Lecture #3: Digital Music and Sound

CS106E Spring 2018, Young

In this lecture we take a look at how computers represent music and sound. One very important concept we'll come across when studying digital music is the difference between analog and digital. This is a fundamental issue that will occur anytime we take something in the real world and try to represent it on the computer.

As we'll see, representing music leads to tradeoffs between music quality and the space taken by the music files. Digitizing music immediately leads to a loss of information. In addition a variety of lossy and lossless compression techniques exist for making music files smaller, with the most popular file formats using lossy techniques based on psychoacoustics—the study of how human's perceive sounds.

Analog vs. Digital

- One item underlying our Digital Images lecture that I didn't really highlight is that Digital images are stored using discrete quantities that don't match the real world.
  - The real world does not consist of only 256 colors, 65,536 colors, or even 16.7 million colors, yet on the computer, that's how we represent them.
  - Similarly an image in real life does not consist of a set number of pixels, but it must inside the computer.

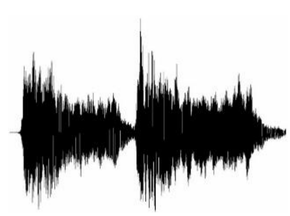
- In the computer, everything is represented by discrete numbers. This is what we mean when we say that computers store information as ***digital*** data.
- In contrast the real world consists of continuous non-discrete entities. We say that the real world is ***analog***.

- Studying how Digital Music works will give us a good opportunity to really see this distinction between discrete digital data and continuous analog real world entities.

How Music and Sound Work (in the Real World)

- Before we can see how music and sound work on the computer we need to understand how they work in real life.
- Suppose the Stanford Symphony Orchestra is playing in front of our classroom. Each instrument is creating vibrations of sound. These vibrations travel through the air and hit our eardrums and we hear music.

  - Here's what the sound wave for the famous initial two "short-short-short-long" motifs in Beethoven's 5^{th} Symphony looks like:



- In order to record music, we place a microphone in the room. That microphone senses the music's vibrations just as our eardrum did. The microphone records the intensity of those vibrations on a magnetic tape or other storage media.
    - If we want to store a stereo recording, we do the same thing but use two microphones and store two different tracks on our magnetic tape.

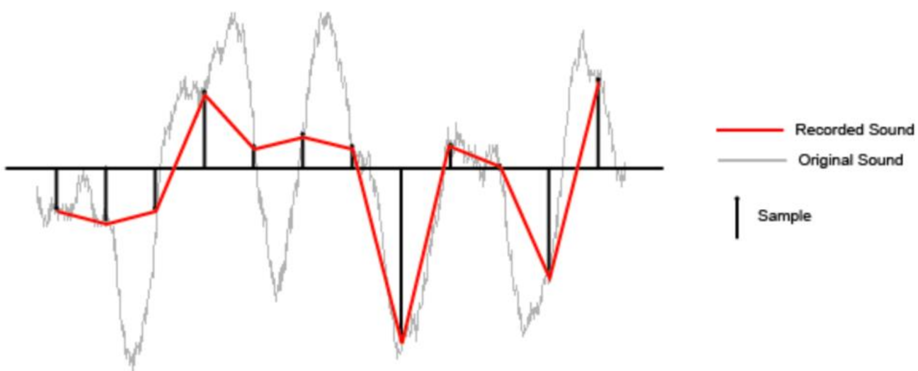
- To playback our music, we place speakers in the room. We play our magnetic tape and try to get the speakers to vibrate creating the exact same set of vibrations as the orchestra had created when the recording was made.

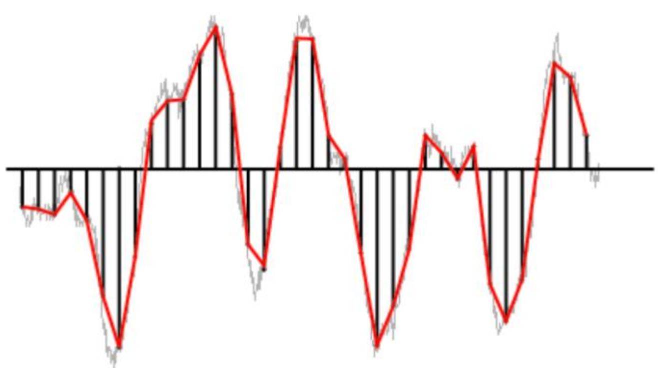
Representing Music Digitally

- We represent music digitally by storing a sequence of numbers, which represent the sound wave we are recording.

