Title: Investigating the effect of MP3 and FLAC compression algorithms on sound files

Research Question: To what extent are MP3 and FLAC compression algorithms usable in the field of bioacoustics (specifically for birds) in terms of audio quality and compression ratio?

Subject: Computer Science

Word count: 3996

Table of Content

Introduction

Animal acoustic sounds have been an important study system for a wide variety of fields in ecology and evolution (Araya-Salas et al.). While one would intuitively think that videos are a better method to draw biological conclusions, sound recordings give scientists a better understanding on how species are interacting or changing in response to climate change, human disturbance, or habitat destruction (Welz). Araya-Salas et al. explains that the particular value of bioacoustic analysis lies in our ability to digitally record acoustic signals, conduct detailed measures to characterize signal structure (tone, pitch, spatial location, etc) with spectrograms, and match acoustic patterns to bird species (Farina and Gage 87). Thus, recording quality could pose a potential challenge to carry out precise analyses. Background and external noise can mask important signals, affecting the information received at the end-result; hence, noise is considered to fundamentally shape acoustic signals.

Bioacoustic recordings can be rather long; i.e. night soundscapes may create recordings of several hours. Furthermore, recording in an uncompressed format results in large sound files, a disadvantage when needing to manipulate, edit, or analyze signals. Thus, compression becomes a strategic solution, as lossy and lossless codecs[1] have the ability to (at least) halve the original file size. Such dramatic decrease in file size might come with a cost of quality; yet a compressed file must keep bird signals relatively unchanged and adequate levels of noise to be usable in an analysis of bird signals.

Therefore, in this study, I evaluate whether MP3 (lossy) and FLAC (lossless) algorithms can be applied for bioacoustic analysis. To do so, I review the literature to understand the principles of digital audio and compression and to study the MP3 and FLAC encoding processes in depth. Then, I conduct a quantitative and qualitative experiment to evaluate the algorithms's effects on compression ratio and audio quality, measured by the noise in the recording and deviation in bird signals. These two algorithms

[1] A codec compresses audio files

are chosen uniquely for their global popularity, as they are the leading algorithms in their type (Nugent). Finally, based on my research and results, I discuss to what extent these algorithms can be practically used in a bird acoustic analysis.

Digital Audio

Analog vs. Digital recording

Once sound is emitted, it produces a pressure wave that propagates through the air which in turn creates a vibration in the diaphragm of a recording device. With the help of transducers, the vibration creates an electrical signal, in voltage, that varies continuously with the amplitude of the wave in the air (Bartlett and Bartlett 548). This continuous and proportionate variation creates an analog wave, representing a change of amplitude over time (*Figure 1*). In an analog system, the voltages, representing amplitude, are stored in wax, vinyl disks, or magnetic tapes in a continuous manner ("Understanding Sample"). However, in a digital system, the continuous voltages are sampled and quantized at discrete intervals, allocating a binary word that quantifies the voltage for a sample (*Figure 1*). To decode the digital sound and produce playable analog waves, the same process is inverted.

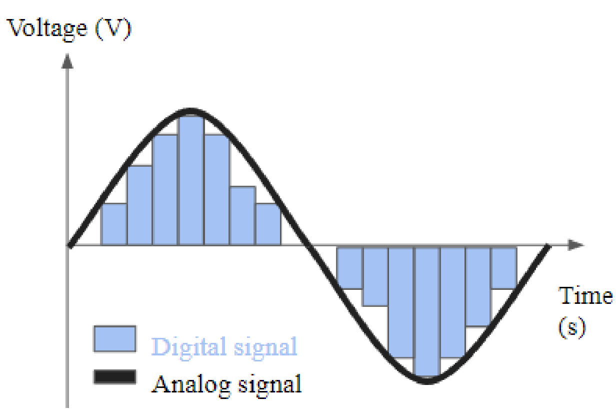


Figure 1 - Analog vs. Digital signal. By author based on Gans

How is a sound recorded?

Compressed formats, such as MP3 and FLAC, derive from uncompressed recordings, which use Pulse-Code-Modulation[2] (PCM) to digitally represent analog signals. This section explains sampling and quantization: two PCM processes that are determining factors in the audio quality of an uncompressed and compressed file.

