Salisbury

Honor College at Salisbury University

Honors Thesis

An Honors Thesis Titled

Audio Compression : Psychoacoustics of MP3

Submitted in partial fulfillment of the requirements for the Honors Designation to the

Honors College

of

Salisbury University

in the Major Department of

Math and Computer Science

by

Morgan Iachini

Date and Place of Oral Presentation: SUSRC Spring 2019

Signatures of Honors Thesis Committee

	Signature	Print
Mentor:	Lori Carmack	Lori Carmack
Reader 1:	Barbara Wainwright	Barbara Wainwright
Reader 2:	Colleen Clark	Colleen Clark
Dean:		

Audio Compression: Psychoacoustics of MP3

Honors Thesis Project

Morgan Iachini

Mentor: Dr. Lori Carmack, Mathematics Department

Reader 1: Dr. Barbara Wainwright, Mathematics Department

Reader 2: Professor Colleen Clark, Music Department

Salisbury University

Abstract

Digital audio compression is used to minimize the size (number of bits) of an audio file in order to transmit the file effectively and clearly. Only necessary data points are encoded which will continue to represent the audio signal well. Examples of audio compression can be found in music and cell phone transmission. The main compression standard studied is MP3 digital audio compression, which is an effective tool for reducing the size of audio files. Reducing file size, however, may put sound quality in jeopardy. I created an audio test and questionnaire to see if humans can distinguish a difference in quality between similar compressed MP3 files. MP3 audio encoding implements a Psychoacoustic Model, which recognizes that most humans cannot perceive frequencies outside of the range 20Hz to 20kHz, and these frequencies are therefore not encoded. The audio test analyzes this technique used in MP3 encoding by compressing portions of a song at various sampling rates (samples per second) while keeping bitrate (number of bits read per second, corresponding to file size) constant at 96kbps (kilobits per second). Sampling rate is the number of samples per second that is read from a continuous signal and converted to a discrete signal, hence the more samples, the more data, and the higher the quality. The greater the bitrate and sampling rate, the higher the audio quality, but correspondingly the greater the storage necessary to save. Subjects are asked to listen and compare two different sound files and then identify which sound they deem has a higher quality. The goal of this project is to attempt to identify an optimal sampling rate for compression that reduces the file size while also maintaining high quality. MP3 encoding removes information from the digital audio file, and this can result in some humans detecting a reduction in the quality of the audio.

Audio Compression

Introduction to Audio Compression

Audio compression is used to reduce the size of digital audio files, in order to transmit audio data quickly and efficiently. Examples of audio compression can be found in radio and cell phone transmissions, along with picture signal transmissions to your TV. Digital audio encoding is most often studied in the digitalizing of music to create small downloadable sizes.

There are two different types of audio compression: lossy and lossless. Lossy compression results in the permanent loss of audio data, which could decrease signal quality. Lossless formats, on the other hand, allow for an audio file to be completely reconstructed and quality is not diminished. Within this project, lossy formats are studied, specifically the MP3 codec (program that encodes and decodes data into a lossy form).

Introduction to MP3

It takes a large amount of data to store high quality audio, which is not efficient for transmission. MP3 digital audio encoding has been an effective solution by limiting the amount of data in audio signals by disregarding "unnecessary" pieces of information within a signal. This method of compression will be discussed later in this paper.

MP3 stands for MPEG-1 Layer 3. Moving Pictures Expert Group (MPEG) developed various techniques to transmit not only audio signals, but video broadcasts as well. DVD's are examples of MPEG-2 standard. MPEG-1, however, is the standard used for compression of audio signals. MPEG-1 Layers 1 & 2 didn't divide the signal enough and failed to capture CD-quality. MP3, however, partitions the audible frequency bands and refines the partitions to ensure high quality, using the CD sampling rate of 44, 100 Hz (Wilburn, 2007, p.1-3).

The unique component of MP3 is that it implements a Psychoacoustic Model of compression to determine which signals are the "unnecessary" sounds. Psychoacoustics is the study of human perception from a psychological and physiological standpoint. Why do humans hear what they hear? After sound reaches the eardrum, it is passed to the cochlea. Within the cochlea are tiny hairs that convert the vibrations caused by the sound into electrical signals for the brain to interpret. But some humans can hear frequencies that others cannot.

Most humans cannot perceive frequencies outside of the range 20-20,000 Hz. MP3 compression takes this idea into account. Frequencies in an audio signal that fall outside of this range are deemed "unnecessary" and are not encoded into the compressed version. As stated by Rassol Raissi in his article *The Theory Behind MP3*, "while playing a CD it is impossible to percept all data reaching your ears, so there is no point in storing the part of the music that will be inaudible" (Raissi, 2002, p.4). The encoder embedded within MP3, is designed around the previous statement, and deletes data that is not crucial to a human's perception of a song.

Definitions

Before describing the internal algorithms used within MP3, a few definitions are required for the reader's understanding:

- A *sample* is a value of an audio function at a given point in time.
- The *sample frequency* (or *sampling rate*) is the amount of times the amplitude of a signal is stored per second. Sample frequency is measured

in Hertz (Hz). The greater the sample frequency, the higher the quality of audio. CD-quality has a sample frequency of 44,100 Hz.

- The *sample depth* (or *bit depth*) is defined as the number of bits used to store each sample. The sample depth essentially describes the y-axis of an audio signal. With bit depth of n, the y-axis is divided into 2^n partitions. For example, CD-quality has a bit depth of 16, and therefore there are $2^{16} = 65{,}536$ possible values along the y-axis. The greater the sample depth, the higher the quality of audio.

