Digital Sound
Jim Williams
HONP-112
Week 9

What we will cover

- Part 1: Understanding Sound
- Part 2: Digitizing Sound
  - Sampling Rate
  - Sample Size
  - Problems/conversions/calculations/etc.
- Part 3: Compression

Part 1: What is sound

- Before we can come up with a way to digitally represent sound, we need to understand what sound is.
- Then we have to break down into the smallest meaningful part, determine how many possible values there can be for each individual part, string them together and we have our representation of a sound.
Sound

- A compression wave is a rhythmic disturbance (or vibration) in a medium (for our discussion, our atmosphere – or “air”).
- It is important to understand that these vibrations move in **two directions** (forward and backward).
- Remember that a single disturbance will propagate other disturbances (how many will depend on how “intense” the initial disturbance is). These vibrations can “travel” through air at approximately 1125 ft/sec.
- There are certain vibrations that can be detected as “sound.” This depends on how often a single cycle of the vibration happens each second.

Sound Detection

- If a **complete forward/backward cycle** happens anywhere from 20 to 20,000 times in a second, this vibration can be detected by the human ear as “sound.”
- The eardrum acts as a “transducer”, transforming these vibrations through the small inner ear bones and cochlea into sound “signals” that are carried by the auditory nerve to the brain.
- Any vibrations outside this range (subsonic, ultrasonic) are not detectable as sound by humans (though this is different for other animals).

What is a Cycle

- As stated earlier, a compression wave moves the air, from a certain static point, forward, then backwards, then returns to starting point.
- Each full movement – from starting point, to fully forwards, then backwards (passing the starting point) fully backwards, and then back to the starting point, is called a cycle.
  - You might ask yourself why the wave moves forwards and backwards … remember “for every action there is an equal and opposite reaction.”
  - If there were a vacuum (no air), there would be no medium to “react” the disturbance, and hence no sound.
Cycles over Time

- As you can imagine, this movement is happening over TIME.
- Imagine this happening perhaps hundreds or thousands of times every second.
- A measurement of cycle(s) per second is also called the Hertz (or Hz).
- So for instance 2,000 Hz (or 2kHz) means the compression wave fully “cycles” 2,000 times every second.

Frequency

- As it turns out, the Hz (and kHz for thousands) measurement is used to measure Frequency.
- In this case the word means “how frequently does one complete cycle occur during a single second?”
- The frequency measurement is related to our concept of “pitch.”
- The higher (faster) the frequency, the higher the perceived pitch. The lower (slower) the frequency, the lower the perceived pitch.

Amplitude

- Another important measure of sound is the Amplitude.
- This measurement tells us how “intense” the disturbance in the medium is. We can also look at this in terms of the volume of air that is displaced in either direction.
- We perceive amplitude in terms of “loudness.” For human hearing, amplitude is measured in a scale called the decibel (or dB) scale. We will not get into the specifics of the dB scale but just keep in mind that it is not “linear” (in other words a 10dB difference does not sound “10 times as loud”).
Amplitude and hearing

- The concept of amplitude is not easy to conceptualize as a physical phenomena since we cannot “see” the compression wave.
- But, we can “feel” the wave. If you have ever been up very close to a very loud speaker for instance (which I am not recommending!), you can actually feel the movement.
- We can also detect when a sound is soft, loud, very loud, too loud, etc. Exposure to very loud sound can cause permanent hearing loss!!!

Amplitude over time

- Remember too that changes in amplitude are happening over time.
- In fact, since the amplitude is really telling us how “far” the disturbance in the air is “moving” in both directions, and this disturbance is happening at a certain frequency, there is also a time element involved with amplitude as well.
- Rarely does amplitude stay static – it usually changes as well (for instance if a sound gets louder, or softer, or “fades out”.

Visualizing sound over time

- Since sound is a physical phenomena that occurs over time, it is helpful if we are able to visualize what the phenomena “looks” like.
- To do this we can use a transducer (in this case, a microphone) to convert the physical energy into electrical energy. Then we can display it through an oscilloscope.
- We will look at some examples in the next few slides.
A sound “wave” graphed over time

- The X axis represents time.
- The Y axis represents levels of amplitude.
- The 0 point (no amplitude) is the “starting point.
- A movement from start to top, then back down to bottom, then back up to the 0 point again is a single Cycle.

Cycles

- In the diagram I added the small vertical lines to point out the individual cycles.
- The X axis would of course be measured in seconds. (But there are so few cycles in the diagram that this would be far less time than a single second if were audible!).