The core conversion lies in the analog-to-digital (ADC) converter where sampling and quantization occur. Sampling represents the time of a measurement, and the number of samples per second can vary between files (*Figure 2*) (Pohlmann 28). These rates can go as high as 192kHz per second in raw recording, resulting in a top-level frequency bandwidth, but also resulting in a large file. However, one rule exists; the Nyquist Theory states that the sampling rate must be at least twice the frequency of the signal being sampled to successfully reconstruct the same analog signal (*Figure 3*) (Elsea). To respect this rule prior to sampling, the PCM process starts with an anti-aliasing filter that gets rid of the frequencies that are above the frequency limit (Bartlett and Bartlett 548). Thus, the

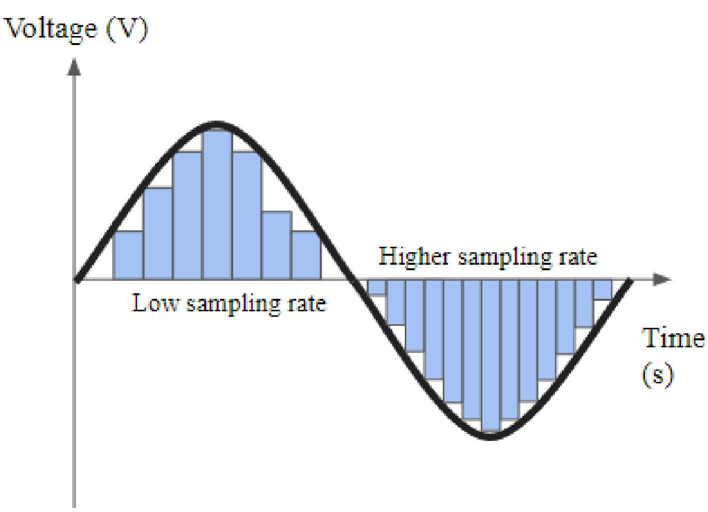


Figure 2 - Rates. By author based on Bartlett

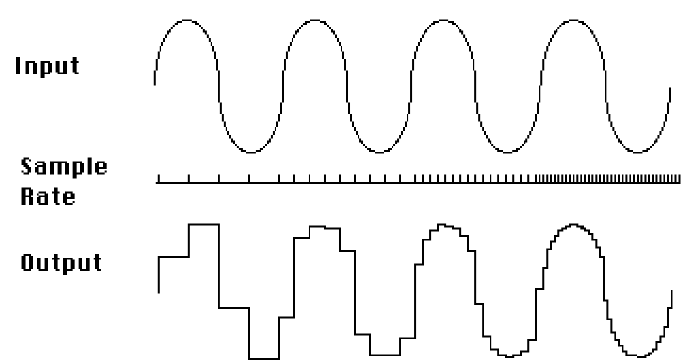


Figure 3 - Effect of increasing sampling rate (Elsea)

---

[2] Method that represents analog signals digitally (Pohlmann 28)

standard sampling frequency became 44.1kHz, as it maintains a well-rounded trade off between file size and frequency bandwidth, while respecting the Nyquist Theory, considering that humans have a bandwidth of 20Hz to 20kHz for audible sounds (Pohlmann 45).

While sampling represents the time of the measurement, quantization represents the value of the measurement, and in the case of audio, the amplitude of the waveform at a sample time. Having a limited number of bits available, quantization maps the analog value to a discrete amplitude level (Pohlmann 29). A higher bit-depth, number of bits available, means more levels of amplitude which improves the accuracy of each measurement and the dynamic range of the recording (*Figure 4*). Still, some measurements would inaccurately represent the actual amplitude of a signal, so quantization error, which measures the difference between the actual analog value and the selected quantized value, occurs. Such error produces unpleasant distortions in the recording; thus increasing the bit-depth, along with other techniques, improves the resolution of the signal and reduces distortion noise ("Quantization - Digital" 5:00). A compromise between resolution and file size was found, and 16 bit-depth became the standard (Pohlmann 45).

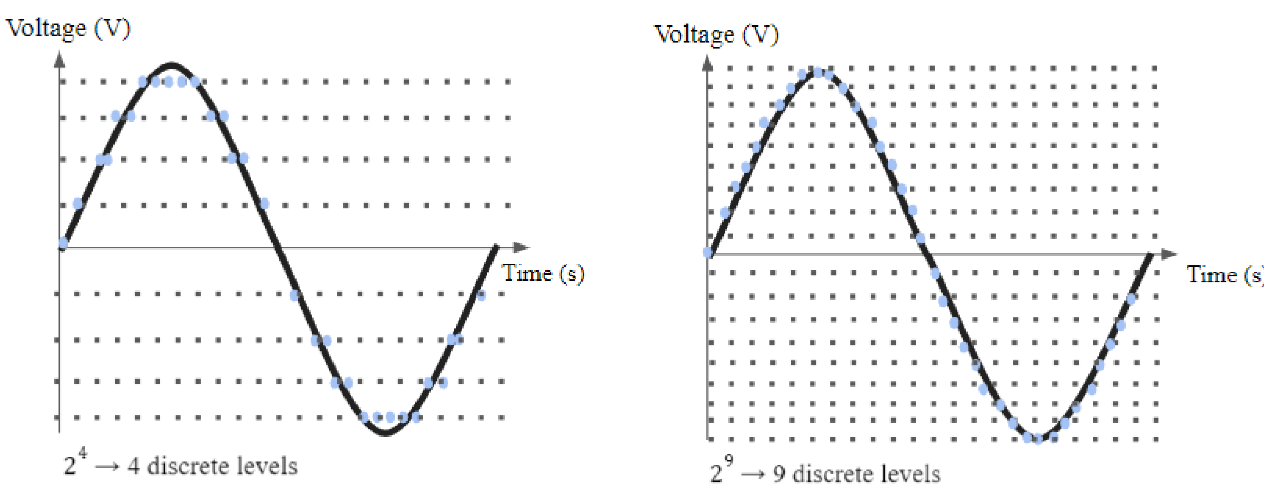


Figure 4 - Difference between bit-depth. By author based on “Quantization - Digital”