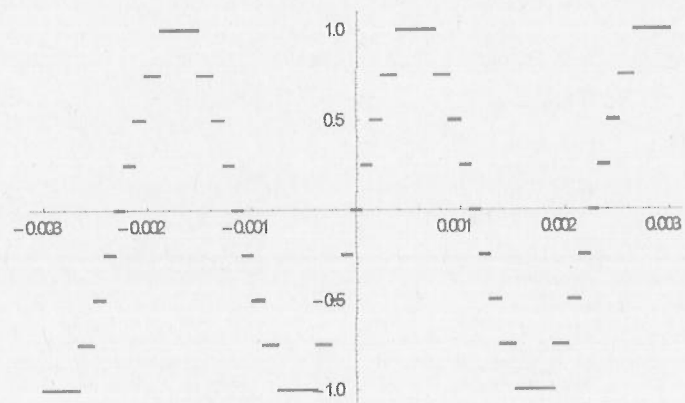

Figure 1: Rounding the tone, $Middle\ A = \sin(2\pi * 440t)$ to the nearest ¼. Hence bit depth 4.

- The *Bitrate* is the product of the sampling rate, sample depth, and number of channels (Raissi, 2002, p.8). For example, the bitrate of CD-quality audio is represented as $Bitrate = 44{,}100\ (sample\ rate) * 16\ (bit\ depth) * 1\ (mono - channel) = 705{,}600\ \frac{bits}{sec}$. Because the bitrate is the product of the sample rate and sample depth, the greater the bitrate, the higher the quality of audio.

MP3 Encoder

Prior to answering the question of whether there exists an optimal sampling rate and sample depth, the following is an outlined synopsis of the algorithms used within the MP3 encoder. The following steps and mathematical methods address the Psychoacoustic Model used in MP3 file compression.

Step 1: Polyphase Filterbank

A Pulse-Code Modulation (PCM) signal is used to digitally represent an original analog sound. Within the MP3 encoder, the PCM signal is divided into 32 frequency subbands of equal length (Raissi, 2002, p.5). The sampling frequency of a PCM determines the width of each subband, but what is a suitable sampling frequency? According to the Nyquist Frequency Theorem, "the sampling frequency should be at least twice the highest frequency contained in the signal" (Olshausen, 2000, p.1). For example, the tone Middle A is mathematically described as $f(t) = \sin(2\pi * 440t)$. The highest frequency of Middle A is 440 Hz, therefore, according to the Nyquist Frequency Theorem, a suitable sampling frequency for Middle A is 880 Hz.

Sampling at greater than the Nyquist frequency can result in oversampling. Sampling at less than the Nyquist frequency is under sampling and consequently loses the higher frequencies. This creates pulsing in the signal which is an unpleasant sound to the listener. Once a suitable sampling frequency is computed, the length of each subband is calculated to be $\frac{Nyquist\ Freq}{32}$ Hz. This algorithm reduces the amount of data in the signal since the entire frequency is not represented in a sample of each subband (Olshausen, 2000).

PulsingAbove.wav

Middle A sampled at 1,000 Hz, just above the Nyquist Theorem. (Double click the icon above to hear tone)

Pulsing.wav

Under Sampling Middle A at 879 Hz, just below the Nyquist Theorem. (Double click the icon above to hear tone)

Step 2: Fast Fourier Transform

Occurring simultaneously to the Polyphase Filterbank in the MP3 compression algorithm is the Fast Fourier Transform (FFT). The FFT is a complex algorithm that estimates a masking threshold for a given signal and will "give higher frequency resolution and information on the spectral changes over time" (Raissi, 2002, p.21). Temporal masking is when a loud sound is very close to a soft sound time-wise in a signal; the louder sound will "mask" the weaker one. Since these "masked" sounds are not typically perceived by humans, they may be removed from audio files with minimal effect on overall audio quality. The masking threshold is part of the Psychoacoustic Model, specific to MP3 compression.

Step 3: Psychoacoustic Model

As mentioned previously, most humans cannot perceive frequencies outside the range of 20-20,000 Hz. This range references the Absolute Threshold of Hearing of humans. The figure below is a good representation of the relationship between frequency and human perception.

Audio Compression

From **Figure 2** below, as frequencies approach 10^3 Hz, the amount of sound pressure (decibels, dB) needed to perceive the frequencies decreases. The dip in sound pressure, in the graph, from approximately 1,100 Hz to 1,400 Hz, demonstrates how easily humans can perceive frequencies in that range, fewer decibels are needed to perceive the frequency. However, as frequencies reach about 10^4 Hz and greater, humans need much more sound pressure to hear those sounds. From **Figure 2**, we see how humans' hearing discriminates against very low (bass) and very high frequencies. The MP3 encoder acknowledges these limitations in human hearing and creates a compression standard around this concept.

