Audio Format

Tom Ko

How to represent a wave signal in computer?

- Signals in the air are analog

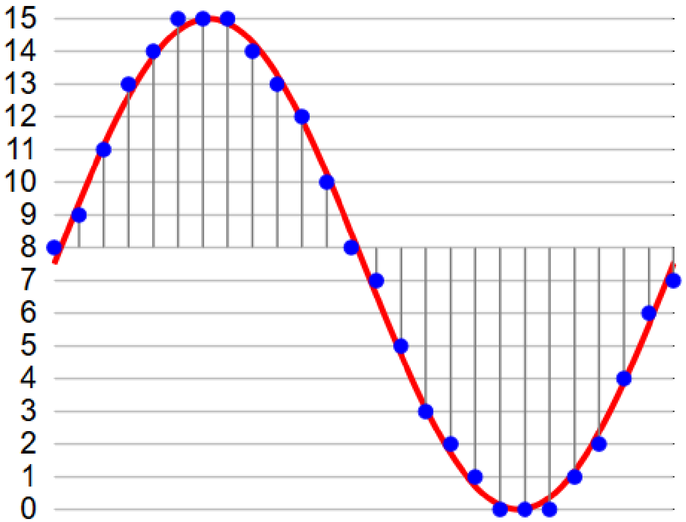
- It is a superposition of many waves.

- Digitize the signal
  - Sampling and quantization of the analog signal

Pulse-code modulation (PCM)

- **Pulse-code modulation** (**PCM**) is a method used to digitally represent sampled analog signals.

Quantization on the amplitude



Sampling on the time

Sampling frequency

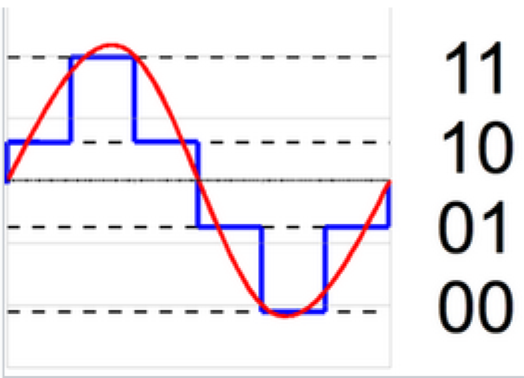
- The **sampling frequency** (or **sample** rate) is the number of **samples** per second in a Sound.
  - **8 kHz** for telephone channel
  - **44.1 kHz** for CD quality
  - **16 kHz** for ASR

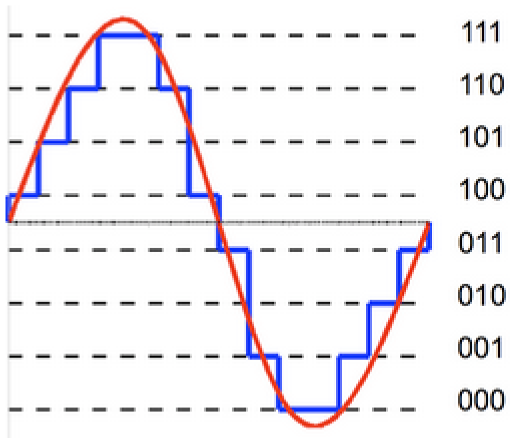
Nyquist frequency

- The **Nyquist frequency** is half of the sampling rate.
- It means that only signals with frequencies lower than the **Nyquist frequency** can be capture.
  - Use 8k Hz for telephone channel because human voice is below 4k Hz
  - Use 44.1k Hz because top limit for human hearing is about 20k Hz

Quantization

- It affects the resolution of the signal.





- How many bits we need to represent one sample point?
  - In ASR, it is common to store each sample with 16 bits.

Loss from quantization

Audio file format

- An **audio file format** is a file format for storing digital audio data
- It is important to distinguish between the audio file format (the container) and the audio data format (the content).