Amplitude

- I added the horizontal lines to point out the amplitudes over time.
- Remember we must swing back and forth for a complete cycle! The amplitude is measured relative from the zero point. In other words, -1 is just as “loud” as +1!
Notice that the “peaks” and “valleys” in the amplitude are “connected” over time by smooth, continuous changes in amplitude level. Not all sound waves “look” like this – some are more “jagged” – but just keep in mind that there is some sort of continuous change over time.

This diagram illustrates two waves. The second wave cycles more times over the same amount of time.

What does that tell you about the frequency/pitch?

How many cycles occur over a period of time tells us the frequency (pitch).

How intense the movement in either direction is tells us the amplitude (loudness).

The important thing to remember is that these changes are happening over TIME.

More specifically, we are really talking about changes in AMPLITUDE over time, and these changes are taking place a certain number of times per second.
Putting it all together

- So, we have to understand that “frequency” per se is not an explicit measurement on the graph.
- It is implicit because the graph will tell us how the amplitude “cycles” over time.
- If we see full cycles occur very “close” to each other on the graph, this means that there are more cycles in a given amount of time, which tells us that the frequency (pitch) is HIGHER.
- When the cycles are “further apart” this means there are less cycles in a given amount of time, which tells us the pitch is LOWER.

What we did NOT cover!

- The study of sound is far more complex than this lesson allows for.
- We did not cover common topics like harmonics, timbre, “noise”, phase, etc.
- This is because when we are talking about digitizing sound, we are really only interested in representing the changes in amplitude over time.
  - If you are interested in studying the topic of sound further, however, I encourage you to do so.
  - Remember: If we were designing a system to manipulate and process sound, we would need to know about these things, and many others too!

Sound – Review

- Know what a compression wave is.
- Understand how to measure amplitude and frequency.
- Know what a cycle is. Know what the Hz measurement means.
- Understand how graphing cycles of amplitude changes over time implicitly gives us information about the frequency/pitch of the sound.
- Know the range of frequencies that the human ear can detect as sound.
Part 2: PCM Audio

- The type of digital audio scheme we will be studying are called Pulse Code Modulation (PCM).
- We will not discuss the technical details of the various file formats that use this type of encoding (some examples include WAV, AIFF, etc.) but we will understand the BASICS of this method and do some work with theoretical example audio files.

Digitizing an audio file

- Imagine we are looking at a sound file on an oscilloscope.
- Since it is a continuous “line”, we know that there are an infinite number of points on that line.
- We also know (hopefully!) that the computer can NEVER handle an infinite number of values!
- So we instead have to take tiny “points” along the line and string them together.
- Remember that these “points” will represent some value of Amplitude.

Taking “points” of the Wave

- The dots represent possible points, or “samples”, that we may take. Remember that each sample will represent a value of Amplitude.
Sampling Rate

- Taking discrete “points” – or “samples” of the wave – is called “sampling.”
- How many samples we take over some set period of time on the line (one second for instance) is called the sampling rate.
- The higher the sampling rate, the more “points” we capture. The lower the sampling rate, the less “points” we capture.
- This corresponds conceptually to our study of “resolution: when talking about images. In fact, sometimes the sampling rate is called the “resolution” of the sound file.

Different Sampling Rates

- Notice that the second “half” of the wave is sampled at a higher rate – in other words, there are more samples taken over the same period of time.
  - This is for illustration only – we would normally never take samples at two different rates for the same sound!

How Many Samples Per Second?

- You should be wondering about now exactly how many samples do we need to take each second to be able to accurately represent the sound.
- There is a set rule for this (called the Nyquist rule). We need to take at least twice as many samples per second as the highest possible frequency of the sound.
  - I’ll leave it to you to figure out why this must be the case...
The sampling rate is measured in **Samples Per Second**.

But to make things confusing, the measurement is usually represented in Hz!

So, a sampling rate of 40kHz means that we are taking **40,000 samples per second**.

Just be aware of this. The Hz in this context does not mean Cycles Per Second, it means Samples Per Second!