- Two features determine the quality of the recording
    - The Sampling Rate determines how frequently we store a number
    - The Bit Depth determines how much space we are setting aside to represent each of these numbers.

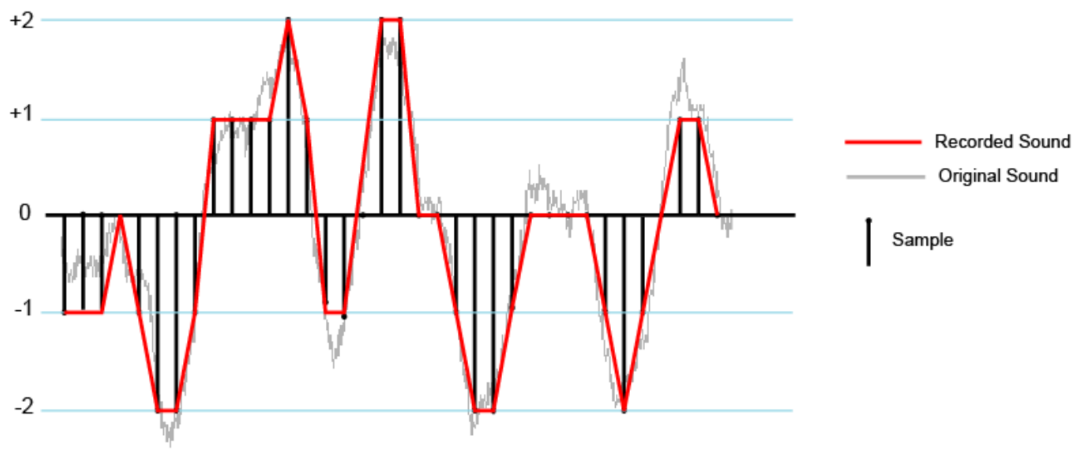
    - Let's take a look at how changing the Sampling Rate effects how closely our digital recording matches the original wave. Here are two different attempts to take the original analog sound wave for Beethoven's 5th Symphony and to take discrete samples to store in our digital recording allowing us to playback the Symphony at a later date.



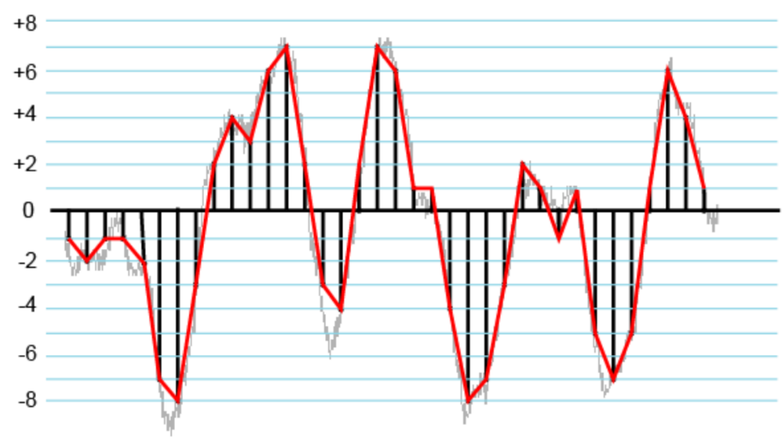


As we can see the second attempt at recording has many more samples taken, and as a result is a more accurate representation of the original wave.

- Let's see how changing the bit depth affects our ability to accurately represent the original wave. In this case we begin by allowing each sample to range from -2 to +2. This leads to the following set of samples:



As you can see this doesn't lead to a very accurate representation of the original wave. If we increase the range by a factor of four, storing numbers from -8 to +8 we can get a much more accurate reproduction:



(Note as we saw in the first lecture, binary representations of signed integers are actually asymmetric. The ranges here should really be -2 to +1 and -8 to +7, but I thought it would make the example to unwieldy.)

- As we can see, increasing the sampling rate increases the accuracy of our recording; similarly increasing the bit depth increases the accuracy of our recording.
  - The tradeoff here is that increasing these values also increases the size of our recording.