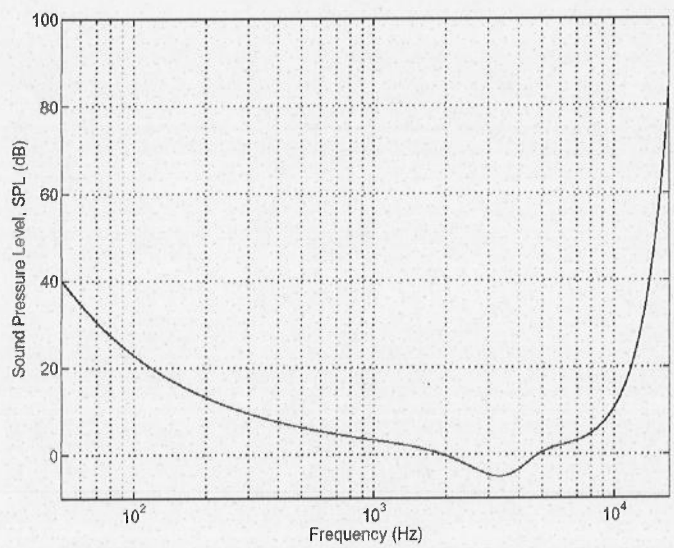

Figure 2: Representation of human's absolute threshold of hearing. (Raissi, 2002, p.4)

Step 4: Modified Discrete Cosine Transform (MDCT) Brief Overview

From the Polyphase Filterbank, the 32 subbands of equal length are partitioned even further by the Modified Discrete Cosine Transform (MDCT).

The MDCT takes those 32 subbands and refines them into 18 finer bands, creating a total of 576 total smaller bands. Therefore, "each band contains $1/576^{th}$ of the frequency range from the original signal" (Wilburn, 2007, p.3). The MDCT then windows each subband.

Windowing divides the frequencies in each subband with either a long or short window. The size of the window is determined from the Psychoacoustic Model. If the Psychoacoustic Model shows little change from the previous frame to the current frame, a long window can be used. On the other hand, if there is considerable change between frames, three short widows are used with 50% overlap between them. The overlap helps to blend frequencies together and avoid distortion. The overlap reduces artefacts caused by the edges of the windows.

Step 5: Quantization

The output of the MDCT is then quantized, which creates artefacts (static, muffled sounds) in the signal. Because a continuous set of data cannot be encoded, the signal must be quantized or discretized. In other words, the signal will be represented as a countable number of points. Rounding continuous data to discrete values will result in discontinuities within the signal, creating what is known as quantization error. There are various techniques that can minimize these errors (Burg, 2014).

One solution is called the Distortion Control Loop. The Distortion Control Loop mitigates the sound of artefacts by making the volume of these artefacts less audible to the human ear.

Another solution is related to the Dynamic Range of a signal. The Dynamic Range is defined as the difference between the loudest volume of a signal and the quietest. From a psychoacoustic standpoint, a human cannot perceive sounds at both ends of the range in a close period of time. However, the Dynamic Range is also limited by quantization error. Dither can be used to increase the Dynamic Range and mitigate those quantization errors (Burg, 2014). **Figure 3** below is an example of adding random noise, Dither, to Middle A.

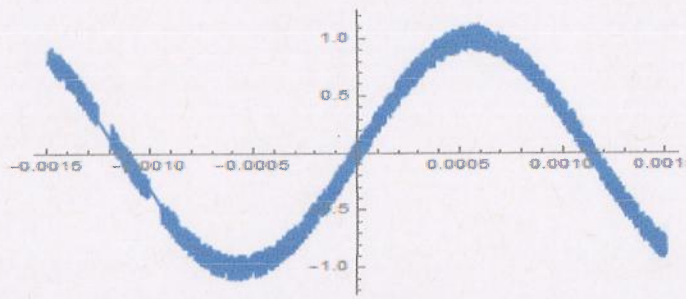

Figure 3: Adding dither to the tone, $Middle\ A = 2\pi\ \sin(440t)$.

Dither is a small amount of random noise that is added to the signal before quantization of the signal. Adding dither spreads out the errors, creating a constant hiss, or white noise, with no tonal changes, hence easy for humans to ignore. Adding dither also amplifies the lower, bass frequencies. Below is an example of dither applied to pixels in a picture.

Figure 4: Associating the concept of dither to pixels. (Redmon, 2015)

As mentioned previously, the MP3 encoder is a complex algorithm that minimizes file size to make optimal sizes for transmission and downloading of music. Within the MP3 Encoder section, I analyzed the pure tone Middle A. Middle A is smooth sine curve with no tonal changes when played, which is why a sampling rate of only 880 Hz is suitable. However, music and songs contain a variety of different sine curves. So, the suggested Nyquist sampling rate may not adequate when recording songs. This idea is explored with the following statistical experiment.

Research Question

Can an optimal sampling rate for compression be determined that minimizes MP3 file size while maintaining high audio quality?

Hypothesis

If MP3 files are compressed at a sampling rate of about 20,000 Hz, near humans' absolute threshold of hearing, then an optimal sampling rate of compression can be determined because most humans cannot comprehend frequencies above 20,000 Hz.

Experiment Description

To test this hypothesis, a hearing test was created to compare MP3 files compressed at various sampling rates. A portion of the song "Knockin' On Heaven's Door" by Bob Dylan was ripped from a CD and saved as a .wav and .mp3 at 44,100 Hz using the software *GoldWave*. The sampling rates of interest are 8,000 Hz, 11,025 Hz, 16,000 Hz, 22,050 Hz, 24,000 Hz, 32,000 Hz. Using *GoldWave*, each sample was saved as .mp3 at the specified rate and uploaded into a PowerPoint, so the test is user friendly for participants. The PowerPoint contains seven total slides with one comparison per slide. The sample comparisons on the slides were randomly selected between the seven slides. The comparisons per slide are as follows:

Slide 1 compares .mp3 22,050 Hz vs. .mp3 24,000 Hz

Slide 2 compares .mp3 16,000 Hz vs. .mp3 11,025 Hz

Slide 3 compares .mp3 11,025 Hz vs. .mp3 8,000 Hz

Slide 4 compares .wav vs. .mp3 44,100 Hz

Slide 5 compares .mp3 24,000 Hz vs. .mp3 32,000 Hz

Slide 6 compares .mp3 32,000 Hz vs. .mp3 44,100 Hz

Slide 7 compares .mp3 22,050 Hz vs. .mp3 16,000 Hz

Audio Compression

Participants were randomly selected students and faculty across the Salisbury University Campus, each between the ages of 20-40. Each participant was asked to listen and compare each of the samples per slide and determine which sample they believe has a higher quality. Participants recorded their answers on a questionnaire.