---

**Examples – sampling rates**

- Given: a sound contains frequencies ranging from 400Hz through 2kHz.
  - What is the minimum required sampling rate?
  - Answer: $2 \times 2kHz = 4kHz$.
    - or 4000 samples per second

- Given: a sound contains frequencies ranging from 1kHz through 15kHz.
  - What is the minimum required sampling rate?
  - Answer: $2 \times 15kHz = 30kHz$
    - or 30,000 samples/sec

---

**Sampling Rate for Voice**

- The human voice can produce frequencies roughly between 200Hz and 4kHz.
- Since the highest frequency is 4kHz, this means our Sampling Rate must be at least 8kHz.
  - Yes I know using Hz in both ways is confusing 😞
- The 8kHz sampling rate for voice / spoken word is the industry standard.
The human ear can detect frequencies roughly between 20Hz and 20kHz.

Since the highest frequency is 20kHz, this means our Sampling Rate must be at least 40kHz.

Yes I know using Hz in both ways is confusing 😐

The industry standard rate for sampling music is in reality 44.1Khz.

Know first of all that we always have to account for the “worst case scenario” – in other words, what is the highest possible frequency in a particular sound?

If a particular type of sound is sampled at a rate LOWER than twice the highest possible frequency, what do you think will happen, and in what circumstances?

Consider a sound that only contains frequencies between 80Hz and 400Hz.

Our rule tells us that we need to sample it at 800Hz.

So what if we only sample it at say 300Hz?

This means that any frequency in the sound that is over 150Hz (half the sampling rate) will NOT be able to be represented properly.

This is why we always need to account for the highest possible frequency.
Review – Sampling Rate

- Know what samples are.
- Know what a sampling rate is.
- Know how it is measured.
- Know the Nyquist rule to determine how many samples per second we need to take given the highest possible frequency in a sound.

Sample Size

- Remember when we discussed graphics? We first discussed resolution, and then we discussed color depth (i.e. how many possible shades can a pixel represent?).
- In sound, we use a similar concept that tells us how many possible levels of amplitude can each individual sample represent.
- This is called the Sample Size, or Bit Depth.

Sample Size and Amplitude Levels

- Remember that each sample is really representing a certain level of amplitude captured at a given point in time.
- In real life, there are an infinite number of amplitude “levels” between any two points on the “line”.
  - There is of course a maximum amplitude (the overall “loudness” but what about all the gradients in between silence and the loudest part of the sound?)
- So our Sample Size tells us how many different possible amplitude “levels” we can represent for each sample.
Amplitude Levels

- Consider a representation of sound that only accounts for very few amplitude levels: (say maybe Very loud, soft, or silence).
- When the sound is converted to samples, the actual amplitude level that is captured must be made to “fit” into one of these levels.
- The sound would be extremely “choppy” and barely recognizable!

How Many Amplitude Levels?

- For general sound/music, a common sample size is 16 bits per sample (or 2 bytes per sample).
- This gives us $2^{16}$, or 65536, different possible amplitude levels.
- In other words each sample we capture can represent one of 65536 possible levels (in gradients from silence to very soft to very loud all the gradients in between) –
  - Remember that we must have “levels” for above and below the 0 on the graph (because the sound wave “moves” in both directions). So our levels must represent both positive and negative values.
- Lower bit depths (like 8 bits/sample) may also be used when we might need as many amplitude levels (like with voice recording for example).

Bit Depth – Review

- Know what the bit depth is, and what it represents.
- Know some standard bit depths.
- Know how a sound sampled at too low a bit depth would be affected.
To digitize audio, we need a piece of equipment called an Analog to Digital converter (ADC).
This takes in the analog audio signal, captures individual samples at a given rate, and of a given depth.
Then these individual samples can be assembled and stored as a digital audio file.

Of course to listen to the digital audio, it must be converted back to a form we can understand. We need a piece of equipment called an Digital to Analog Converter (DAC).
This reads the digital audio file and, knowing the sampling rate and sample size it was recorded with, converts the digital representation into its corresponding amplitude levels over time.
Then this resulting analog signal can be played back through a speaker, headphones, etc.

This diagram gives an overview of the equipment used to digitize sound and play it back. We did not discuss the “filters” but you should be able to identify the other equipment and processes.
Sample Problem 1

- We want to digitize a music recording. The song is 4 minutes long. We will use the standard sampling rate of 44.1Khz, and a standard bit depth of 16 Bits/sample. The recording is STEREO (meaning that there are two independent “tracks” – one left, the other right. What is the file size in Bytes?
- 4 min * 60 sec/min * 44100 samples/sec * 16 bits/sample * 2
- 338,688,000 bits = 42,336,000 Bytes = 42.336 MB

Sample Problem 2

- Given:
  - Audio specifications:
    - Length = 5:30 (5 minutes, 30 seconds)
    - Mono
  - Technical Specifications:
    - Sampling Rate: 44.1 kHz
    - Bit Depth: 16 bits/sample
- What will the resulting file size be, in MB?
  - Length = 330 seconds, Bit depth = 2 bytes/sample
  - 44100 samples/sec * 330 sec * 2 bytes/sample
  - 29106000 bytes = 29.106 MB.
- What if the file were Stereo? Then simply double the total number to 58.212 MB.