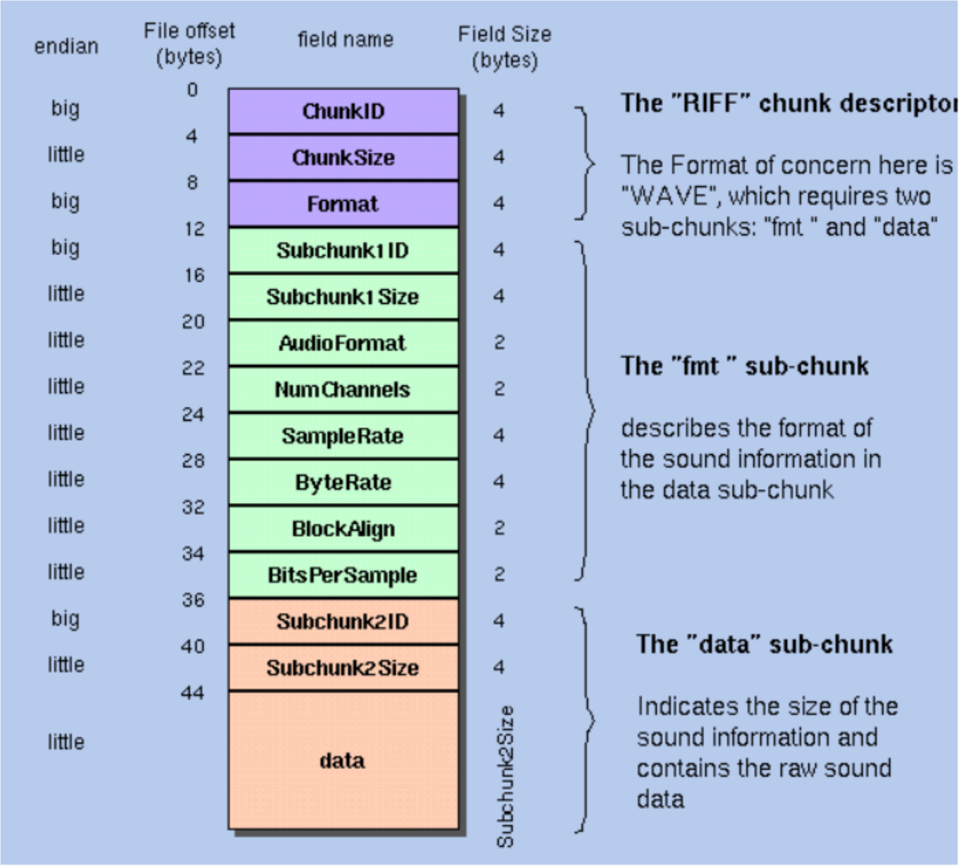
- Three major types of audio data format:
  - Uncompressed
  - Lossless compressed
  - Lossy compressed