A secondary question to this experiment is whether participants with a music background can distinguish between the higher qualities more than those without a music background. During data collection participants were asked to describe their music background, so they can be assigned to either group.

Data and Results

After distributing the test and collecting data, 40 total (music and non-music combined) participants were studied. **Figure 5** is an Excel representation of the proportions of participants who selected the higher quality sample for each slide.

Slide	Comparison	Totals	Proportion Correct
1	.mp3 22,050 vs. 24,000 Hz	33	0.825
2	.mp3 16,000 vs. 11,025 Hz	40	1
3	.mp3 11,025 vs. 8,000 Hz	38	0.95
4	.wav vs. .mp3 44,100 Hz	19	0.475
5	.mp3 24,000 vs. 32,000 Hz	26	0.65
6	.mp3 32,000 vs. 44,100 Hz	26	0.65
7	.mp3 22,050 vs. 16,000 Hz	38	0.95

Figure 5: Proportion of all 40 participants with correct selection.

From **Figure 5**, every participant selected the higher quality sample in the slide 2 comparison. This leads us to believe that there are significant differences between the samples in slide 2. Slide 4, however, shows that less than half of the participants selected the higher quality response. Slide 4 compared the highest two quality samples, which we would expect to be difficult to distinguish. Saving a MP3 file at 44,100 Hz may not be necessary, and a lower sampling rate would suffice, since Slide 4 compared a lossless .wav file to a lossy MP3 file with few distinguishable differences.

Using *Minitab*, we can test for significant differences between proportions for each slide comparison. For each slide the null hypothesis is $H_0: p = 0.5$, the proportion of participants who chose the highest quality audio clip is 0.5. If the proportion is about 0.5 (50%), the participants had a difficult time determining the higher audio. However, we would expect the proportion to be greater than 0.5, since one sample does have a higher quality, so compare the null hypothesis to the following alternative hypothesis, H_A: $p > 0.5$.

The following are the *Minitab* analyses for each slide:

Slide Comparison #1: .mp3 22,050 Hz vs. .mp3 24,000 Hz

Test and CI for One Proportion: Slide 1

Method

Event: Slide 1 = 1
p: proportion where Slide 1 = 1
Exact method is used for this analysis.

Descriptive Statistics

N	Event	Sample p	95% Lower Bound for p
40	33	0.825000	0.696294

Test

Null hypothesis H_0: p = 0.5
Alternative hypothesis H_1: p > 0.5

P-Value
0.000

Minitab outputs a p-value of 0.000. Testing at $\alpha = .05$ significance level, we reject the null hypothesis and conclude that there is statistically significant evidence that more than 50% of the participants can distinguish between the lower and higher recordings. For this slide, about 83% of the participants successfully chose the higher audio quality, as shown in the Excel table of the raw data. Interpreting the confidence interval, we are 95% confident that at least 70% of all students and faculty at Salisbury

University can distinguish noticeable differences within this comparison. From this result of the sample, it can be said that students and faculty at Salisbury University can distinguish a difference between the lower quality, 22,050 Hz, and the higher quality, 24,000 Hz.

Slide Comparison #2: .mp3 16,000 Hz vs. .mp3 11,025 Hz

Minitab outputs a p-value of 0.000. Testing at $\alpha = .05$ significance level, we reject the null hypothesis and conclude that there is statistically significant evidence that more than 50% of the participants can distinguish between the lower and higher recordings. Interpreting the confidence interval, we are 95% confident that at least 93% of all students and faculty at Salisbury University can distinguish noticeable differences within this comparison. From this result of the sample, it can be said that students and faculty at Salisbury University can distinguish a difference between the lower quality, 11,025 Hz, and the higher quality, 16,000 Hz. For this slide, every participant successfully chose the higher audio quality, as shown in the Excel table of the raw data.

Test and CI for One Proportion: Slide 2

Method

Event: Slide 2 = 1
p: proportion where Slide 2 = 1
Exact method is used for this analysis.

Descriptive Statistics

N	Event	Sample p	95% Lower Bound for p
40	40	1.000000	0.927842

Test

Null hypothesis $H_0: p = 0.5$
Alternative hypothesis $H_1: p > 0.5$

P-Value
0.000

Slide Comparison #3: .mp3 11,025 Hz vs. .mp3 8,000 Hz

Minitab outputs a p-value of 0.000. Testing at $\alpha = .05$ significance level, we reject the null hypothesis and conclude that there is statistically significant evidence that more than 50% of the participants can distinguish between the lower and higher recordings. Interpreting the confidence interval, we are 95% confident that at least 85% of all students and faculty at Salisbury University can distinguish noticeable differences within this comparison. From this result of the sample, it can be said that students and faculty at Salisbury University can distinguish a difference between the lower quality, 8,000 Hz, and the higher quality, 11,025 Hz. For this slide, 95% of the participants successfully chose the higher audio quality, as shown in the Excel table of the raw data.

Test and CI for One Proportion: Slide 3

Method

Event: Slide 3 = 1
p: proportion where Slide 3 = 1
Exact method is used for this analysis.

