MP3 perception: Perceived quality of MP3 music

Rajeev Naidu

San Francisco State University

Abstract

MP3s prove to be one of the most ubiquitous and convenient formats of audio media. It has both positive and negative connotations depending on the consumer. For musicians and artists, the MP3 has a damaging effect to original music encoded into this format. Music encoded in MP3 may contain a decrease of important frequencies and loss of three-dimensional sonic space. In order to assess if these changes are perceptual, three songs will be composed in the genres of hip-hop, alternative rock, and film music. These three songs will also be crafted and mixed during the post-production process. This includes enhancing fundamental frequencies to the various instruments and creating the three-dimensional sonic space. After, each song will be played back in three common formats to discover perceived changes in quality. This includes the sample rate and bit depth of 48/24, 44.1/16 (compact disc quality), and MP3 at a bandwidth of 128 kilobits per second (kbps). The tracks will be played back to three academic audiences including an introductory audio class, an advanced audio class and a Sound Aesthetics class. This preliminary study will aim to discover auditory changes in format when music is encoded into the MP3 algorithm.

Quality of music: Formats of music used in distribution

This paper seeks to illustrate perceptual changes in encoded audio concerning high resolution, CD format, and MP3s. Three music pieces in different formats will be played for academic audiences to test for perceived changes in sonic degradation concerning MP3s. Today, the entertainment industry bombards listeners with MP3 encoded audio whether its film, television, or music. On a progressive note, music streaming companies and websites use this format to populate large databases with ease. Downloading and streaming MP3s are feasible because of its minimal file size; however, they lack depth and sound distorted compared to higher resolution formats (Shepard, 2009).

MP3 files are pervasive among different forms of media. Jonathan Sterne states MP3s are "ubiquitous and banal" and preliminary studies suggest listeners prefer the sound of MP3s (Sterne, 2012). Corporations distribute MP3s for speed and convenience. As a result, this type of commercial distribution has created listeners who willfully accept these reduced quality files (Shepard, 2009). These changes evidently have an auditory impact on our culture by altering how we listen over time. Piaget discusses the concept of schemas and the nature of human intelligence in his theory of cognitive development. His theory will be used as a guide to assist in understanding why today's listeners may prefer the sound of MP3 encoded audio. Furthermore, Levitin's discussion about schemas related to music will help create a deeper understanding of auditory preference. The shift from compact discs to MP3s raised various concerns about the future of listeners. The changes in listening behavior due to technological progress should be considered.

Literature Review

Music played on vinyl represents the truest form of the recorded signal. Vinyl is renowned for its indiscernible quality that listeners describe as warm and pure; however, we are limited to the number of songs we can store on a record. Listening to MP3s feels and sounds vastly different. This type of format allows users to digitally store a plethora of files on their hard drive. Unfortunately, this technological advancement created a culture habituated to sonically degraded music, usually undetectable by the listener. Sterne (2012) states, MP3's are created for the distracted listener. Fraunhofer, the creators of the algorithm, intentionally reduced the sonic quality to create a playback algorithm small in file size and for listeners who are not critically listening to audio. Their algorithm was based on psychoacoustic principles, defined as the study of perceived sound. Corbett (2012) explains the quality of music when one listens to MP3 players, radio networks, iTunes, YouTube and other music streaming websites, which undergoes a form data compression known as perceptual coding. Corbett (2012) uses the term lossy to describe a format that negatively affects audio through data compression. This includes, MP3s, AAC, Ogg Vorbis and WMA (Cobett 2012). Many music and visual streaming applications use these formats to store an abundant amount of audio on their databases. According to Corbett (2012), audio in lossy formats experience quality loss and the resulting audio is not the same as the file that initially went into the encoder due to data compression. The encoder undergoes a series of procedures. Masking is one process partially responsible for mitigating sound aesthetics.

Masking and Temporal Masking

Frequency manipulation allows instruments and vocals to sound pronounced, clear, and audible. Increasing certain frequencies allow bass instruments to sound deeper and create the

crisp intelligibly of vocals. The MP3 encoding process diminishes frequencies deemed inaudible according to psychoacoustic principles (Corbett, 2012). According to Corbett (2012), perceptual coding works by splitting the full-range of frequency into thirty-two different parts, or bands, and then into a variable number of detailed sub-bands. Then, the frequency range of the different bands is analyzed for undetectable frequencies unable to be heard by the listener (Corbett, 2012). The noticeable frequency content remains while the algorithm discards imperceptible frequencies.