Sample Problem 3

- Given:
  - Audio specifications:
    - Length = 0:45 (45 seconds)
    - Mono
  - Technical Specifications:
    - Sampling Rate: 8 kHz
    - Bit Depth: 8 bits/sample
- What will the resulting file size be, in KB?
  - Length = 45 seconds, Bit depth = 1 bytes/sample
  - 8000 samples/sec * 45 sec * 1 bytes/sample
  - 360000 bytes = 360 KB.
**PCM Audio – Review**

- Know what sampling rate and sample size (bit depth) mean
- Understand the various typical sampling rates and sample sizes, and applications for them.
- Given a frequency range, be able to calculate the minimum required sampling rate.
- Be able to calculate the FILE size given the sampling rate and sample size, and information about the sound (length and mono/stereo).

**Part 3 – Compression**

- Let’s consider a 3-minute song, digitized in stereo, CD audio quality.
- Remember the sampling rate is 44.1kHz, and the sample size is 16 bits (2 bytes).
- $44100 \text{ samples/sec} \times 60 \text{ seconds/min} \times 3 \text{ min} \times 2 \text{ bytes/sample} = 15.872 \text{ MB}$; of course multiply by 2 for stereo = $31.754 \text{ MB}$
- That is a large file size and will take a very long time to upload/download if we want to use them on websites or e-mail them.
- What do we do?

**Audio Compression**

- As you already know, there are two types of compression: lossless and lossy.
- Unfortunately for us, there is no completely lossless way to compress audio files (for the same reason it is not possible to use lossless compression with high quality image files).
- So, we have to come up with a way to compress the audio file and still maintain a reasonable listening quality.
Psychoacoustic Compression

- This is a type of lossy compression that takes into account the way the human mind and sensory system processes and perceives sound.
- Complex calculations are done to remove certain information from the sound file that are deemed "not perceivable" to our hearing anyway.
- We will not study the details but it is a fascinating subject you may want to explore further on your own.

Psychoacoustic Compression

- Once certain information is discarded, the file can then be further compressed using various lossless methods.
- There are several types of compression schemes that rely on the psychoacoustic model to discard sound information from the file.

MPEG Compression

- The most common form of lossy audio compression is called MPEG. This stands for the Motion Picture Experts Group. This is a real organization made up of smart people who came up with some complicated mathematical formulas to compress digitized motion pictures, which consist of both moving video, and a soundtrack.
- The soundtrack compression scheme is called MPEG-1, Layer 3. Sound files compressed using this scheme are therefore called MP3 files.
- The degree of compression affects the sound quality.
- Generally speaking, a sound file can be reduced to 1/10 of its original size using MP3 compression. Some loss will be noticeable at this rate, but it is "acceptable" for most uses.
A final word to the wise...

- Hopefully you understand a little more about lossy audio compression than you did before.
- With that in mind... if you are working with editing / saving sound files in any way, I recommend that you:
  - First save the file in a "raw" uncompressed format (common format is WAV but there are others – research them for your particular computer system).
  - Do all your editing/changing/saving/etc. ONLY in uncompressed format, each and every time.
  - Only when you are all done editing, save it as a MP3 file and distribute as you see fit.
- Otherwise, you can lose more quality each time you save the file in a lossy compression format!

... and another thing...

- Scenario: a friend lends you an audio CD. You “rip” MP3s from it (for your MP3 player) and return it (this is a purely hypothetical scenario of course 😊).
- Now say you want to play this music on a normal audio CD player. So you “burn” them to an audio CD as “regular” decompressed audio files.
- But all you did was make “regular” audio files out of ones that already suffered a loss in quality! Remember that even after it is decompressed, the discarded information is gone forever!
- Then someone “borrows” your homemade CD, rips MP3s from it, the cycle continues ... quality degrades with each generation.
- The lesson: Keep everything in “raw” WAV format whenever possible.

Audio Compression: Review

- Understand the concepts of audio compression
- Understand its applications and limitations.