Descriptive Statistics

N	Event	Sample p	95% Lower Bound for p
40	38	0.950000	0.850848

Test

Null hypothesis	H_0: p = 0.5
Alternative hypothesis	H_1: p > 0.5

P-Value
0.000

Audio Compression

Slide Comparison #4: .WAV vs. .mp3 44,100 Hz

Test and CI for One Proportion: Slide 4

Method

Event: Slide 4 = 1
p: proportion where Slide 4 = 1
Exact method is used for this analysis.

Descriptive Statistics

N	Event	Sample p	95% Lower Bound for p
40	19	0.475000	0.337696

Test

Null hypothesis $H_0: p = 0.5$
Alternative hypothesis $H_1: p > 0.5$

P-Value
0.682

This slide compares .mp3 at 44,100 Hz, which is considered high CD quality, to a .wav file. A .wav file is a lossless form of audio compression as previously discuss. *Minitab* outputs a p-value of 0.682. Testing at $\alpha = .05$ significance level, we **fail to reject** the null hypothesis and conclude that there is **not statistically significant evidence** that more than 50% of the participants can distinguish between the lower and higher recordings. Interpreting the confidence interval, we are 95% confident that at least 34% of all students and faculty at Salisbury University can distinguish noticeable differences within this comparison. From this result of the sample, it cannot be said that students and faculty at Salisbury University can distinguish a difference between the lower quality, .mp3 44,100 Hz, and the higher quality, .wav file. For this slide, only about 48% of participants successfully chose the higher audio quality, as shown in the Excel table of the raw data. This is an interesting result in that participants struggled differentiating between a .wav lossless format (more data points) and a lossy format (fewer data points), .mp3. This suggests that a high quality .mp3 file such as 44,100 Hz may be used in replace of a lossless format without the majority of people being able to distinguish the difference in quality.

Audio Compression

Slide Comparison #5: .mp3 24,000 Hz vs. .mp3 32,000 Hz

Test and CI for One Proportion: Slide 5

Method

Event: Slide 5 = 1
p: proportion where Slide 5 = 1
Exact method is used for this analysis.

Descriptive Statistics

N	Event	Sample p	95% Lower Bound for p
40	26	0.650000	0.508054

Test

Null hypothesis	H_0: p = 0.5
Alternative hypothesis	H_1: p > 0.5

P-Value
0.040

Minitab outputs a p-value of 0.040. Testing at $\alpha = .05$ significance level, we reject the null hypothesis and conclude that there is statistically significant evidence that more than 50% of the participants can distinguish between the lower and higher recordings. Interpreting the confidence interval, we are 95% confident that at least 51%, barely over 50%, of all students and faculty at Salisbury University can distinguish noticeable differences within this comparison. From this result of the sample, it can be said that students and faculty at Salisbury University can distinguish a difference between the lower quality, 24,000 Hz, and the higher quality, 32,000 Hz. For this slide, 65% of the participants successfully chose the higher audio quality, as shown in the Excel table of the raw data. These conclusions, however, are base on $\alpha = .05$ significance level. If analysis is done at $\alpha = .01$, we would fail to reject the null hypothesis. So this comparison has a boardline result of whether to reject the null hypothesis or not.

Audio Compression

Slide Comparison #6: .mp3 32,000 Hz vs. .mp3 44,100 Hz

Minitab outputs a p-value of 0.040. Testing at $\alpha = .05$ significance level, we reject the null hypothesis and conclude that there is statistically significant evidence that more than 50% of the participants can distinguish between the lower and higher recordings. Interpreting the confidence interval, we are 95% confident that at least 51%, barely over 50%, of all students and faculty at Salisbury University can distinguish noticeable differences within this comparison. From this result of the sample, it can be said that students and faculty at Salisbury University can distinguish a difference between the lower quality, 32,000 Hz, and the higher quality, 44,100 Hz. For this slide, 65% of the participants successfully chose the higher audio quality, as shown in the Excel table of the raw data. These conclusions, however, are base on $\alpha = .05$ significance level. If analysis is done at $\alpha = .01$, we would fail to reject the null hypothesis. So this comparison has a boardline result of whether to reject the null hypothesis or not. 32,000 Hz and 44,100 Hz are both high sampling rates, so this is an interesting result in that depending on the significance level, the Salisbury University participants may or may not be able to distingiush differences.

Test and CI for One Proportion: Slide 6

Method

Event: Slide 6 = 1
p: proportion where Slide 6 = 1
Exact method is used for this analysis.

Descriptive Statistics

N	Event	Sample p	95% Lower Bound for p
40	26	0.650000	0.508054

Test

Null hypothesis	H_0: p = 0.5
Alternative hypothesis	H_1: p > 0.5

P-Value
0.040

Slide Comparison #7: .mp3 22,050 Hz vs. .mp3 16,000 Hz

Test and CI for One Proportion: Slide 7

Method

Event: Slide 7 = 1
p: proportion where Slide 7 = 1
Exact method is used for this analysis.

Descriptive Statistics

N	Event	Sample p	95% Lower Bound for p
40	38	0.950000	0.850848

Test

Null hypothesis	H_0: p = 0.5
Alternative hypothesis	H_1: p > 0.5

P-Value
0.000

Minitab outputs a p-value of 0.000. Testing at $\alpha = .05$ significance level, we reject the null hypothesis and conclude that there is statistically significant evidence that more than 50% of the participants can distinguish between the lower and higher recordings. Interpreting the confidence interval, we are 95% confident that at least 85% of all students and faculty at Salisbury University can distinguish noticeable differences within this comparison. From this result of the sample, it can be said that students and faculty at Salisbury University can distinguish a difference between the lower quality, 16,000 Hz, and the higher quality, 22,025 Hz. For this slide, 95% of the participants successfully chose the higher audio quality, as shown in the Excel table of the raw data.