Uncompressed audio format

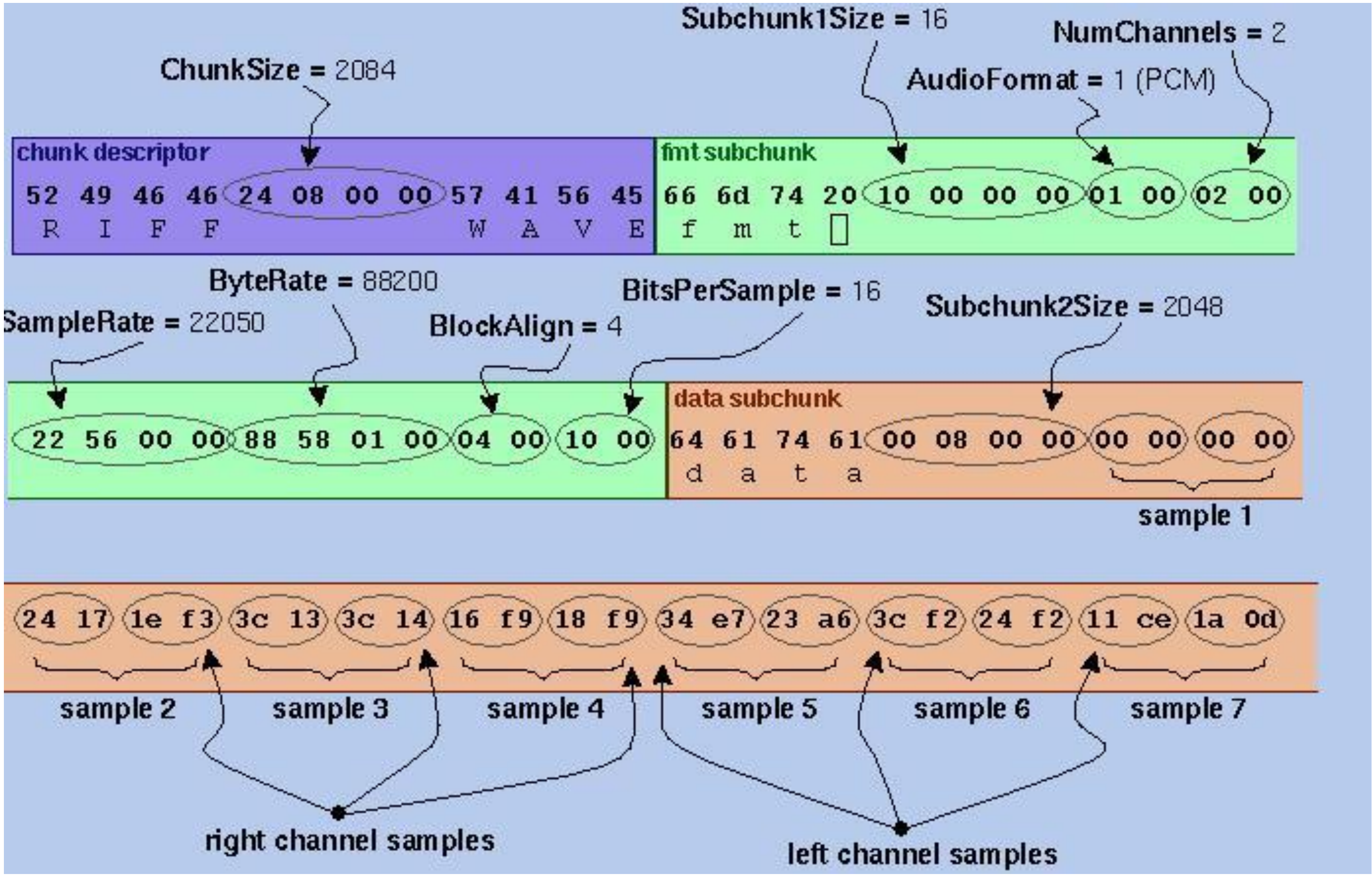
- **PCM is regarded as the uncompressed (raw) audio format**
- Uncompressed audio file formats, such as WAV, AIFF, AU or raw header-less PCM
- **WAV (Waveform Audio File Format) files are often indicated with .wav filename extension**

.wav files

- **Every .wav files has a 44-byte file header to store the metadata of the audio data.**