Signal to noise ratio (SNR)

In the field of bioacoustic, recording equipment and background noises may interfere with the recording of a particular bird. This external sound along with quantization error creates a noise floor. In a recording, the energy in the noise floor can be compared to the level of signal power, or the dynamic range, to obtain the signal-to-noise ratio (SNR) (*Figure 5*) (Pohlmann 349). Expressed in decibels, it measures the proportion between the strength of the unwanted noise and the audio signal emitted. In this context, a SNR of at least 6dB, indicates that the noise does not hold enough amplitude to critically interfere with bird signals (Araya-Salas et al. 57). Hence, understanding this concept becomes crucial to evaluate the data in the experiments, as noise levels determine the audio quality of a sound file and the ability to analyze bird frequency patterns ("Signal to Noise").

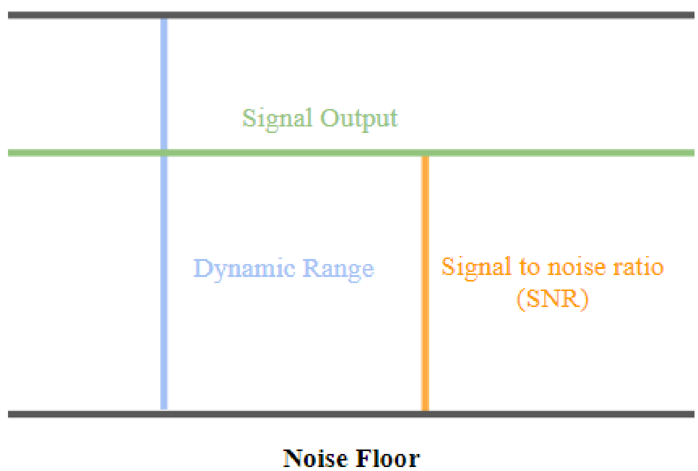


Figure 5 - SNR. By author based on Hahn

Audio Compression

What is audio compression?

Audio compression is the process of decreasing an audio file's size by reducing its amount of data to varying extents. The many forms of audio compression techniques offer a range of encoder and decoder complexity, compressed audio quality, and amounts of data compression. This essay studies and compares the two specific processes: lossy and lossless audio compression (Yes Pan 2).

Why is audio compression needed?

In a world where users can stream millions of songs on the go, compression becomes pretty crucial. In its utter purpose, it reduces the size of audio files, allowing more efficient storage and data transmission across networks (Yes Pan 1). People are still interested in listening to adequate audio efficiently, and compression finds the right balance between audio quality and file size.

More specifically in the context of Bioacoustics, audio compression becomes intuitively more essential. Though recordings may be short, ornithologists generally record soundscapes in an uncompressed format for long periods of time (i.e. during the night), resulting in large files that are difficult to transfer, store, edit, and analyze. To illustrate the amount of data stored in a raw audio file, let's calculate the size (in MB) of a 2 hour long file, recorded with a sampling rate of 44.1kHz, a 16 bit-depth, and with 2 channels.

Formula:

$$\frac{Sample\ Rate(Hz) \times Length\ of\ recording\ (seconds) \times BitDepth\ (bits) \times \#\ of\ channels}{8{,}000{,}000\ (for\ MB)}$$

$$\frac{44100 \times 7200 \times 16 \times 2}{8{,}000{,}000} \approx \ 1,270.08MB$$

The calculation results in 1270.01 MB for 2 hours. This is a very large value. It is inefficient to record large audio files, considering that some scientists record longer audio sequences. Thus, investigating whether lossy and lossless compression can be used for bioacoustic analysis is a pertinent issue, as both algorithms can compress an audio file to half its size, or more (Pohlmann 404).

Lossy audio algorithms

In today's world, there is a cost for storage, and for applications, it is essential that files are coded as bit-efficient as possible. By applying the principles of psychoacoustics[3] to model the perceptual qualities of the human ear, lossy compression algorithms reduce the binary word length of imperceptible signals, drastically decreasing the file size at no perceptual cost (Kavitha 1-2).

In general, lossy compression considerably decreases the bit-rate (amount of bits per seconds) of an audio file, oftentimes to more than 50%. Such a difference may seem absurd, but lossy algorithms code signals in a more intelligent manner. Thus, the art and science behind lossy coding conveys high-quality audio for the human ear, while requiring a fraction of the information needed by an uncompressed format (Pohlmann 352-353).