In general, from these *Minitab* analyses, students and faculty from Salisbury University can distinguish differences in audio quality between sampling rates as high as 32,000 Hz and 44,100 Hz. So, can an optimal sampling rate be within 32,000 Hz and 44,100 Hz?

Another question that was considered during data collection and analyses is whether individuals who have a music background (music major, play an instrument, study music in school, teach music topics, etc.) perform better on the hearing test compared to those individuals who have little to no music experience.

During data collection, participants were asked to describe their music experience. Of the 40 total participants, 21 were "non-music" and 19 were "music". The following raw data is an Excel representation of the proportions of non-music and music participants who selected the higher quality sample for each slide.

Slide	Totals	Comparison	Proportion Correct
1	17	.mp3 22,050 vs. 24,000 Hz	0.80952381
2	21	.mp3 16,000 vs. 11,025 Hz	1
3	19	.mp3 11,025 vs. 8,000 Hz	0.904761905
4	9	.wav vs. .mp3 44,100 Hz	0.428571429
5	13	.mp3 24,000 vs. 32,000 Hz	0.619047619
6	11	.mp3 32,000 vs. 44,100 Hz	0.523809524
7	19	.mp3 22,050 vs. 16,000 Hz	0.904761905

Figure 6: Proportion of Non-Music Group with correct selection.

Slide	Comparison	Totals	Proportion Correct
1	.mp3 22,050 vs. 24,000 Hz	16	0.842105263
2	.mp3 16,000 vs. 11,025 Hz	19	1
3	.mp3 11,025 vs. 8,000 Hz	19	1
4	.wav vs. .mp3 44,100 Hz	9	0.473684211
5	.mp3 24,000 vs. 32,000 Hz	13	0.684210526
6	.mp3 32,000 vs. 44,100 Hz	15	0.789473684
7	.mp3 22,050 vs. 16,000 Hz	19	1

Figure 7: Proportion of Music Group with correct selection.

From **Figure 6** and **Figure 7**, every participant selected the higher quality sample in the slide 2 comparison. Each participant in the music group, however, was able to determine the higher quality on slides 2, 3, and 7. The music group also appears to perform slightly better on all seven slides when comparing these initial proportions. It appears the music participants of Salisbury University, may be better at distinguishing the higher quality audio than those non-music participants. So, an audio quality of 32,000 Hz to 44,100 Hz may not be high enough.

The following *Minitab* analyses are tests between proportions for each of the seven slides comparing musicians and non-musicians. Data in *Minitab* was coded with 0's and 1's. 0's denote those participants who are non-music and 1's denote the music group. Analyzing the differences between two proportions is testing the null hypothesis that $H_0: p_1 - p_2 = 0$ where p_1 is the proportion of non-music participants and p_2 is the proportion of music participants. We would expect that $p_2 > p_1$ so we test the alternative hypothesis $H_A:\ p_1 - p_2 < 0$.

Music vs. Non-Music Comparisons: Slide 1

Minitab outputs a p-value of 0.559 for Fisher's Exact. Testing at $\alpha = .05$ significance level, we **fail to reject** the null hypothesis and conclude that there is **not statistically significant evidence** that the proportion of music participants from Salisbury University who selected the higher quality response is greater than the proportion of non-music participants

Test and CI for Two Proportions: Slide 1, Music

Method

Event: Slide 1 = 1
p_1: proportion where Slide 1 = 1 and Music = 0
p_2: proportion where Slide 1 = 1 and Music = 1
Difference: $p_1 - p_2$

Descriptive Statistics: Slide 1

Music	N	Event	Sample p
0	21	17	0.809524
1	19	16	0.842105

Estimation for Difference

Difference	95% Upper Bound for Difference
-0.0325815	0.164394

CI based on normal approximation

Test

Null hypothesis $H_0: p_1 - p_2 = 0$
Alternative hypothesis $H_1: p_1 - p_2 < 0$

Method	Z-Value	P-Value
Normal approximation	-0.27	0.393
Fisher's exact		0.559

The normal approximation may be inaccurate for small samples from Salisbury University. For this slide, about 81% of the non-music participants successfully chose the higher audio quality and 84% of the music participants successfully chose the higher audio, as shown in the Excel table of the raw data. Interpreting the confidence interval, we are 95% confident that at least 16% of all musicians and non-musicians at Salisbury University can distinguish noticeable differences within this comparison. From this result of the sample, there is not a significant difference between music and non-music Salisbury University students' and faculty's ability to distinguish a difference between the lower quality, 22,050 Hz and the higher quality, 24,000 Hz.

Music vs. Non-Music Comparisons: Slide 2

Minitab outputs a p-value of 1.000 for Fisher's Exact. Testing at $\alpha = .05$ significance level, we **fail to reject** the null hypothesis and conclude that there is **not statistically significant evidence** that the proportion of music participants from Salisbury University who selected the higher quality response is greater than the proportion of non-music participants from Salisbury University.

Test and CI for Two Proportions: Slide 2, Music

Method

Event: Slide 2 = 1
p_1: proportion where Slide 2 = 1 and Music = 0
p_2: proportion where Slide 2 = 1 and Music = 1
Difference: $p_1 - p_2$

Descriptive Statistics: Slide 2

Music	N	Event	Sample p
0	21	21	1.000000
1	19	19	1.000000

Estimation for Difference

Difference	95% Upper Bound for Difference
0	*

CI based on normal approximation

Test

Null hypothesis $H_0: p_1 - p_2 = 0$
Alternative hypothesis $H_1: p_1 - p_2 < 0$

Method	Z-Value	P-Value
Normal approximation	*	*
Fisher's exact		1.000

The normal approximation may be inaccurate for small samples.