CD Audio

- ***CD Audio***, sometimes referred to as Red Book Audio:
    - takes 44,100 samples per second, so we say it has a sampling rate of 44.1kHz
    - uses 16-bits to store each sample, so each sample can range from −32768 to +32767
    - is in stereo, so we have two separate sequences of numbers, one for the right speaker and one for the left speaker.

- It is possible to increase the fidelity of our recordings. Various DVD-Audio variants exist which support sampling rates up to 192kHz, bit depths up to 24-bits, and up to 6 channels.

- Okay, so we're done right? We now know how digital music is represented right? Well no.
    - CD Audio format takes up a fair amount of space.
    - A 5 minute CD Audio song takes up about 50 megabytes
        - 5 minutes x 60 seconds per minute x 44,100 samples per second x 2 channels (for stereo) x 16 bits
        - This gives us 423,360,000 bits or 403.7 megabits
        - Which is 50.47 megabytes
    - When music first started being shared online, that 50.47 megabyte file would have taken over 5 hours to download over a telephone line Internet connection

MP3 (and Apple AAC and Windows WMA)

- MP3 is designed to store digital audio in a much more compact format

- We'll use two techniques to reduce the size of our original CD Audio file
    - First, we'll use ***psychoacoustics*** to reduce the information in the original file
        - We use something called a Fast Fourier Transform to take the original sound wave and convert it to a different representation of the wave which divides out all the different frequencies of the wave.
        - Psychologists tell us that not all the information in the original wave is actually noticed by human beings.
            - Humans can't hear very high frequencies, so we chop out all those frequencies.
            - If two frequencies are near each other in the frequency spectrum and one is really loud, we won't notice the quieter frequency, so we can toss it out.

    - After we've reduced the amount of information using psychoacoustics, we'll reduce it even further using something called a Huffman Encoding.
        - Huffman Encodings re-encode the information reducing the amount of space taken by considering which information occurs more frequently.
        - I like comparing it to Morse Code.
            - You may recall that Morse Code is used to transmit letters via telegraph. Probably the most famous Morse Code sequence is … --- … (dot-dot-dot, dash-dash-dash, dot-dot-dot) which represents SOS (or Save Our Ship).
                - … represents the letter S and --- represents the letter O
            - One interesting aspect of Morse Code is that letters in the alphabet that are transmitted frequently are encoded using shorter sequences.
                - e is represented as . (a single dot)
                - a is represented as .- (dot dash)
            - Letters that are used infrequently are encoded using longer sequences
                - q is represented as --.-

            - z is represented as --..
        - Using Morse code sending an e is four times faster than sending a q or a z.
    - Huffman Encoding does something very similar, it re-encodes so that frequently occurring data is encoded in less space than infrequently occurring data

- ***MP3 is a Lossy Format***
    - This should be obvious, since we know that the psychoacoustic part of the compression is throwing out some of the original information.

    - We can control the amount of loss, choosing how many bits per second to store.
        - common rates include 128 kilobits per second, 192 kbps, or 256 kbps

    - Let's compare a 5-minute MP3 file with our 5-minutes of CD-Audio.
        - 5 minutes of 128 kbps MP3 gives us:
        - 128 kilobits per second x 60 seconds per minute x 5 minutes = 38,400 kilobits
        - this is 4800 kilobytes which is a little over 4.5 Megabytes

        - this is less than $1/10^{th}$ of the size of our original CD-Audio file.

- Apple's AAC format and Microsoft Window's WMA formats work very similarly to MP3. Both AAC and WMA will result in smaller files than MP3 with better quality. However MP3 is more universally used.

FLAC and Lossless Formats
- While MP3 (and AAC and WMA) are all lossy, it is possible to reduce the size of a music file without loss
- FLAC stands for Free Lossless Audio Codec.
    - A file stored with FLAC has all the original information found in the original file.
    - There is no loss of musical quality in the conversion from CD Audio to FLAC.

- FLAC Compression Rates Vary
    - There is a tradeoff between amount of compression and time to compress/decompress
    - generally a FLAC file will be between ~65-70% the size of the original file

    - Comparison
        - 5-minute song CD Audio = 50.47 Megabytes
        - 5-minute song MP3 at 128 kbps = 4.5 Megabytes
        - 5-minute song FLAC (assuming fast compression time) = 35.67 Megabytes

- Other Lossless formats exist including Apple Lossless Audio Code (ALAC) and Windows Media Audio Lossless (WMA Lossless)

Analog vs. Digital
- Let's go back to our original discussion on Analog vs. Digital
- We can think of the original musical performance as consisting of a complex soundwave generated by the instruments passing through the air and reaching our eardrums.
    - This original soundwave is complex, analog, and continuous.
- Our representation of this original wave consists of discrete measurements representing the wave at various points in time.