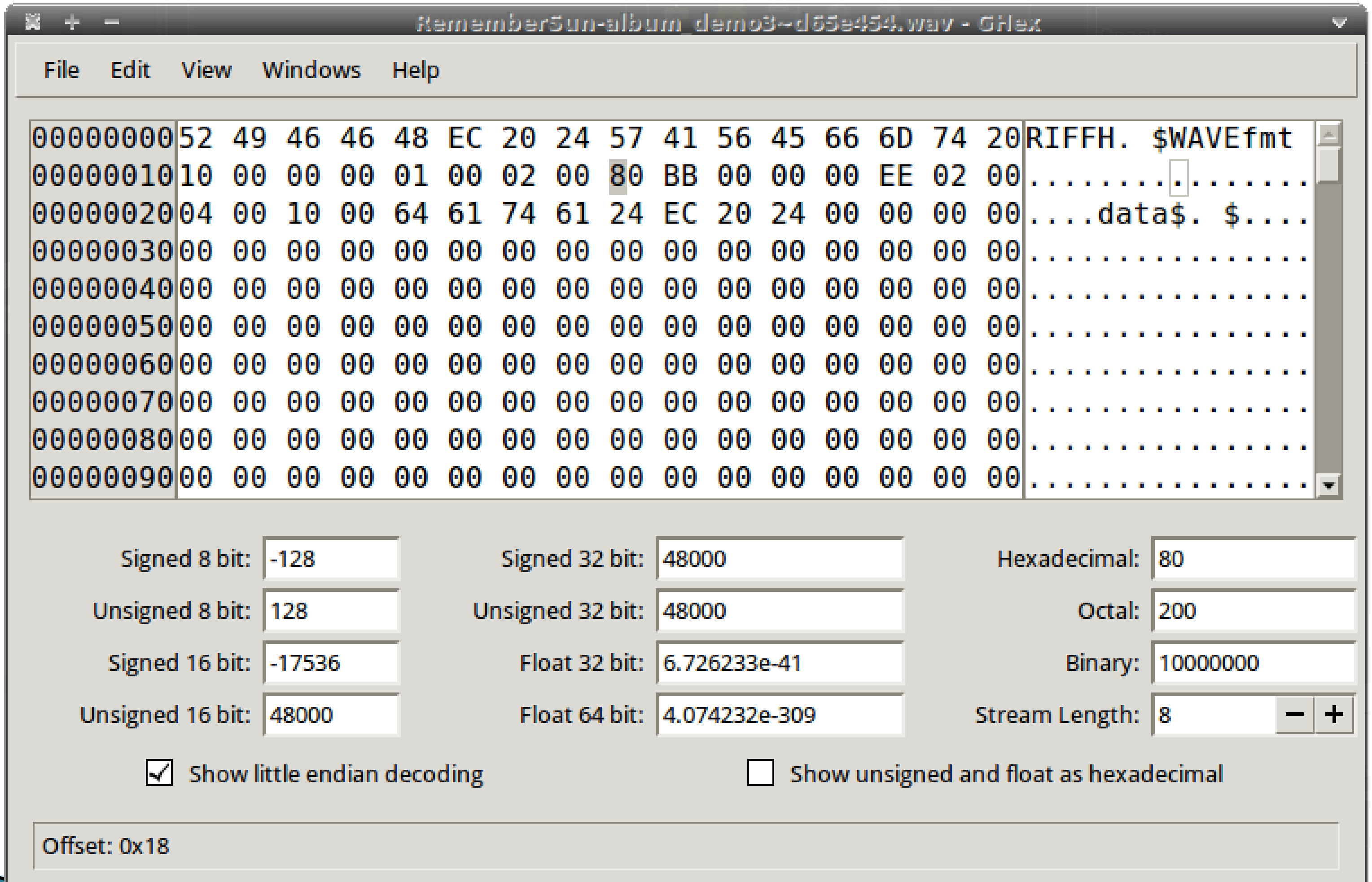


.wav files



ChunkSize = 2084
Subchunk1Size = 16
NumChannels = 2
AudioFormat = 1 (PCM)
chunk descriptor
52 49 46 46 24 08 00 00 57 41 56 45
R I F F W A V E
fmt subchunk
66 6d 74 20 10 00 00 00 01 00 02 00
f m t
ByteRate = 88200
SampleRate = 22050
BlockAlign = 4
BitsPerSample = 16
Subchunk2Size = 2048
22 56 00 00 88 58 01 00 04 00 10 00
data subchunk
64 61 74 61 00 08 00 00 00 00 00 00
d a t a
sample 1
24 17 1e f3 3c 13 3c 14 16 f9 18 f9 34 e7 23 a6 3c f2 24 f2 11 ce 1a 0d
sample 2
sample 3
sample 4
sample 5
sample 6
sample 7
right channel samples
left channel samples

