The Untrained Ear

By:

Navid Galt

March 12, 2017

Department of Software Engineering- INFS 515- Dr. He

George Mason University

Fairfax, VA 22030

The Untrained Ear

Navid Galt

Abstract

MP3 and compression algorithms have brought a new light to the technology world and for years have allowed for efficient memory storage and capability. In this paper, we will be researching different audio file formats including MP3, WAV, and FLAC. Conversion of live sound to a digitized rendition, explanation of different hardware's used to house audio files, and different de facto standard audio formats that are used worldwide today will be explored. A comparison of the three formats and their results of data degradation caused from compression algorithms will also be explained.

Introduction

Mozart, Beethoven, Chopin; all the composers of this time had one thing in common - they didn't have to worry about degradation of music by different compression algorithms and audio formats. However, in today's day and age, file format can mean the difference between a blurred hi hat to a sanded over drum beat. To the untrained ear the difference is almost indistinguishable. However, to those with a discerning ear and an appreciation for dynamic timbres, these different formats can mean the world of difference. In this paper, we will probe further into uncharted territory where science meets art as we seek to understand and describe the subtle differences between FLAC, WAV and MP3 formats. In doing so, we will graphically compare these formats measuring their frequencies and decibel ranges and compare how these subtle differences in frequencies can have a profound effect on the quality of music the listener hearkens to. In this technological revolution, music is undergoing a tribulation caused by societies obsession with sacrificing quality for quantity.

Before we jump into comparison of FLAC, WAV, and MP3 formats, let us first lay the foundation and establish the conversion of live sound to a digitized rendition. Live sound is typically collected from a microphone, amplifier, or directly from musical instruments connected to a computer by wires where a program can read in their sound waves. Original sound waves from live instruments begin as analog gradients. Analog basically refers to audio that is recorded using methods that replicate the original sound waves. The sound is recorded by electronically sampling the analog waveform at regular time intervals. At each sample the amplitude of the wave is measured by an electronic circuit, which then converts this analog measurement to its binary equivalent [3]. Binary code is a series of 0's and 1's used by the computer to transfer, import and export data. The circuit that performs this function is known as an A-to-D converter. The loudest possible sound creates a positive peak and is represented by the maximum binary number and the most negative peak is represented by the largest negative number. Binary number 0 falls in the middle. The sampling rate used by the converter is normally around 50 kilohertz which basically means that the program is sampling an average audio signal at about fifty thousand times a second! To aid in visualizing this conversion, below is an illustration that depicts this conversion.

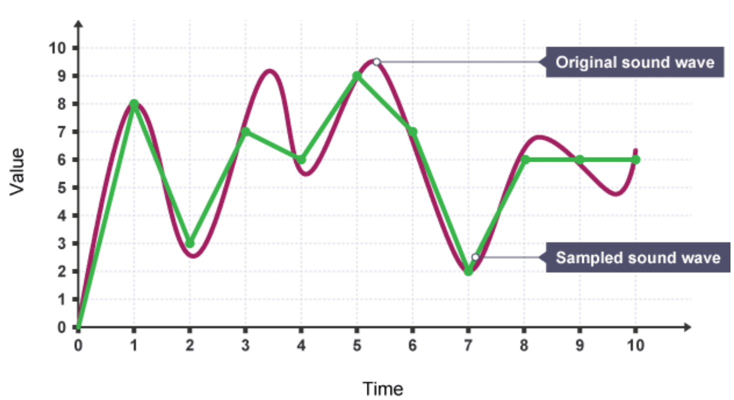


Figure 1 – Illustration of an original sound wave and a sampled sound wave via A-D converter.
(Source: http://www.bbc.co.uk/education/guides/zpfdwmn/revision/3)

You can see from figure 1 that the converter is discretizing the analog waveform into a coarser representation of the same wave, but using bins which are spaced 1 second apart. This process of 'fitting' the continuous signal with a discrete approximation of the waveform is the conversion process that creates the digitized audio. After these waves are converted to digitized data the receiving computer's program processes this code and reproduces these waveforms to recreate the sound waves initially converted. However, even in a lossless compression, where supposedly zero information is lost in conversion, the algorithm is still unable to exactly replicate an analog curve. In the next section, a brief review of hardware strategies used for audio storage is provided and different compression algorithms applied through software to audio are described.