Also, certain frequencies are eliminated due to a process known as masking. Sterne (2012) states, auditory masking is based on a principle that when two sounds have similar frequency content, and one is played louder than the other, the listeners will only hear the louder sound. For example, if a bass guitar and kick was playing in unison, the perceptual coding algorithm will attenuate frequencies on the quieter instrument. An instrument can lose clarity and definition during this process. A bass instrument loses its sub-harmonic power while vocals becomes less intelligible. Additionally, the encoding process encourages loss of three-dimensional sonic space. The process generates the flat sound created by MP3s by altering reverb frequencies and delays.

More specifically, delays are affected by another process called temporal masking. This process attenuates a quieter signal over a period of time. Sterne (2012) states when two sounds play together, up to five milliseconds apart, the MP3 algorithm encodes only the louder signal. Often times, post-production engineers generate three-dimensionality through the use of short delays. This technique can create a sense of depth and ambience, similar to reverb, and should be of concern with MP3 encoded music.

The MP3 Sound

As stated previously, the MP3 encoder eliminates reverb frequencies. Post-production engineers use reverb to craft acoustic space around instruments (Gibson, 1997). For example, one can digitally place a guitar signal in a gymnasium using a reverb effect from their computer or music equipment. Guitar players utilize floor pedals or guitar amplifiers to achieve this effect. The masking process removes specific frequencies related to reverb, mitigating the realistic impression of the effect. Thus, the encoded process perceptibly reduces the three-dimensional sonic space.

Corbett (2012) examined this phenomenon and compared two versions of a song produced by Now Now Sleepyheads called *Milestones*. One version had a sample rate and bit depth of 44.1/16 (compact disc quality) and the other played back the MP3 version at 128 kilobits per second. He (2012) noticed the tail of the vocal reverb lost detail and texture when listening to the MP3 file. Decay time, a reverb parameter, creates the tail of a reverb. Decay time is defined as the length of the reverb until it's inaudible (Gibson, 1997). Corbett (2012) describes the reverb decay as being transparent, grainier and less smooth. Corbett (2012) also notices how reverb is more noticeable in some parts of the song compared to others.

Shepard (2009) discusses similar issues with MP3s. He (2009) writes how they sizzle and produce an unnatural, metallic noise that unintentionally sits in the mix. Shepard (2009) metaphorically relates the sizzle to chime bars or a mosquito buzzing in ones ear. He (2009) states sizzle comes from the lack of headroom in MP3s. This proves MP3s cannot handle the dynamic range of CD's; therefore, these tracks will result in an added distortion and sound more two-dimensional. Reverb usually play back quietly and is considered ambience. Being able to

discern the reverb in a mix can be difficult due to the distortion and metallic noise enveloped by MP3s.

Shepard (2009) also states, "There is a loss of '3D' stereo imaging, a blurring and flattening of the encoded audio. Often, MP3s sound as if they have less reverb than the original. Complex sounds lose their interest, the audio overall is less rich and involving - the result is harsher and more crude." Similar to Corbett's study, Shepard (2009) states MP3s sound two-dimensional due to the reduction of sonic information. Shepard (2009) states that up to 90% of original file is extracted. The encoder eradicates the subtle and detailed touches of a song. This includes ambience, space, realism and other delicate sounds. As a result, the lush and rich three-dimensional sound diminishes to a flat, two-dimensional sounding track, and, it is evident that the young generation seems to have gotten accustomed to the MP3 sound, resulting in a shift in quality preference.

Experimentation in Audible Preference

Jonathan Berger describes the MP3 effect as a penchant towards the MP3 sizzle (Gopinath & Stanyek, 2014). The Stanford professor discovered a rise in preference for MP3s and realized students often favored the MP3 sizzle (Gopinath & Stanyek, 2014). Since the early millennium, Berger conducted informal experiments, over a period of eight years, by playing back different encodings of the same song (Gopinath & Stanyek, 2014). Some scholars denote the MP3 effect as a phenomenon prevalent towards Generation Y listeners. Berger agrees and concludes this holds true because of their familiarity with the sizzle sound (Shepard, 2009).

