Audio formats

EECS 4462 - Digital Audio

November 5, 2018

Audio formats

- Audio files contain:
  - The digital representation of an audio signal, usually in PCM (also called the **bit layout** of the audio data)
  - Metadata (sample rate, bit depth, endianness, number of channels)
- The bit layout may be:
  - Uncompressed
  - Compressed in a lossless way
  - Compressed in a lossy way

Audio Codecs

- Codec: A portmanteau of coder-decoder
- A pair of software pieces
- Coder: Encodes PCM audio to a more compressed format
- Decoder: Decodes compressed audio back to PCM, so it can be played back
- Latency of the codec is important for real time applications
  - Latency is analogous to the number of samples that must be analyzed before a block of audio is processed

Uncompressed Audio Formats

- WAV (industry standard)
- AIFF (mostly on Macs)
- Others, like AU or raw PCM, are much less common
- They all use PCM
- WAV is little-endian, AIFF was originally big-endian, but little-endian versions exist now
- Space requirements: 10MB for 1 minute
  - Stereo, 44.1kHz, 16-bit

YORK UNIVERSITÉ UNIVERSITY

Broadcast Wave Format (BWF)

- The WAV format was extended in 1997 to include additional metadata useful for motion pictures, radio and TV production
  - Most notably a timestamp reference for synchronization with video
- The file extension is still WAV
- Audio players can handle either WAV type
- Need to be aware of this when extracting audio data from a WAV file

Compressed formats (lossless)

- Uses less space to store audio data without losing information
- The original, uncompressed data can be recreated from the compressed version
- Similar ideas to file compression are used
  - Silence can be compressed to almost no space at all
  - Next sample value is predicted, and only difference between actual and predicted value is stored
- Most common format: FLAC
  - Compression ration of about 2:1

Mid-Side coding

- A lossless alternative to stereo signal encoding is Mid-Side (MS) as opposed to Left-Right (LR)
- $M = L + R$ $\qquad S = L - R$
- $L = (M + S) / 2$ $\qquad R = (M - S) / 2$
- The side signal is often amenable to heavy compression
- Many professional audio plugins use mid-side processing as opposed to left-right processing

Compressed format (lossy)

- Removes some audio information to achieve larger compression ratios
- Exploits psychoacoustics to remove part of sound that cannot be heard: Perceptive Encoding
  - Weaker sounds that are masked by louder ones
  - Frequencies at the edges of human hearing
  - Bass frequencies may be stored as a mono signal as opposed to stereo
    - Humans are not great at determining the direction of bass
- Demo…