Background – Storage Mediums

In general, new technology is often expected to offer an improved experience for consumers. This improvement could be categorized as a perceptual gain (quality) or a performance gain (quantity). Following Moore's Law, digital audio devices have improved over time by minimizing audio playback latency (fast retrieval) while maximizing the space available (denser transistors) for the user to fill. However, before the digital revolution in audio/visual technology, vinyl reigned supreme in terms of quality. This medium is still considered by many to yield superior audio quality, and offers listeners a larger control over the waveform. This is due in part to the fact that vinyl does not require a conversion from analog to digital – the needle which makes contact with the vinyl is directly reading the waveform of the audio sample. A vinyl record is a two-sided polyvinyl chloride disc which contains an inscribed groove that is physically etched onto the surface in a modulated spiral shape. This continuous groove initially has a uniform depth from the surface of the vinyl. If one attempted to play this uniform audio using a needle, it would have zero variation in frequency and pitch, and might possess a degree of white noise imposed by impurities in the record surface as well as on the needle itself. [15]

Audio is written onto this disc through a process which includes microphones, magnets, a motor, and a sharp needle. To briefly summarize this process, a needle is placed at the start of the outer spiral, and a motor is started which eventually brings the needle, bound by the grooves, towards the center of the disc. The analog audio waveform captured with microphones is amplified with vacuum tubes, and this amplified signal is used to drive an electromagnetic recording head. This allows for a wide frequency range to be recorded, without any constraints on playback volume. During playback, the record player's needle acts as one component in a transducer – a transducer is a device that converts one form of energy to another. In the case of a record player, this transducer is a cartridge which is composed of a stylus, cantilever, magnets, coils, and body. This cartridge is then able to convert the mechanical energy of the recorded vibrations back into sound waves, which are then amplified and broadcasted through speakers [16].

Vinyl's storage medium for audio files in this circular pattern of data compression is a common theme we encounter through multiple channels of data storage. A propagation of vinyl's technology can be seen in other data storage devices including the CD, which we will discuss next. Data storage throughout most mediums can be seen as a modulated spiral track due to this arrangement's ability to compress the most data in the most convenient and cost-efficient form.

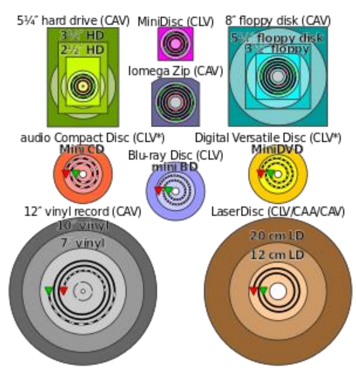


Figure 2- Comparison of several forms of disk storage showing tracks
(Source: https://en.wikipedia.org/wiki/Gramophone_record#LP_versus_CD)

The next big jump of hardware mediums used to house audio files are CDs. A CD-ROM is a simple piece of plastic, and its ability to hold music is still considered magic to some even in this day and age! CDs use microscopic laser etched crevices to register data and audio files in it's storage. A CD is etched with a single track of data, shaped in the form of a spiral, from the inside to out shown below in figure 3. After these bumps are etched into this clear piece of polycarbonate a reflective aluminum layer is positioned on top of the bumps. Once in the CD player, a laser beam is used on the reflective side of the CD.

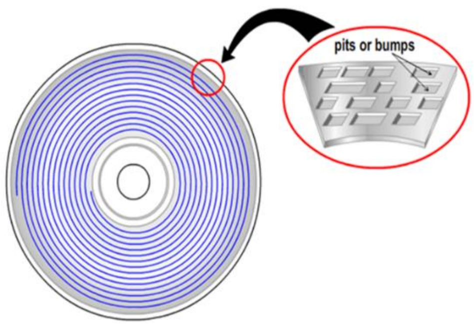


Figure 3 - CD record/playback surface with enlarged region showing pits/bumps within grooves.
(Source: http://www.futurekids.com.sg/doyouknow.html)

After this is done a light detector, otherwise known as an opto-electronic sensor, is used to determine whether it is reading a pit or pit-less mark along the spiral track. Pit's translate into 0's in binary and a flat area is translated into 1's. This set of binary code composed of 1's and 0's is used by the computer to understand the data from the CD as mentioned earlier. The computer is then able to take these numbers compare them to a table of characters and voilà! the computer is able to re-create the music which was burnt onto the CD.