On the contrary, Sean Olive (2012) conducted a similar research among 18 high school students. He (2012) concluded that 70% of his participants appreciate compact disc quality over MP3. Olive's examinations were guided in a controlled atmosphere, which includes an acoustic

listening room and high quality, accurate speakers. Olive (2012) discusses the improvement of audio encoders since the time Berger carried out his tests. He (2012) explains how listeners have grown accustomed to better encoders over time attributing to a change in listening behavior. Most participants declared a change in reverb traits. They claimed the CD quality reverb sounded raw, natural, brighter, and stronger (Olive, 2012). In addition, they stated the MP3 quality files were slightly distorted with less audible high frequencies (Olive, 2012). These participants describe the sound features of the encoder's algorithm discussed earlier. Besides the controlled settings, it should be noted as to what caused this change in audible opinion.

Gopitnath and Stanyek (2014) discuss how audible preference partially stems from listening habituation. They state that listeners from the 1960s gravitate towards the hissy noise from vinyl (Gopinath & Stanyek, 2014). In comparison, Generation Y listeners are routinely exposed to the crunchiness notable in MP3 audio, which explains their predilection towards this format. The authors also discuss a contextual preference towards audio algorithms (Gopinath & Stanyek, 2014). They write how some listeners favor specific playback qualities depending on their environment. Vinyl lovers prize the crackle of a record and desire that sound in a suitable acoustic room (Gopinath & Stanyek, 2014). On the contrary, it would be appropriate to listen to a plethora of MP3s for the distracted listener, or those who are not attentive to the music. Furthermore, Gopinath and Stanyek (2014) discuss the influence on music due to playback technology. They (2014) discuss listening to bass culture through treble culture. More specifically, mix engineers and musicians would often alter bass lines to make the instrument audible in broadcast technology that cannot represent low frequencies, such as in-ear headphones, laptops, and smartphone speakers (Gopinath and Staynek, 2014). Technology

affects both consumers and engineers in how they listen and create. One should investigate how preference originates and how it changes as technology progresses.

Theory

Piaget (1963) discusses the nature and development of human intelligence in his theory of cognitive development. He (1963) was interested in how an organism adapts to an environment, which he defines as intelligence. One of the three components in Piaget's theory of cognitive development is the concept of schemas. Piaget (1952) defined schemas as, "a cohesive, repeatable action sequence possessing component actions that are tightly interconnected and governed by a core meaning." Schemas are the basic building block of intelligent behavior and the way we organize knowledge. Piaget states that schemas can be thought of as a unit of knowledge and are used to understand, interpret and respond to situations (Cherry 2013).

Additionally, Daniel Levitin (2007) discusses the Piagetian notion of schema, but relates it to music. He (2007) theorizes that our appreciation for certain music develops at a young age and we develop schemas during these mentally formative years. Levitin believes that schemas function as musical filters. When we reach adulthood, we begin to filter out all the music that does not fit into our schema that we developed when we were younger. Musical tastes and expectations derive from schemas. Our proclivity and fondness of specific chord progressions and melody are developed through schemas. In addition, schemas explain why certain generations of listeners prefer specific playback qualities.

In terms of the pervasive and ubiquitous MP3s, younger listeners may have developed schemas that gravitate them towards the sizzly and distorted quality. Generation Y listeners have placed the MP3 as their primary form of listening because of quick download speeds, ease of transfer, cheapness and availability. These listeners may have miniscule listening experiences

with other formats such as vinyl and compact discs; thus creating a shift in listening behavior and preference. Piaget (1952) states, schemas contribute to stereotypes and make it difficult to retain new information that does not agree with our established ideas of the world. Therefore, listeners will have difficulty obtaining a proclivity towards vinyl and compact discs if they grew accustomed to MP3s. According to Levitin's idea of schemas, listeners develop their tastes in music during their teen years (Levitin, 2007). Consumers expect the hissy noise of vinyl if they were raised during that era. Comparatively, Generation Y listeners expect the crunchy sound of MP3s. Various researches have been done both concluding and falsifying this claim. Different methodology and practices should be examined to contribute to this study.

Method

Recall the purpose of this project was to investigate for perceived audible differences in particular formats. More specifically, this project aims to compare auditory dimensions of high-resolution audio with MP3 files concerning genres, frequency, and three-dimensional sonic space. The three audio formats include: high-resolution format with a sample rate of 48,000 and bit depth of 24, compact disc resolution with a sample rate of 44,100 and a bit depth of 16, and encoded MP3 format at 128 kilobits per second. The types of genres composed are rap/hip-hop, alternative rock and film score music, accompanied by the actual film.