[3] Scientific study that explains human auditory perception (Pohlmann 404).

Lossless audio algorithms

In lossy compression, the final output differs from the original file in ways that “may or may not be audible” (Pohlmann 384). This drastically differs in lossless compression, as a decompressed file matches bit for bit with the original data. Lossless encoders create a smaller coded file which is decoded into the original uncompressed audio file at playback. In other words, the lossless coded file acts as an intermediate optimization of storage on a device, as the quality remains the same at decoding (Maher 257). However, keeping the same quality comes at a cost, as lossless files weigh heavier than their counterparts. For some, the cost is worth it.

The difference between lossy and lossless algorithms lies in their methods for compression; as mentioned earlier, lossy compression relies on data irrelevancy, but lossless compression “operates strictly on redundancy” (Pohlmann 384). The codec compresses redundant data using indexing methods, reducing file size in temporary storage, and then reverses the process to retrieve the original data.

Algorithms

In order to effectively answer the research question, it is important to study and understand the methodologies of compression for the MP3 and FLAC compression algorithms.

MP3

In 1993, the MPEG-1-Audio-Layer-3 (MP3) lossy algorithm received unprecedented levels of popularity as the developers made the sample code available to the public, enabling the entire world to encode music (Hacker 25; Fraunhofer). This paper discusses the MP3, as it is the most commonly used lossy codec.

Process

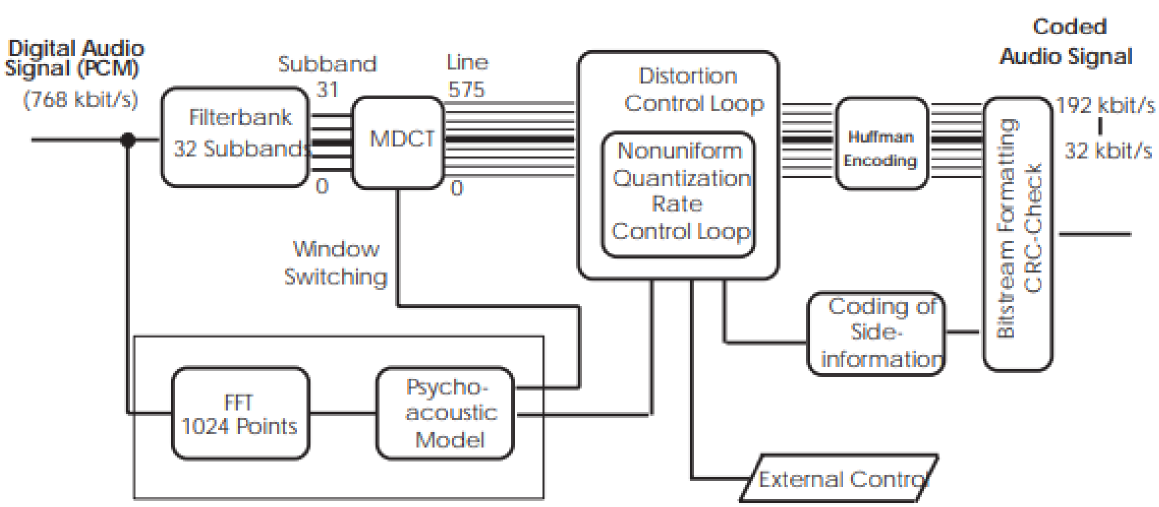


Figure 6 - Process of MP3 (Sripada 15)

Analysis

First, in MP3 compression, samples of a PCM file are processed in frames of 1152 samples, and are run through the polyphase filterbank. In this analysis, the samples are divided into 32 equal frequency sub-bands; i.e if the Nyquist frequency is 22.05 kHz, then each subband will be approximately 689 (22,050 / 32) Hz wide (Raissi 20). In MP2, the 32 values are used for encoding, yet MP3 further divides these sub-bands into 576 smaller bands, by applying a modified discrete cosine transform (MDCT) to each frame of subband samples (Wilburn; Raissi 20). MP3 needs spectral information from the MCDT to compress a signal instead of operating with waveforms, as spectral information resembles the way humans interpret audio. Not only does spectral separation increase the potential for data removal, but dividing frames into 576 bands, each containing 1/576 of the frequency range, is a crucial first step in realizing a powerful and successful compression (Wilburn).

As seen in *Figure 6,* two parallel processes take place in the first stage of MP3 compression; as the 1152 sample frame runs through the filterbank and the MDCT, the same frames are transformed into information that can be analyzed by the perceptual model (Raissi 21). The Fast Fourier Transform uses complex mathematical operations to change the domain of samples from time to frequency, making it easier for the analysis of a signal consisting of multiple frequencies (Maklin). Now in the frequency domain, the FFT output becomes the input of the psychoacoustic model block.