Background – Audio Formats

As we dive further and further into a hardware we come across different software also known as formats in this instance. The first format to be discussed is WAV otherwise known as Waveform Audio File Format. WAV is an audio file format standard that was designed by Microsoft. WAV's equivalent on Macintosh computers is known as AIFF format. WAV format is raw and typically uncompressed audio, making it a lossless format, meaning no data is sacrificed in conversion from PC to PC. Because of the large file sizes it requires, WAV is less popular among audio files as it takes about 10MB per minute of music [9]. The linear pulse code modulation (LPCM) format is the technology that encodes and digitally represents these analog wave signals. There are two properties of a pulse-code modulation stream that determines the stream's fidelity to the original analog signal, the sampling rate and the bit depth. The sampling rate refers to the number of times per second that samples are taken. This rate is measured in hertz and represents the range of discrete frequency bands that compose the audio. The larger the sampling rate, the more samples that are taken, and the more superb quality of audio.

This ensures more accurate precision of the high and low notes throughout the track which translates to the listener being able to hear more dynamic nuances. The number of possible digital values that represent each sample being taken is referred to as the bit depth. The higher the bit rate (bit depth), the higher quality of audio. For instance an 8-bit audio will sound grainy and harsh, while a 16-bit audio sounds much better. Standard CD format has a 44.1k sampling rate combined with a 16-bit rate[2]. The sampling rate along with the bit depth together encompass a term that might be more familiar to most, kbps, which is the kilobytes per second of the song. This refers to the number of bytes that are sampled per second. So, kbps essentially refers to the distinct frequencies that are packed into a single second. Bit rates range from 96 to 320 kilobytes per second (kbps). The higher the bit rate the less the encoder will discard during the compression process equating to a higher quality of music. WAV files are typically set around 320 kbps, ensuring sound quality is not sacrificed. Even at this level of quality WAV formats, compared to vinyl records, still miss transitions that might not be picked up when transferring analog data to digitized data. Albeit this format being lossless, this digital recording is still not capturing the complete sound wave, it is quantizing it with a series of steps. Some sounds that have instantaneous transitions, such as a drum beat or a trumpet's tone, will be distorted because they change too quickly for the sample rate. [14].

MP3 stands for MPEG Audio Layer III and was developed by The German company Fraunhofer-Gesellshaft [1]. MP3s use lossy compression algorithms to decrease file size, lossy meaning data is lost in conversion. These lossy compression algorithms operate under the science of psychoacoustic theories. Psychoacoustics refers to the study of how people perceive sound. These theories state that the typical human ear can only perceive sounds with frequencies from 20 Hz to 20,000 Hz. However, under ideal laboratory conditions, humans can hear sound as low as 12 Hz and as high as 28 kHz [11]. They also operate under the assumption that the user can accept a certain amount of data degradation as a trade-off for the savings in a critical resource such as storage requirements or data transmission time [3]. However this degradation of data is leading to significant changes on the way we view music and is altering the music industry. “Some musicians and audio engineers say that the MP3 format is changing the way music studios mix recordings. They say that the MP3 format "flattens" out the dynamics -- the differences in pitch and volume -- in a song. As a result, much of the new music coming out of the industry has a similar sound, and there's not as much of a focus on creating a dynamic listening experience. Why work that hard on creating a complex sound if no one can detect it?” [12]. However detrimental this ubiquitous format might be to the uniqueness of music it has become the de facto standard (commonly practiced standard) due to its space complexity. “MP3s gained popularity because of the large number of encoded recordings that were posted on the Web in this format and also because of the availability of low-cost portable devices that can download, store, decode, and reproduce MP3 data” [3]. Like its counterpart the key component that directly affects the size and compression of a MP3 format is the bit rate also known as the number of bits per second encoded in the file. MP3’s are typically compressed into 128 kbps files which is directly related to the size of the file. Unlike it’s 320+ kbps counterpart MP3 might be saving space, but this, like all good things comes with a price. When an audio file is compressed certain frequencies that are deemed “undetectable” to the human ear are truncated.