The Three Dimensional Mix

The goal of these mixes was to produce three-dimensional sonic space. Additionally, the project aims to examine if listeners can perceive changes in the MP3 files. Reverb and delays were utilized to create three-dimensional space. These effects varied on the type of music. Reverb and delays created the sense of depth and pushed instruments further back in the mix. The amount of effects was based on intricacy and rhythm. The rap/hip-hop song contained less

reverb due to its fast percussion. The alternative rock and film score encompassed a greater amount of reverb due to greater musical space.

The use of delays created depth and width. Utilizing short delays, between five and thirty-five milliseconds, gave instruments a sense of depth. Also, applying delays allowed instruments to be heard from the sides, rather than the middle of the stereo image. After, specific frequencies were manipulated to improve the instruments intelligibility. For example, I increased the bass frequency of kicks to pronounce the low end of the sound. On the other hand, I raised the high frequencies of vocals and guitars to produce clarity. Each instrument, in each song, was treated to produce the best three-dimensional space in terms of width, depth and symmetry.

San Francisco State University provided audio equipment relating to the post-production process. The primary DAW's that was used to create the mixes and record the instruments were the programs Logic Pro 9, Pro Tools 10 and Nuendo 6. I used outboard reverbs including TC Electronics M One XL and M 5000. The plug-ins that was used to equalize, compress and create delays was provided by the stock plug-ins of Pro Tools 10 and Nuendo 6.

Demonstration

A supplementary goal of this study was to experiment how to approach this investigation to obtain accurate results. All things considered, the demonstration altered between each of the three academic audiences. The total sample size was approximately 72. Each audience was exposed to a PowerPoint explaining details of the test and why it was carried out. This presentation included differences in each format and their uses. Also, a brief overview of theory was covered. After the presentation, a short duration, between thirty seconds and one minute, was played back to the audiences. The order of format differed depending on the course and student knowledge. Each sample of the song was played three times, in the three formats. The

formats included a sample rate and bit depth of 48/24 (high resolution), 44.1/16 (compact disc format), and MP3 at 128 kbps. Students then took a brief survey after the each song played three times. The survey included questions regarding which format they heard, which they prefer, the differences in genre and quality, and if they're concerned with modern distribution of digital music.

Each class had distinctive playback systems. The beginning audio class utilized two low budget Tannoy speakers without any subwoofers. The second, Sound Aesthetics, class used two Tannoy 501 Active Speakers with subwoofers. Lastly, the Advanced Audio class used two expensive and accurate Tannoy Gold monitors, used in high-end recording studios. This room also had Tannoy subwoofers.

Analysis

My research demonstrates dissimilar results based on how the three formats were presented. I revealed the format before playing each song to the introductory audio class and asked if they can distinguish sonic differences. Most claimed they were able to differentiate between 48/24 and MP3 format. Furthermore, they admitted to challenges in separating 48/24 and the 44.1/16 compact disc version. Generally, genre did not play a role in separating the qualities. In other words, these students could not discern format due to song type. Also, the introductory audio class preferred the highest quality format. Lastly, most students took issue with the dominant MP3 distribution.

I challenged the Advanced Audio Class and chose not to tell them the format when presenting the three songs. Students' chose which algorithm played. Interestingly, they had problems distinguishing the three qualities; however, technical difficulties persisted when playing the first song. The playback device was improperly altered causing artifacts in the music.

Similar to the introductory audio class, most students could not decipher between compact disc and high-resolution formats. However, some distinguished which songs were encoded in MP3. Intriguingly, genre seemed to make a difference in discerning quality. 5 of the 7 students were able to tell which track was MP3 playing the film score. 4 of the 7 students were able to distinguish the MP3 version of the alternative rock song. And, 2 of the 7 students separated the MP3 version of the hip-hop/rap song. Finally, most students care about the digital age and pervasive MP3 dispersal.

I received similar results in the Sound Aesthetics class. I asked the classroom to guess which format of each song. The majority of the students guessed incorrectly; however, I received results from a miniscule sample that accurately distinguished the MP3 files. Also, genre did not make a drastic impact on differentiating MP3's from other qualities. 8 of the 33 students were able to discern the film score MP3. 11 of the 33 students separated the alternative rock MP3. And, 10 of the 33 students correctly guessed the rap/hip-hop MP3. Common amongst all audiences, most felt that quality of music is an issue and people should be able to retrieve high quality music as feasibly as MP3 files.