.wav files


RememberSun-album_demo3~d65e454.wav - GHex
File
Edit
View
Windows
Help
00000000 52 49 46 46 48 EC 20 24 57 41 56 45 66 6D 74 20 RIFFH. $WAVEfmt
00000010 10 00 00 00 01 00 02 00 80 BB 00 00 00 EE 02 00 ....
00000020 04 00 10 00 64 61 74 61 24 EC 20 24 00 00 00 00 ....data$. $....
00000030 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ....
00000040 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ....
00000050 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ....
00000060 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ....
00000070 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ....
00000080 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ....
00000090 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ....
Signed 8 bit: -128
Unsigned 8 bit: 128
Signed 16 bit: -17536
Unsigned 16 bit: 48000
Signed 32 bit: 48000
Unsigned 32 bit: 48000
Float 32 bit: 6.726233e-41
Float 64 bit: 4.074232e-309
Hexadecimal: 80
Octal: 200
Binary: 10000000
Stream Length: 8
Show little endian decoding
Show unsigned and float as hexadecimal
Offset: 0x18

Question

- **If I change the sample rate of a .wav file from 8k to 16k without modifying the data content, what is the effect when I listen to the wave?**

Lossless compressed audio format

- **Similar to the zip compression**
- **FLAC**

Lossy compressed audio format

- **MP3,** stands for MPEG Layer 3
- **AAC in MPEG-4**
- A typical uncompressed wave file might be as big as 30 MB for a typical 3 minute song. But after being run through the MP3 compression algorithms that might drop down to 3 MB without any serious loss of quality.

MP3 file header

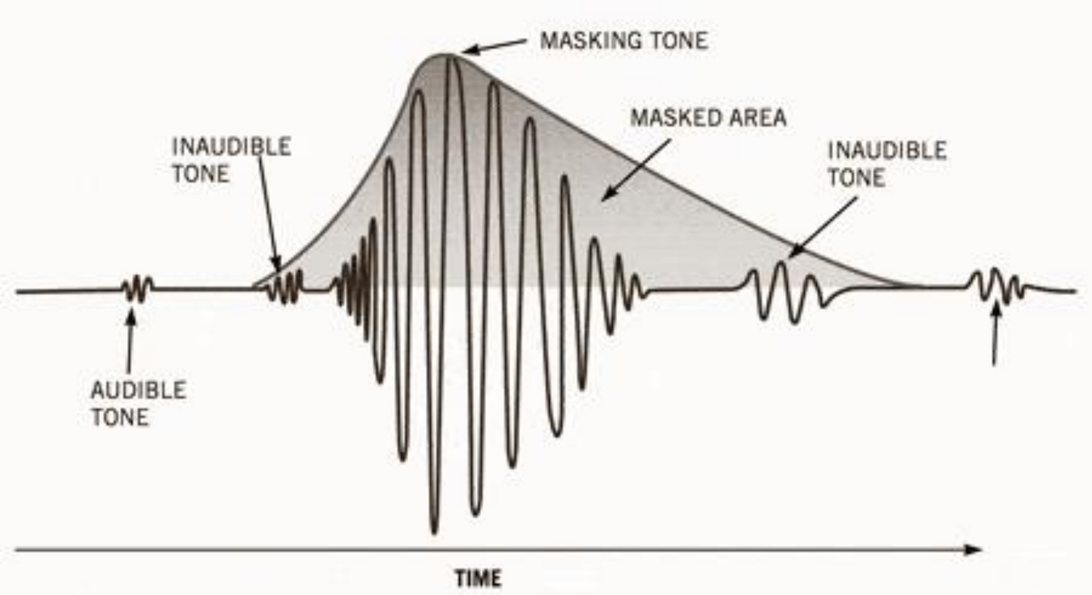
Detail of an MP3 Header

Bits	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Binary	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	0
Hex	F				F				F				B				A			
Meaning	MP3 Sync Word												Version	Layer		Error Protection	Bit Rate			
Value	Sync Word												1 = MPEG	01 = Layer 3		1 = No	1010 = 160			

21	22	23	24	25	26	27	28	29	30	31	32
0	0	0	0	0	1	0	0	0	0	0	0
0				4				0			
Frequency		Pad. Bit	Priv. Bit	Mode		Mode Extension (Used With Joint Stereo)		Copy	Original	Emphasis	
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	

MP3 algorithm

- Main idea: only store the data that affect how human hear the sounds



Useful tools in Linux

- **ffmpeg** – for converting audio format
- **Sox –** for a lot of audio operation