Let’s also take a look at another less common, but far superior in quality audio format, FLAC. FLAC is an acronym used to describe Free Lossless Audio Codec. FLAC first emerged as an open-source alternative to other lossless formats emerging in 2001. These included Apple Lossless (ALAC), Microsoft's WAV (Waveform Audio Format), and WMA Lossless [6]. FLAC uses a lossless compression algorithm where quality is not sacrificed for space. “It is similar to how a zip works, except with FLAC you will get much better compression because it is designed specifically for audio [7]. However, FLAC has its drawbacks and limitations. It is not commonly supported by average PC and MAC software or portable player devices and doesn’t truly have a popular purchasing source (like Apple Store or Amazon). It is also very consuming space wise and even with multiple different compression algorithms to choose from and encode your audio file with, it still has a larger than normal space complexity if you were to compare it to a MP3; “about six times larger than an MP3” [6]. Experts such as Malcolm Hawksford, professor of psychoacoustics at Essex University, say that despite competition from other formats FLAC is still competitive. "FLAC has a place in the future for high-quality audio. It is good for transporting files on the Internet as it typically halves download time. It is unlikely that for lossless compression there will be significant improvements” [8]. So it seems that FLAC is here to stay and will continue to grow so long as memory becomes cheaper and the need for higher quality, originally quantized audio files arises.

Methods and Materials – Comparing Audio Formats

Now that we’ve introduced the three different formats, let’s take a look at a side by side comparison that depicts the differences of these three format’s frequencies and decibel ranges. Decibels are the units used to measure the intensity of sounds. To illustrate these differences, a free software, Audacity, was used to read in these three formats discussed above. Audacity is able to provide several measures of audio properties for users of different levels of signal analysis experience. These three samples of MP3 (orange), WAV (yellow), & FLAC (blue) are of different genres and therefore have different dynamics throughout the audio. Below, figure 4 shows three different formats of an R&B genre audio file, embedded into a plot of Hertz (Hz) vs. Decibels (db), where a range of frequencies (Hz) comprise the x-axis and a range of volumes (db) form the y-axis.

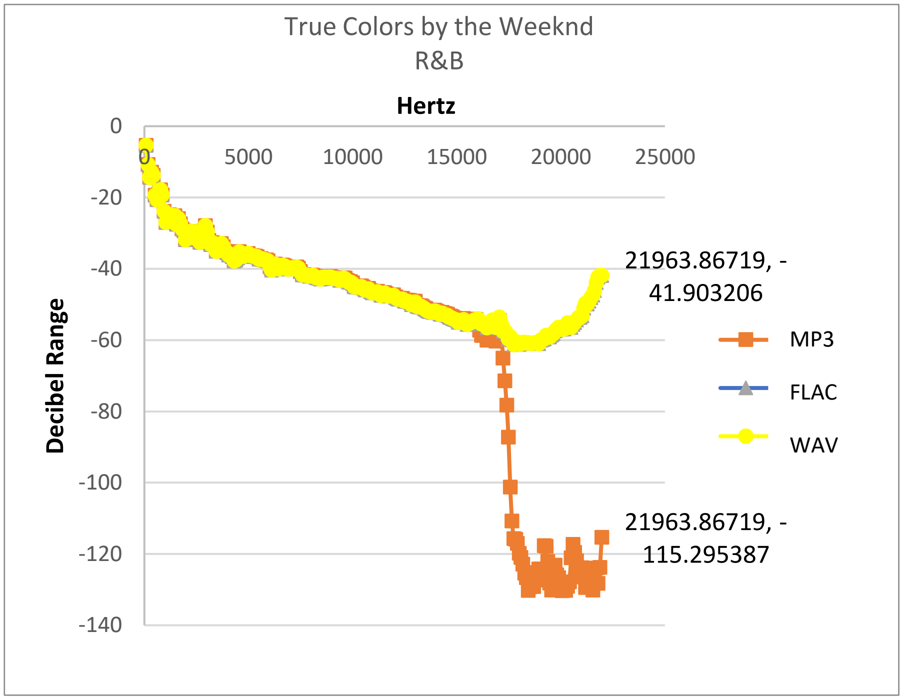
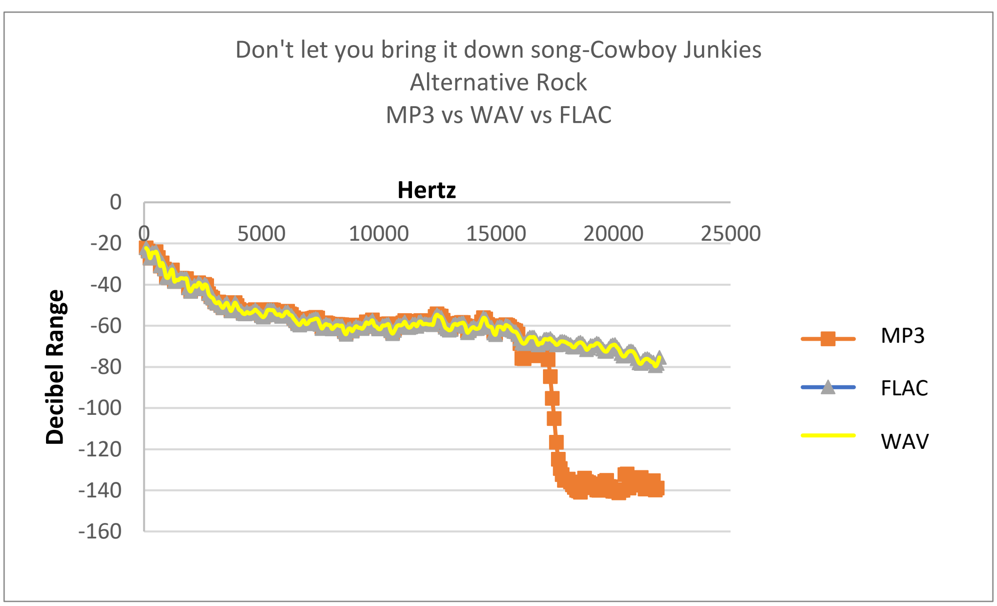