The psychoacoustic model is the algorithm that distinguishes an encoder's quality. Its output determines the window types that the MDCT applies and provides the instructions for quantization in the Nonuniform Quantization block (Raissi 21). For the MP3, a perceptual model is the core element that defines which signals are kept based on various psychoacoustic principles. For example, temporal masking (as seen in *figure 7*) dictates that a stronger signal will mask a weaker one if they appear within a small (milliseconds) interval of time, subsequently assigning less data to the weaker signal (Raissi 5).

In addition to masking, the psychoacoustic model may also set a minimum audition threshold to toss out the quieter signals that the human ear cannot specifically perceive. In this way, perceptual algorithms produce information that determines the sounds that can be safely discounted, which is the nature of lossy compression. (Wilburn).

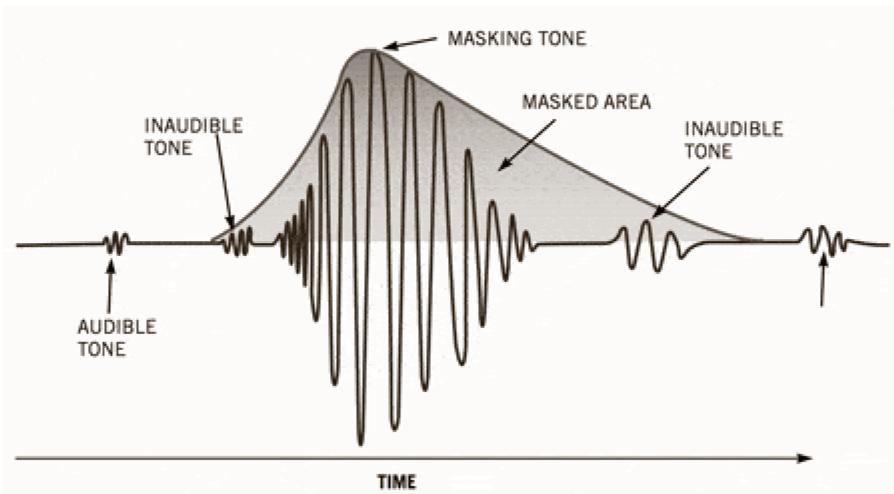


Figure 7- Temporal Masking (Jared)

Compression

Quantization is where the bit allocation takes place and Huffman coding indexes shorter code words to quantized values based on their recurrence in the frame. These two processes work together in a system of two nested iteration loops (Raissi 23). Since noise shaping is done to reduce the quantization noise before Huffman coding, a global variable (that determines the quantization step size) and scale factors (which determines the noise-shaping factors for each band) are established (Brandenburg and Popp 8). Thus, the goal of the two nested loops is to find a balance between the quantization step size and scale factors depending on the bit-rate chosen. *Figure 8* shows the two loops. Finally, in the bitstream block, the frame, representing 1152 encoded PCM samples, is put together with the information to decode it (*Figure 9)* (Brandenburg and Popp 31).