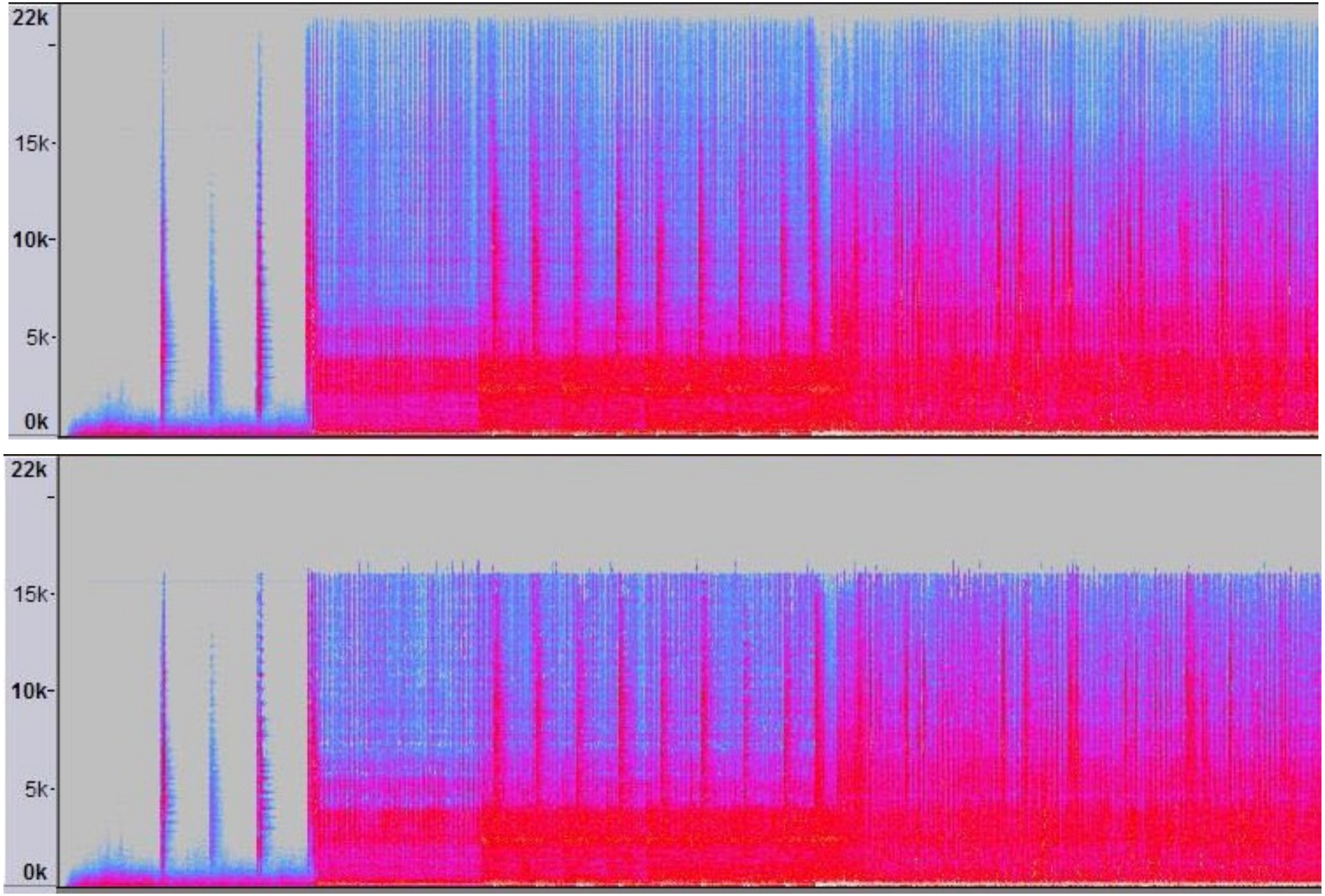
Bit rate

- The numbers of bits required to encode one second of audio
- For PCM, it's equal to
Sample rate X Bit depth X Number of Channels
- CD Audio has a bit rate of
44100 x 16 x 2 = 1,411,200 bits/sec = 1411 Kb/sec
- MP3 can be encoded at different bit rates up to
320 Kb/sec
- The less bits per second, the more information removed, the lower the sound quality

MP3

- Formally MPEG-1 Audio Layer III or MPEG-2 Audio Layer III
- Originally designed for speech rather than music
- Can achieve 75%to 95% reduction in size depending on bit rate
- Quality of encoded sound file also varies with bit rate
  - Difference is imperceptible for high bit rates
- Responsible for revolutionizing the music industry

WAV vs MP3 Spectrogram



https://en.wikipedia.org/wiki/Data_compression#/media/File:AudiodatenkompressionManowarThePowerOfThySword.jpg

AAC

- Stands for Advanced Audio Coding
- Used by iTunes, YouTube and many others
- Better sound quality than mp3 at the same bit rate
  - Better reproduction of high frequencies
- Supports up to 48 channels and sample rates up to 96kHz
- True Stereo

YORK UNIVERSITÉ UNIVERSITY

CBR vs VBR vs ABR

- Different kind of sounds can be compressed with different efficiency
  - Sounds with sharp attack, like a clap, cannot be significantly compressed without loss of quality, while silence or a drone can easily be compressed
- More sophisticated MP3 encoders can produce variable bit rate audio (VBR)
  - The audio is analyzed in chunks and an appropriate bit rate for each part is determined
  - The final file size of VBR audio is hard to predict
- In average bit rate (ABR) audio, the bit rate is allowed to fluctuate, but only around a pre-determined average value

Compression artifacts

- Logic Demo
- https://www.youtube.com/watch?v=DwpS7gOt554
- Video has several loud parts, so watch at low volume to begin with
- Note that subtracting two waveforms is not an accurate estimation of the difference between them due to phase issues