Figure 4 – Frequency Spectrum vs. Intensity plot of a song from the R&B genre with 3 audio compression formats.

You're able to distinguish the MP3 (orange) from the other two formats as the MP3 truncates hertz ranges slightly above 15,000, which is still technically well within human perception range according to Psychoacoustic experts. With a side by side comparison of decibel range in excel we can also see that a considerable number of decibels are truncated throughout the process of a MP3 compression with a total sum of -15024.32685 decibels for the MP3 format vs. a total sum of -11402.89416 for WAV format and -11402.94049 for FLAC format. With this truncation MP3's are able to considerably reduce file size and data allowing for larger storage capabilities. However, WAV and FLAC files are almost indistinguishable if not exactly simlar. So since FLAC files use a working loseless compression algorithim to store their music, are those files the way to go if searching for a high quality audio file? Below, figure 4 depicts a graphical comparison of frequency spectrum for different music genres.



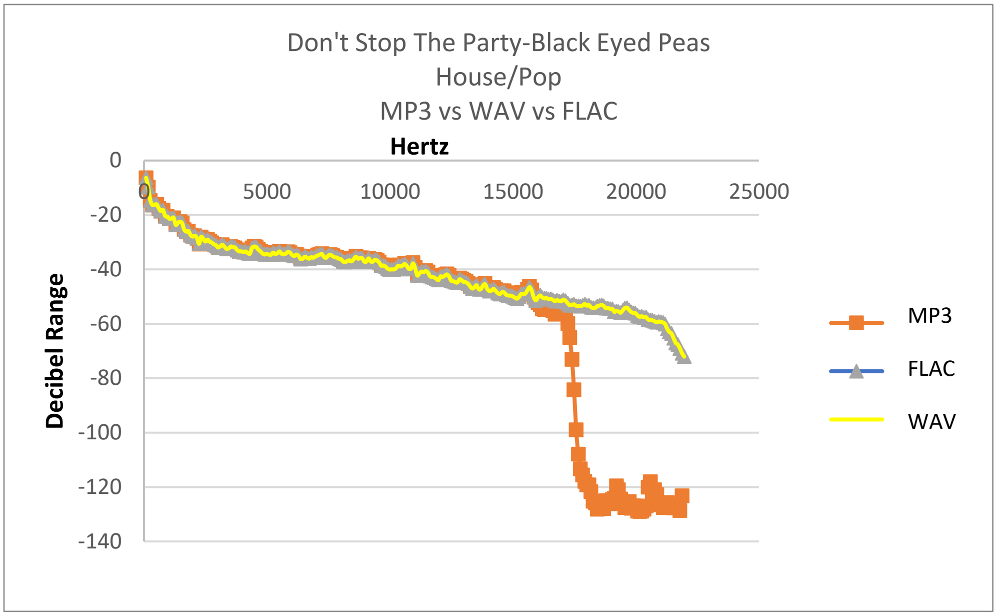
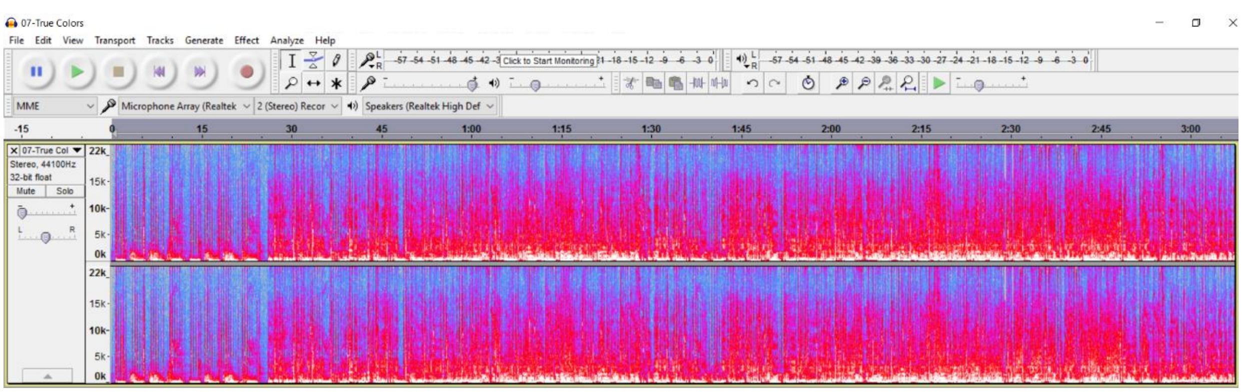