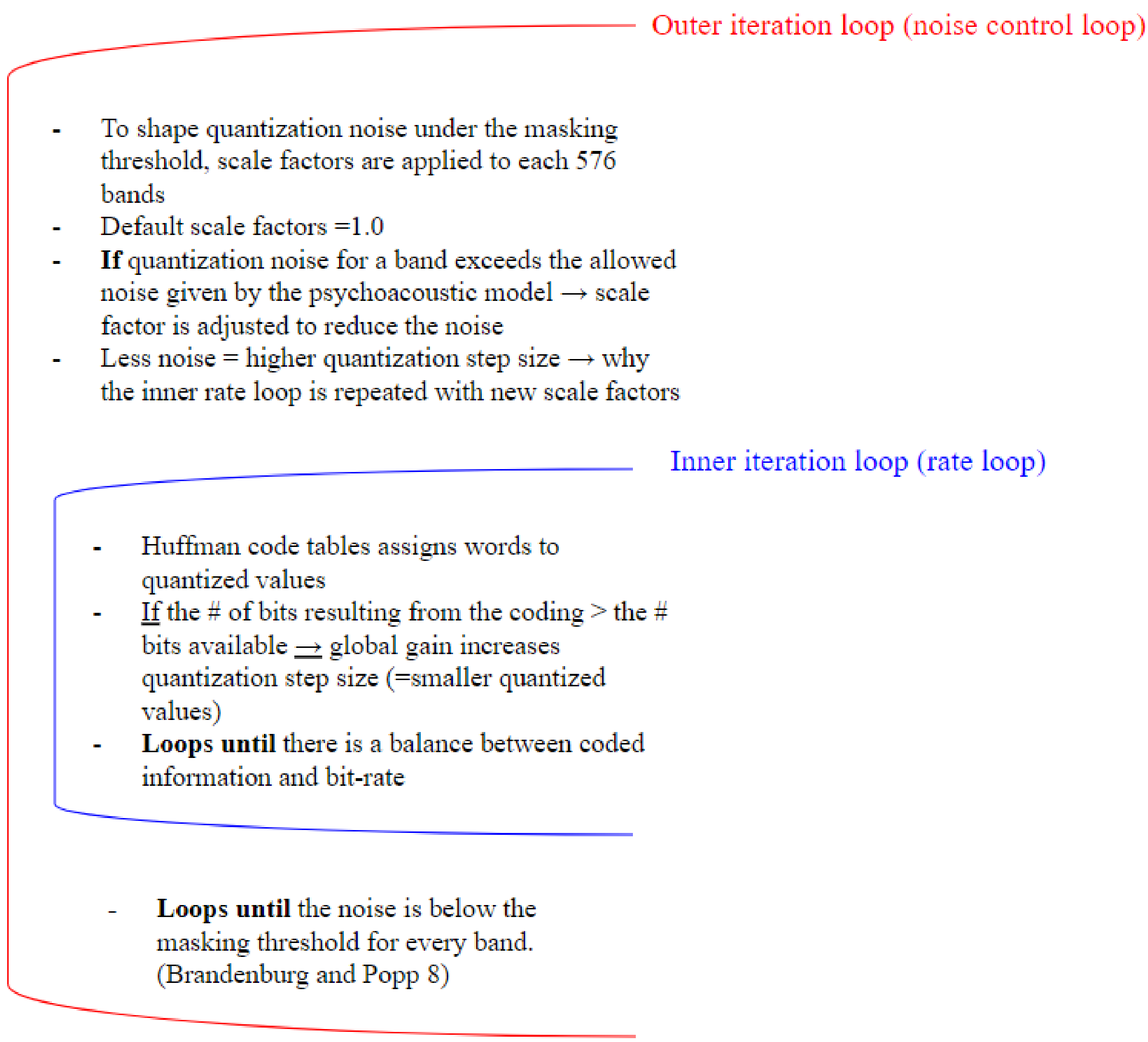


Figure 8 - Two nested loops. By author based on Brandenburg and Popp

Header	CRC	Side Information	Main Data	Ancillary Data

Figure 9 - MP3 frame after algorithm. By author based on Brandenburg and Popp

FLAC

Process

Blocking

To start the compression process, blocking divides the audio signal into blocks of a specific size. A block can range between 16 and 65,535 samples, yet audio signals are not always split in constant block sizes (Muin 6). For example, a block may be so large that the encoder will be unable to find a suitable prediction in later stages. On the other hand, a block may be too small which would waste bits on frame headers (Van Beurden and Weaver 8). Thus, the algorithm needs to determine the block divisions that will result in an adequate file size. Then, block data is passed to the Inter-Channel Decorrelation stage one sunblock at a time.

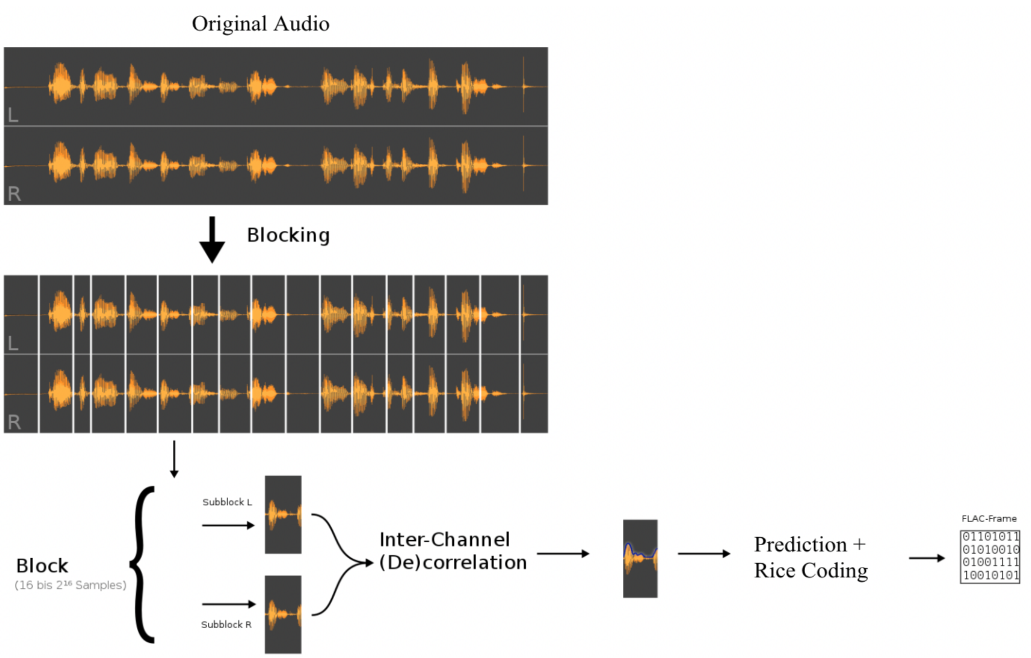


Figure 10 - Illustration of FLAC process. By author based on Kleinesfilmröllchen

Inter-channel Decorrelation

An audio file may include multiple different or identical channels. Stereo files have 2 channels, whereas mono files have 1 channel of sound. For stereo files, the inter-channel decorrelation block can average or take the difference of the samples of both sub-blocks (channel 1 and channel 2) to create a new channel; thus this process removes redundancy between channels at compression for multi-channel files (Van Beurden and Weaver 8).

