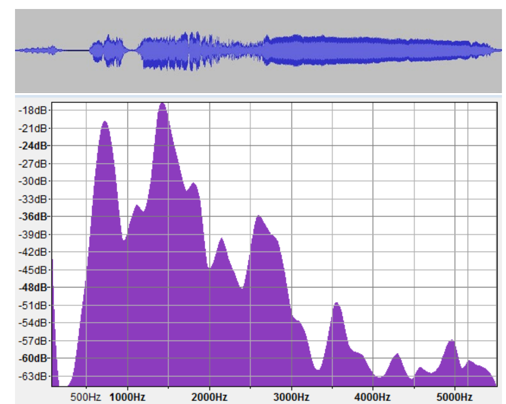
Elephant



Horse



Sheep



Rooster

Waveform Audio File Format

Audio Format	File Extension	Advantage	Disadvantage
Wave	.wav	Good sound quality Supported in browsers without plug-in	Audio data could be stored in raw, uncompressed format (PCM), so files are very large
MP3 (also called MPEG-1 Layer 3)	.mp3	Good sound quality even though the file is compressed Can be streamed over the Web	Requires a stand-alone player or browser plug-in or a HTML5-enabled browser
RealAudio	.ra .ram .rx	High degree of compression produces small files Can be streamed over the Web	Sound quality is not up to the standards of other formats Requires a player or plug-in
WMA (Windows Media Audio)	.wma	Compressed format, very good sound quality	Requires Windows Media Player
AAC (Advanced Audio Coding) (the successor to the MP3)	.aac	Lossy compressed format with good quality and small file size Default audio format of Apple's iPhone, iPod, iTunes	Heavily patent, may limit the usage potential
Ogg Vorbis	.ogg	Open, patent-free format with performance rivaling mp3. Supported by Android and many other projects	May not be supported by old portable players

The Need for Audio Compression

- What is the file size for an uncompressed audio file recorded with the following properties?
  - Sampling rate: 44,100Hz
  - Number of bits per sample: 16
  - Number of channels: Stereo
  - Duration of audio: 3 minutes

Duration of 3 minutes = 3 × 60seconds = 180 seconds

In 1 second, there are 44,100 samples per channel

Stereo means that there are 2 channels

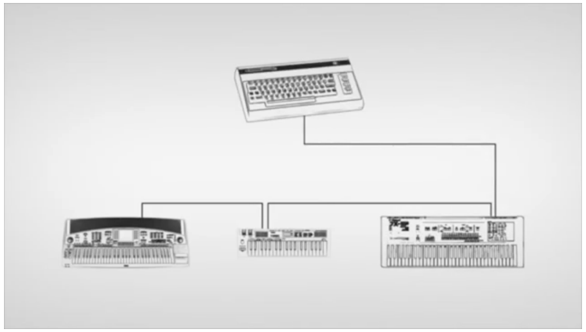
Each sample is represented by 16 bits

So file size = 180 second × 44,100 sample / channel / second × 2 channel × 16 bits / sample

= 254,016,000 bits = 31,752,000 bytes ~ 30.3 MB

MIDI

- MIDI stands for Musical Instrument Digital Interface
  - specifies a standard way to record, store and play back music on digital sound synthesizers
  - Unlike waveform audio files, MIDI files (.mid) do not represent sound directly but contain instructions on how the sound should be created
  - Information includes the pitch of a musical note, note-on time, note-off time, note volume, the type of the instrument
  - Most sound cards are equipped to generate music from MIDI files
- Advantage
  - MIDI files are very small
- Disadvantage
  - Does not produce high quality "real" sound



Lesson Summary

- Music, voice and sound effects can all be recorded and digitally stored as waveform audio, where the amplitude of the sound waves are sampled at small intervals (sampling) and stored as binary numbers (quantization). This process is known as analog-to-digital conversion
- Various sound effects can be created by manipulating the audio waveform
- Popular waveform audio file formats include WAV, MP3, RA, WMA and AAC
- Unlike waveform digital audio, MIDI music is artificially created and contains information about how the sound should be created. MIDI files usually has an extension of .mid

Reference

[1] Ze-Nian Li, Mark S. Drew and Jiangchuan Liu : Fundamentals of Multimedia, 2nd Edition, Springer, 2014, ISBN: 978-3-319-05290-8

[2] How do analog-to-digital converters work?

– http://www.planetoftunes.com/digital-audio/how-do-analogue-to-digital-converters-work.html

[3] What is a Sound Spectrum?

– https://newt.phys.unsw.edu.au/jw/sound.spectrum.html

[4] Audio File Format

– https://en.wikipedia.org/wiki/Audio_file_format