Limitations

The findings of this study are limited to a number of factors. First and foremost, mixing for music proves to be a difficult and lengthy practice. The necessary equipment and acoustic environment is needed during the post-production process to create proper three-dimensional sonic spaces. The final mix depends on the engineers' skill, experience, time spent, and equipment. The results of this study are also restricted to original music. The use of commercial songs may have altered the results and responses.

This study was also limited to the playback system the audience used to hear the mixes. Some speakers have the ability to expose music with precise representation. Accurate speakers detail the frequencies and three-dimensional sonic space. Audience responses differ depending on the stereo image, and quality the speakers depict. Listeners cannot hear the complete illustration of a song from low quality speakers; however, not all MP3 consumers use high-end playback systems.

The methodology of this study did not consider room acoustics. Recording studios are designed and acoustically treated to receive the most accurate playback from speakers. This includes studios containing sound panels, bass traps, sound diffusers, nonparallel walls, and other acoustic factors. This assists a listener to hear a defined representation of a sound, musical instrument or recording. One can also distinguish between two-dimensional and three-dimensional mixes in a proper acoustically treated room. Distinguishing MP3s in this environment would be simpler. Most importantly, separating MP3s from other formats is restricted to listening skill. Everyone has a specific ear for music; but not all consumers listen to the same aspect of a mix. Trained musicians and engineers will have selective hearing compared to untrained listeners. Some listeners distinguish MP3s better than others and naturally posses sensitive ears.

Discussion

Apparently, the characteristics of formats appear to concern most participants. Most did not have the ability to discern between the formats; however, results changed depending on how the tests were demonstrated. Perhaps, listening behavior varies on knowledge and pre-existing frameworks. It's also important to consider age. Age factors into listening skills and the attributes of hearing. It's possible that people of distinctive age groups listen to certain musical

features. Furthermore, certain generations of listeners prefer particular qualities of music. Generation X, and other previous generations, may prefer classic rock music. This type of music may me more susceptible to the negative traits of MP3 encoding. This explains why more participants were able to differentiate the alternative rock MP3. Generation Y listeners have a large rap/hip-hop following. Overall, a smaller number of participants were able to discern the rap/hip-hop MP3. Perhaps hip-hop encoded MP3s are acceptable.

The convenience and feasibility of MP3s allowed us to explore an abundant amount of music. We now have the ability to skip songs seamlessly. This efficiency of listening changes how we listen. Generation X listeners are accustomed to vinyl, which can only store a limited amount of music. They were prone to listening to each track without skipping over; thus, their listening attitude and behavior differs from our current state. Tastes in quality of music depend on time. Aristotle stated, “Time crumbles things; everything grows old under the power of Time and is forgotten through the lapse of Time.” Style, tastes, fads, quality, and preferences change over time. Also, our listening behavior evolves over time due to technology. Even if MP3’s are the dominant format today, in a few years it may not be. Hopefully, we will find a way to disperse music in high-resolution formats just as easily. Only time will tell.

References

Cherry, K. (2013). *What is schema in psychology?* Retreived from: http://psychology.about.com/od/sindex/g/def_schema.htm

Corbett, I. (2012). *What data compression does to your music.* Retrieved from: http://www.soundonsound.com/sos/apr12/articles/lost-in-translation.htm

Gibson, D. (1997). *The art of mixing*. Mixbooks: Auburn Hills, MI.

Gopinath, S., & Stanyek, J. (2014). *The oxford handbook of mobile music studies.* New York, NY: Oxford University Press.

Levitin, D. (2007). *This is your brain on music*. New York, NY: Penguin Group.

McLeod, S. (2009). *Jean Piaget*. Retrieved from: http://www.simplypsychology.org/piaget.html

Olive, S. (2012). Some new evidence that teenagers and college students may prefer accurate sound reproduction. *132 AES Convention Paper, 8683.*

Piaget, J. (1963). *The psychology of intelligence.* New York, NY: Routledge.

Piaget, J., & Cook, M. T. (1952). *The origins of intelligence in children*. New York, NY: International University Press

Shepard, I. (2009). *Do the kids prefer "MP3 sizzle"? Bullshizzle!* Retrieved from: http://mastering-media.blogspot.com/2009/03/do-kids-prefer-MP3-sizzle-bullshizzle.html

Sterne, J. (2012). *MP3 the meaning of format.* Durham and London: Duke University Press