For this slide, every participant selected the higher audio response, corresponding to a p-value of 1.000. From this result of the sample, there is not a significant difference between music and non-music Salisbury University students' and faculty's ability to distinguish a difference between the lower quality, 11,025 Hz and the higher quality, 16,000 Hz.

Music vs. Non-Music Comparisons: Slide 3

Minitab outputs a p-value of 0.269 for Fisher's Exact. Testing at $\alpha = .05$ significance level, we **fail to reject** the null hypothesis and conclude that there is **not statistically significant evidence** that the proportion of music participants from Salisbury University who selected the higher quality response is greater than the proportion of non-music participants from Salisbury University. For this slide, about 90% of the non-music participants successfully chose the higher audio quality and all the music participants successfully chose the higher audio, as shown in the Excel table of the raw data. From this result of the sample, there is not a significant difference between music and non-music Salisbury University students' and faculty's ability to distinguish a difference between the lower quality, 8,000 Hz and the higher quality, 11,025 Hz.

Test and CI for Two Proportions: Slide 3, Music

Method

Event: Slide 3 = 1
p_1: proportion where Slide 3 = 1 and Music = 0
p_2: proportion where Slide 3 = 1 and Music = 1
Difference: $p_1 - p_2$

Descriptive Statistics: Slide 3

Music	N	Event	Sample p
0	21	19	0.904762
1	19	19	1.000000

Estimation for Difference

Difference	95% Upper Bound for Difference
-0.0952381	0.010125

CI based on normal approximation

Test

Null hypothesis $H_0: p_1 - p_2 = 0$
Alternative hypothesis $H_1: p_1 - p_2 < 0$

Method	Z-Value	P-Value
Normal approximation	-1.49	0.069
Fisher's exact		0.269

The normal approximation may be inaccurate for small samples.

Music vs. Non-Music Comparisons: Slide 4

Minitab outputs a p-value of 0.382 for Fisher's Exact. Testing at $\alpha = .05$ significance level, we **fail to reject** the null hypothesis and conclude that there is **not statistically significant evidence** that the proportion of music participants from Salisbury University who selected the higher quality response is greater than the proportion of non-music participants from Salisbury University. For this slide, about 43% of the non-music participants successfully chose the higher audio quality and 47% of the music participants successfully chose the higher audio, as shown in the Excel table of the raw data. From this result of the sample, there is not a significant difference between music and non-music Salisbury University students' and faculty's ability to distinguish a difference between the lower lossy quality, .mp3 44,100 Hz and the higher lossless quality, .wav file.

Test and CI for Two Proportions: Slide 4, Music

Method

Event: Slide 4 = 1
p_1: proportion where Slide 4 = 1 and Music = 0
p_2: proportion where Slide 4 = 1 and Music = 1
Difference: $p_1 - p_2$

Descriptive Statistics: Slide 4

Music	N	Event	Sample p
0	21	9	0.428571
1	19	10	0.526316

Estimation for Difference

Difference	95% Upper Bound for Difference
-0.0977444	0.161200

CI based on normal approximation

Test

Null hypothesis H_0: $p_1 - p_2 = 0$
Alternative hypothesis H_1: $p_1 - p_2 < 0$

Method	Z-Value	P-Value
Normal approximation	-0.62	0.267
Fisher's exact		0.382

Music vs. Non-Music Comparisons: Slide 5

Minitab outputs a p-value of 0.461 for Fisher's Exact. Testing at $\alpha = .05$ significance level, we **fail to reject** the null hypothesis and conclude that there is **not statistically significant evidence** that the proportion of music participants from Salisbury University who selected the higher quality response is greater than the proportion of non-music participants from Salisbury University. For this slide, about 62% of the non-music participants successfully chose the higher audio quality and about 68% of the music participants successfully chose the higher audio, as shown in the Excel table of the raw data. From this result of the sample, there is not a significant difference between music and non-music Salisbury University students' and faculty's ability to distinguish a difference between the lower quality, 24,000 Hz and the higher quality, 32,000 Hz.

Test and CI for Two Proportions: Slide 5, Music

Method

Event: Slide 5 = 1
p_1: proportion where Slide 5 = 1 and Music = 0
p_2: proportion where Slide 5 = 1 and Music = 1
Difference: $p_1 - p_2$

Descriptive Statistics: Slide 5

Music	N	Event	Sample p
0	21	13	0.619048
1	19	13	0.684211

Estimation for Difference

Difference	95% Upper Bound for Difference
-0.0651629	0.182123

CI based on normal approximation

Test

Null hypothesis $H_0: p_1 - p_2 = 0$
Alternative hypothesis $H_1: p_1 - p_2 < 0$

Method	Z-Value	P-Value
Normal approximation	-0.43	0.332
Fisher's exact		0.461

Music vs. Non-Music Comparisons: Slide 6

Minitab outputs a p-value of 0.076 for Fisher's Exact. Testing at $\alpha = .05$ significance level, we **fail to reject** the null hypothesis and conclude that there is **not statistically significant evidence** that the proportion of music participants from Salisbury University who selected the higher quality response is greater than the proportion of non-music participants from Salisbury University. For this slide, about 52% of the non-music participants successfully chose the higher audio quality and about 79% of the music participants successfully chose the higher audio, as shown in the Excel table of the raw data. From this result of the sample, there is not a significant difference between music and non-music Salisbury University students' and faculty's ability to distinguish a difference between the lower quality, 32,000 Hz and the higher quality, 44,100 Hz.