Figure 5 – (top) Frequency Spectrum vs. Intensity plot for Alternative/Rock music, (bottom) Frequency Spectrum vs. Intensity plot for House/Pop music show a similar spectral profile for mp3 which is consistent with Psychoacoustic theory.

Results and Discussion

You can see that as above this result is the case for several different genre types depicted. MP3's lossy algorithm seems to be outmatched by it's two higher quality cousins. Another question one might ask is are MP3 lossy compression algorithims to loose with what they truncate? These compression algorithms even though considered effective might not always be successful for certain types of genres. With different genres of music there is different timbres and dynamics that sets that specific genre of music apart from the rest. Should MP3 compression algorithims account for this? If so how will that be regulated/done? This careless blanket approach needs to be nixed, because as of now music quality is suffering and being negatively affected by these generalized compression algorithms. Since the high ranges are effected more by these MP3 algorithms maybe these algorithms can be customized for certain genres and account for these higher ranges to maintain a certain level of quality.

Let's dive deeper and take a look at two versions of the first song in a spectrogram view. A spectrogram is a visual representation of the spectrum of frequencies in a sound or in this case an audio file, as they vary with time [10]. Below, this is depicted in figure 5 with a comparison of the R&B audio file, over time.



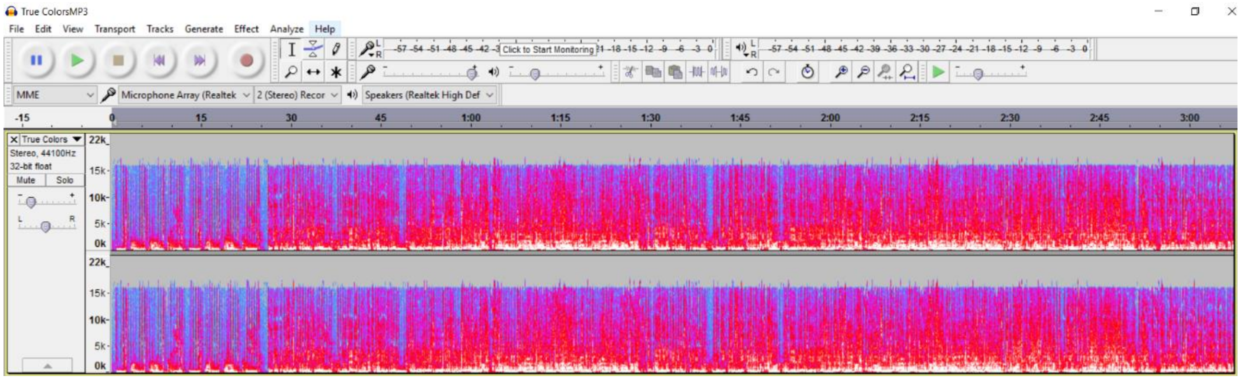


Figure 6 – Spectrogram view of R&B genre song – (top) uncompressed version, (bottom) compressed mp3 version shows clipped range of frequencies from 15kHz – 22kHz due to the algorithm.