Prediction

Once sub-blocks have gone through the Inter-Channel Decorrelation, the compression process starts with the prediction block. A predictor consists of a simple mathematical description that predicts a sample based on previous samples. This prediction is rarely exact and the error of this prediction is passed to be coded in the entropy coding block (Van Beurden and Weaver 7). In this way, the FLAC compression models each block based on its error to the real signal, as compressing the error usually takes less memory than to store the original signal. This error is stored in entropy coding and decompressed at playback, as the decoder knows the methods of prediction.

Entropy Coding

The error coding stage is where the difference between the prediction and the original signal is encoded. Once the difference is computed, it uses Rice coding to compress the relatively small error number from the prediction stage. Rice coding divides the numerical value by a Rice parameter, creating a quotient and a remainder; with a series of operations based on these two values, it computes a small binary word that indexes the original value. If the residual value is close to 0 and the Rice parameter is successfully chosen, this method needs less bits to store the binary code than the original value (Van Beurden and Weaver 10).

Investigation

The research question will be investigated by conducting two experiments: one with a quantitative analysis and the other with a qualitative analysis.

Methodologies

At first, I hoped to use the bird recordings that I took in the forest at early light; yet, a quick look at the spectrogram of my recording deemed my recording unusable for bioacoustic analysis. *Figure 11* demonstrates the high levels of noise present, due to poor recording equipment, in my recording in contrast to a professional recording. Thus, I obtained 3 professional bioacoustic recordings of varying length and qualities: 1 from Raven pro and 2 from the "Xeno-Canto" bird sound database as seen in *Table 1* (RavenPro; "Xeno-Canto").

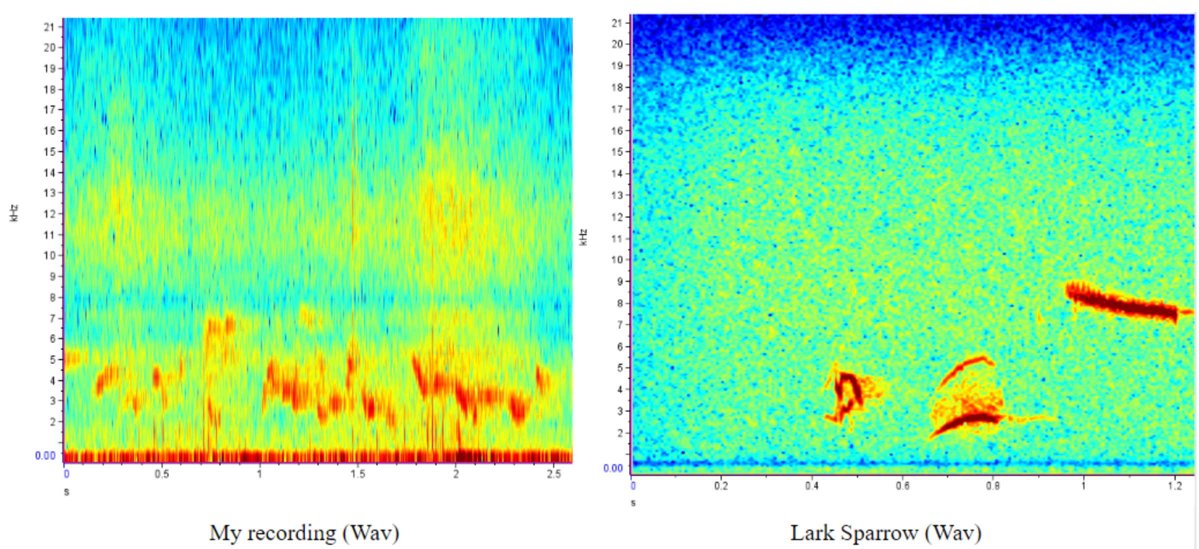


Figure 11 - Spectrogram comparison where time is displayed on the x-axis and frequency on the y-axis, while loudness corresponds to the intensity of the color. By author

Trial	Bird Type	Length	Annotation	Quality (A-E)
1	Lark Sparrow (“LarkSparrow”)	5.6 seconds	From the Raven Pro software as example recording.	A
2	Red Crossbill ("XC748417")	33.6 seconds	From the Xeno-Canto database, a public website to publish bioacoustic recording.	C
3	Boreal Owl (“XC632817”)	3 hours	Longest recording on the Xeno-Canto database.	A

Table 1 - List of recordings

Original audio files were recorded in either mono or stereo with bit depths of 16 or 24 bits and a sampling rate of 44.1. Then, prior to compression, I converted all files to 44.1 kHz, 16 bits, mono WAV files using the WAV format in the software Audacity 3.1.3 (Audacity).