Test and CI for Two Proportions: Slide 6, Music

Method

Event: Slide 6 = 1
p_1: proportion where Slide 6 = 1 and Music = 0
p_2: proportion where Slide 6 = 1 and Music = 1
Difference: $p_1 - p_2$

Descriptive Statistics: Slide 6

Music	N	Event	Sample p
0	21	11	0.523810
1	19	15	0.789474

Estimation for Difference

Difference	95% Upper Bound for Difference
-0.265664	-0.029438

CI based on normal approximation

Test

Null hypothesis $H_0: p_1 - p_2 = 0$
Alternative hypothesis $H_1: p_1 - p_2 < 0$

Method	Z-Value	P-Value
Normal approximation	-1.85	0.032
Fisher's exact		0.076

The normal approximation may be inaccurate for small samples.

Audio Compression

Music vs. Non-Music Comparisons: Slide 7

Minitab outputs a p-value of 0.269 for Fisher's Exact. Testing at $\alpha = .05$ significance level, we **fail to reject** the null hypothesis and conclude that there is **not statistically significant evidence** that the proportion of music participants from Salisbury University who selected the higher quality response is greater than the proportion of non-music participants from Salisbury University. For this slide, about 90% of the non-music participants successfully chose the higher audio quality and all the music participants successfully chose the higher audio, as shown in the Excel table of the raw data. From this result of the sample, there is not a significant difference between music and non-music Salisbury University students' and faculty's ability to distinguish a difference between the lower quality, 16,000 Hz and the higher quality, 22,025 Hz.

Test and CI for Two Proportions: Slide 7, Music

Method

Event: Slide 7 = 1
p_1: proportion where Slide 7 = 1 and Music = 0
p_2: proportion where Slide 7 = 1 and Music = 1
Difference: $p_1 - p_2$

Descriptive Statistics: Slide 7

Music	N	Event	Sample p
0	21	19	0.904762
1	19	19	1.000000

Estimation for Difference

Difference	95% Upper Bound for Difference
-0.0952381	0.010125

CI based on normal approximation

Test

Null hypothesis $H_0: p_1 - p_2 = 0$
Alternative hypothesis $H_1: p_1 - p_2 < 0$

Method	Z-Value	P-Value
Normal approximation	-1.49	0.069
Fisher's exact		0.269

The normal approximation may be inaccurate for small samples

In general, from the above *Minitab* results, no significant differences exist between non-music participants of Salisbury University and music participants. Note that the sample sizes for each the music and non-music groups are considered small. Given more time to collect data, significant results may have been determined.

Conclusion and Discussion

According to the Absolute Threshold of Hearing, humans hear frequencies within the range of 20 Hz – 20,000 Hz; however, as humans age it becomes more difficult to perceive high frequencies. According to the Nyquist frequency theorem, 32,000 Hz and 44,100 Hz should have resulted in significant differences. But are those the frequencies most people who are 20-40 years old typically hear? This may explain why some participants had trouble distinguishing between the higher frequencies.

From this research, we see that the participants in the sample had difficulty distinguishing between higher and lower sampling rates when 44,100 Hz was reached. As of now, CD's are compressed at 44,100 Hz with a sample depth of 16 bits, so this study supports the compression standard used within CD's. From background research, however, most music is now recorded at sampling rates even greater than 44,100 Hz and sample depths larger than 16 bits. As future research, it would be interesting to study even higher sampling rates and greater bit depths and compare the results to those in this study, which were at low sampling rates and constant sample depth.

Even CD's are not the primary form of listening music anymore, many musicians and music enthusiasts listen and use downloadable forms of songs. Therefore, a smaller file size that preserves high quality is critical in the distribution and streaming of music. Even though a human's absolute threshold of hearing says that many people cannot perceive frequencies greater than 20,000 Hz, this study shows that audio files must be saved at higher sampling rates (44,100 Hz and greater) to ensure the highest quality that individuals can enjoy.

Future Projects and Questions

- How can a variable bitrate be implemented that is compatible with MP3? How does this change the encoding process of MP3? Will it be better than using a constant bitrate?
- Some people prefer the lower quality of the original sound - vinyl for example. How does that effect the compression standard used?
- Problems that correspond to the Nyquist Theorem and how they can be alleviated i.e. dither. In general, I would like to look more into dither and its other purposes in audio signals and even images.
- Operations of the MP3 decoder.

References

Burg, J, Romney, J. (2014). *Digital Sound & Music: Concepts, Applications, and Science.* Retrieved from http://digitalsoundandmusic.com/

Olshausen, B. A. (2000). *Aliasing* (PSC 129). Berkeley, California: Redwood Center for Theoretical Neuroscience.

Raissi, R. (2002). The Theory Behind Mp3. *MP3' Tech.* Retrieved from http://www.mp3-tech.org/

Redmon, N. (2015). Digital Audio. *EarLevel Engineering.* Retrieved from http://www.earlevel.com/main/?s=Dither

Sayood, K. (2012). *Introduction to Data Compression. Chap. 17 Audio Coding.* Waltham, MA: Morgan Kaufmann Publisher.

Wilburn T. (2007, 10, 4). The AudioFile: Understanding MP3 compression. Retrieved from https://arstechnica.com/features/2007/10/the-audiofile-understanding-mp3-compression/