The blue wave pattern represents low decibels in the negative range and the red depicts the higher decibels near the zero range. If you take a close look you can see that in the first image the MP3 compression algorithm has truncated these frequencies at just over 15,000 hertz which is significantly lower than percieved audible range, 20,000 hertz or for some 28,000 hertz. In the FLAC audio file you can see these decibal ranges well surpassing percieved audible frequencies. From these two pictures we can distinguish a major difference in the MP3 file where there is less variation of sound bits and a considerable cut off. Here, the quality/frequencies pitches have been sacrificed in order to truncate file size.

How does this data degradation effect sound quality and to what degree can humans perceive it? To evaluate these questions, we conducted a naïve ABX auditory test - a method of comparing two choices of sensory stimuli to identify detectable differences between them - on three different formats of the same song. Disclaimer: these tests are based on the sounds the author and other subjects perceived therefore are totally subjective. It is worth mentioning that the author is a practiced musician with a discerning ear, and this may assist with picking out subtle differences. Also, it should be noted that these sound tests were conducted on a recommended audio player that supports WAV and FLAC files. These audio files were also projected through a high quality home theater system with Klipsch RF-82 floor speakers so that minimalistic differences would be exacerbated. Results of the auditory test are shown in the table below.

Format	**Typical file sizes (5 minute stereo sample in all the formats)**	**Perceptual differences (quality)**
WAV	• 44.1 kHz @ 16 bits →50.4 MB • 44.1 kHz @ 24 bits → 75.7 MB	• The WAV format sounded more dynamic, hi hats would gradually strengthen and draw back. • All test subjects, including author, were able to perceive different gradients and a more robust sound of timbres. • The output was more crisp and clear with sharp high and low tones. • Each instrument had a unique crescendo.
FLAC	• 44.1 kHz @ 16 bits →6.69 MB • 44.1 kHz @ 24 bits → 27.9 MB • (note: Sizes of FLAC files can vary greatly depending on music dynamics, conversion tool used, what format the file was converted from and Level of compression used. The information here is general and currently being revised for more accuracy.)	• Almost identical to WAV, so close that all test subjects, including author, could not notice a difference.
MP3 (128 Kbps)	• 44.1 kHz @ 128 kbps →3.43 MB • 44.1 kHz @ 320 kbps → 11.4 MB	• Hi hats were blurred into one continuous monotonous sound • Instrumental sounds seemed processed and unnatural/artificial

		• No gradients noticeable • End tones were more subdued instead of discrete and sharp • Degradation of quality and harsh vocal tones ("s" started to sound like "sh") • MP3 version sounded noticeably gargled, yet tolerable compared to radio standards

Table 1 – Quantity vs. Quality: Tradeoffs over different compression rates and audio formats.
(Source: http://dsd-guide.com/size-comparison-chart-various-formats-dsd-wav-flac-mp3#.WMSWEYWcGHk)

Conclusions

In this paper we have taken a look deeper into the process of audio digitization. Unmasking the science behind storage mediums, audio conversions from analog to digital, and compression algorithms through comparison of three different audio formats. As seen graphically and through these implications, MP3's compression algorithm is shown to possibly truncate ranges that are perceivable by humans along with ranges that are unperceivable by humans. This truncation can lead to a significant and noticeable degradation in the audio file that is changing the music industry to conform to a monotonous standard. So, what can we do to shift from this disastrous path MP3's have set us on and save music? Neil Young a famous and well-renowned artist, producer and director calls for a necessary "end to end reboot of the consumer digital audio ecosystem, from file formats to playback devices" [13]. Who knows maybe eventually the incremental improvement in compression to be achieved will no longer justify the additional cost in processing or the degradation of the result [3]. It seems that with FLAC, and other high quality compression algorithms we're heading in the right direction. As technology becomes more advance and storage hardware becomes cheaper and smaller, FLAC and other high quality formats will become more common and compatible with portable devices.