- We ***digitize*** the music by taking that original complex, continuous wave and converting it to a sequence of numbers.
- This process of digitization happens with photographs as well. A digital camera takes the infinite shades of color and shapes and converts it.
    - The result is limited both by consisting of a set number of pixels and by limiting the colors to the 1.67 million available with 24-bit color.

- Note that in converting from the real Analog World to our computer's Digital Representation, we are losing information.
    - The digital representation is not as rich as the real analog world.
    - We can control the amount of loss.
        - With music, we can increase the sampling rate and use more bits for each sample.
        - With images (or video) we can increase the pixel density of our camera and can increase the number of bits we allow to store the intensities of each red, green, and blue color channel.

    - Ultimately though, regardless of how many bits we use to store the information, ***our digital representation is not as good as the original analog.***

Why Use Digital?

- This brings up a question, which is why are we using digital at all? If our digital music copy represents a loss from the original analog signal, shouldn't we just stick to our analog methods?

- The issue with analog is that analog reproduction is never 100% accurate.
    - Equipment used to reproduce a copy will somewhat distort the original signal.
    - The original recording itself can also degrade over time.
        - Analog copies of Hollywood movies for example, have faded, and are no longer the vibrant copies that they were when they had originally been created.

    - In contrast, with digital reproduction we're just copying a sequence of numbers. ***We can reproduce that sequence of numbers 100% accurately every time.***

Digital Rights Management

- In the 1970s, if I had a tape of music I really liked, I could make a copy of that tape for a friend, and their friend might ask for a copy of that tape copy, but by the time a friend of a friend of a friend of a friend copied the copy of the copy of the copy of the original tape, the quality of the reproduction would have dropped enough that they probably would have gone to the store to get an new, crisp, original copy.

- As we've seen in the digital world, that doesn't happen, each copy is as good as the original.
- Combined with the ability to put a copy online and share with people that don't even know you, media companies have a big potential problem.

- ***Digital Rights Management (DRM)*** systems are one potential solution to this. These systems encrypt the digital information and prevent copying the information.

- While music companies appear to have given up on DRM, it's still widely used by movie companies as well as electronic books.

MIDI

- There are other ways to store music, MIDI (Musical Instrument Digital Interface) provides an interesting example
- In a MIDI file, we wouldn't try to store the soundwave associated with a song, instead we would store the individual musical notes.
    - In addition, MIDI can store more detailed information such as the velocity at which a musical note should be struck or how it should be sustained.
    - MIDI information can be entered manually using a computer, in addition MIDI instruments allow a musician to play an instrument and have the notes recorded by software.
        - In addition to the obvious MIDI keyboards which resemble piano keyboards, MIDI guitar controllers, as well as wind controllers exist.
        - Wind controllers sense not only what notes the musician is holding the fingering for, but in some cases can measure lip pressure and strength of breath.
    - The quality of MIDI playback depends on the ability of the computer playing the file to synthesize realistic sounding instruments.

- While CD Audio (and derived methods such as MP3 and FLAC) will generally provide a richer listening experience, MIDI has some real advantages of its own.
    - For a composer, the ability to see the notes on the screen and to be able to change and manipulate them gives the MIDI format capabilities completely unavailable with CD Audio.
    - In addition for someone storing music files as part of a large experience (say for example as part of a game or a larger multimedia experience), MIDI files are much smaller than their CD Audio counterpart.
        - Beethoven's Fur Elise, for example, contains perhaps 5,000 notes. Each note can be one of 88 values (the number of keys on a piano) and we can store 88 combinations in 7-bits.
            - So we can store Fur Elise in roughly 5,000 bytes (we would probably actually need more to store note length and other details). But certainly we could store it in 20-25 kilobytes or less.
            - In contrast Fur Elise is an ~3 minute song. A CD audio file of Fur Elise would take over 30 Megabytes
            - An mp3 copy using 128 kbps would take about 2.8 Megabytes