Throughout this study, we have compared compression algorithms of audio formats specifically related to recreational music, however, this practice goes beyond music and can have some serious real world implications. For future investigation into audio compression and the effect on perception, there are different areas that can be researched. Music therapy would be one area of interest, such as analyzing certain frequencies and how these can be used to treat certain disorders and diseases. A study to be done in cases of war veterans and others suffering from Post-Traumatic Stress Disorder (PTSD) would be of great interest. For instance, designing a filter for personal music players that minimizes frequencies known to induce a relapse or mental breakdown, while maximizing frequencies associated with meditation and relaxation. Another field of interest would be to investigate best practices of compression algorithms assisting those using certain programs to help mask or filter out unwanted noise (i.e. Live concert recordings, Intelligence community using a tool to help distinguish the voice of a person of interest from among a crowd). A myriad of further research topics arises from these interesting finds and can be used to further advance how humanity studies and perceives sounds. For now, I'll leave you with the wise words of Ludwig Van Beethoven "Don't only practice your art, but get at the very heart of it; this it deserves, for only art and science raise men to the level of the gods".

References

[1] Bellis, Mary. "Meet the Men Who Created the MP3." *Thought Co*. N.p., 13 June 2015. Web. 08 Mar. 2017. <https://www.thoughtco.com/history-of-mp4-1992132>.

[2] Agarwal, Rajiv. "Demystifying Audio Formats: What Format Should You Record In?" *Audio Issues*. N.p., 20 Mar. 2013. Web. 09 Mar. 2017. <http://www.audio-issues.com/homerecording-studio/demystifying-audio-formats-what-format-should-you-record-in/>.

[3] Englander, Irv. The Architecture of Computer Hardware and Systems Software: An Information Technology Approach. Hoboken: Wiley, 2014. Print.

[4] "WAV." Wikipedia. Wikimedia Foundation, 06 Mar. 2017. Web. 09 Mar. 2017.

[5] "Pulse-code Modulation." Wikipedia. Wikimedia Foundation, 04 Mar. 2017. Web. 09 Mar. 2017.

[6] Pendlebury March 2, 2017 1:24 PM PST, Ty. "What Is FLAC? The High-def MP3 Explained." CNET. N.p., 02 Mar. 2017. Web. 09 Mar. 2017. <https://www.cnet.com/news/what-is-flac-the-high-def-mp3-explained/>.

[7] Coalson, Josh. "FLAC - Faq." FLAC - Faq. N.p., n.d. Web. 09 Mar. 2017. <https://xiph.org/flac/faq.html#general__what_is>.

[8] "The Definitive Guide to 24-bit FLAC." Bowers & Wilkins Blog. N.p., n.d. Web. 10 Mar. 2017. <http://blog.bowers-wilkins.com/speakers/definitive-guides/the-definitive-guide-to-24-bit-flac/>.

[9] Murphy, David. "Audio Format Glossary; AIFF (Audio Interchange File Format) - Apple's Answer to Microsoft's WAV." PC Magazine Jul 19 2005: 119. ProQuest. Web. 10 Mar. 2017.

[10] "Spectrogram." Wikipedia. Wikimedia Foundation, 17 Feb. 2017. Web. 10 Mar. 2017. <https://en.wikipedia.org/wiki/Spectrogram>.

[11] "Hearing Range." Wikipedia. Wikimedia Foundation, 10 Mar. 2017. Web. 10 Mar. 2017. <https://en.wikipedia.org/wiki/Hearing_range#cite_note-7>.

[12] Brain, Marshall. "How MP3 Files Work." HowStuffWorks. N.p., 01 Apr. 2000. Web. 10 Mar. 2017. <http://computer.howstuffworks.com/mp32.htm>.

[13] Calore, Michael. "Why Neil Young Hates MP3 - And What You Can Do About It." Wired. Conde Nast, 02 Feb. 2012. Web. 10 Mar. 2017. <https://www.wired.com/2012/02/why-neil-young-hates-mp3-and-what-you-can-do-about-it/>.

[14] "Is the Sound on Vinyl Records Better than on CDs or DVDs?" *HowStuffWorks.* N.p., 11 Oct 2000. Web 11 Mar. 2017. < http://electronics.howstuffworks.com/question487.htm>.

[15] "Gramophone Record." Wikipedia. Wikimedia Foundation, 10 Mar. 2017. Web. 11 Mar. 2017. <https://en.wikipedia.org/wiki/Gramophone_record>.

[16] Makers, Disc. "How Does Vinyl Work? Disc Makers." Disc Makers Blog. N.p., 21 May 2015. Web. 12 Mar. 2017. <http://blog.discmakers.com/2015/02/how-vinyl-records-work/>.