To compress the audio files, I practically implemented MP3 and FLAC algorithms by using Windows’s command shell, rather than using an encoding application. For MP3 compression, I downloaded the LAME command-line encoder (considered as the best MP3 encoder) and compressed each recording in 5 different bit-rates (96-128-160-256-320 kbps) using a command seen in *Appendix A* (Delgado). Similarly, I downloaded the FLAC algorithm from the “Xiph Foundation”, the developers of FLAC, and compressed each trial once because of FLAC’s lossless nature with a command line (*Appendix B*) (Xiph Foundation). Both download encoders were around 1.1MB, highlighting the complexity of the algorithms.

Analysis

For Experiment #1, I analyzed the WAV, MP3, and FLAC recordings with Raven Pro. Raven Pro is a software, developed by the Cornell Lab of Ornithology, that was specifically developed for the bioacoustic analysis of bird sounds. I obtained a student license to access the advanced functionalities. One of them was the calculation of SNR. *Figure 12* illustrates the method to measure the SNR of a signal; I selected a signal to obtain its SNR. Across all trials, I obtained the file size and SNR of the audio files, and with this data, I could calculate the compression ratio and compare the amount of noise in WAV and compressed files.

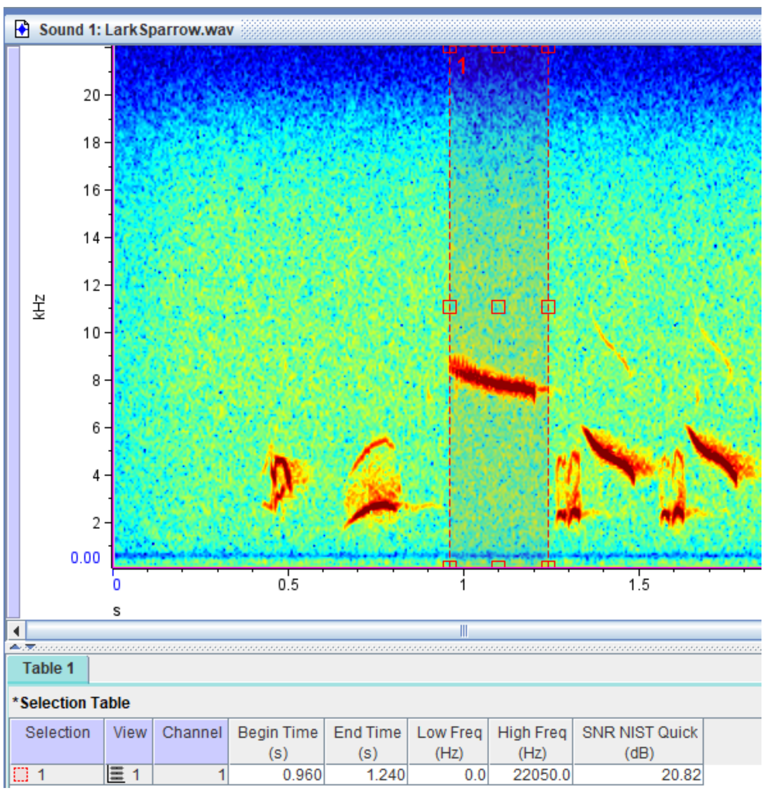


Figure 12 - RavenPro analysis. By author

In Experiment #2, I subtracted a WAV file by a MP3 or FLAC file to visually illustrate the acoustic information lost at compression in spectrograms. Using Mike Russel's method in Adobe Audition, I perfectly aligned the WAV and compressed tracks and then applied an inversion effect to the compressed track; finally, I mixed the two tracks in a mono recording ("Phase Invert"). In this way, the new mono recording portrayed the difference of information between the two files. With a description of the file's spectrogram, I extended the results of Experiment 1, analyzing the qualitative value of lost signals and noise introduced, as a result of compression (noise refers to the signals that are not bird signals).

This methodology enabled me to analyze the quality of a compressed file by describing the deviations in bird signals and by measuring the amount of noise present after compression. A high SNR and clean bird signals are determining factors in finding out whether a recording is usable for Bioacoustic investigations, so the 2 experiments work together to answer the investigation. For the MP3, the experiments should theoretically demonstrate a correlation between bit-rate, quality of bird signals, and noise occurrence, and for FLAC, the two latter factors should remain unchanged after compression.

Experiment #1

MP3

Trial 1 - LarkSparrow - 0.5 seconds				
	Bit-rate (kbps)	Size (kb)	Compression ratio	SNR (dB)
WAV	WAV	434		20.8
1	96	60	7.2:1	13.05
2	128	80	5.4:1	13.75
3	160	100	3.4:1	15.31
4	256	160	2.7:1	15.87
5	320	200	2.1:1	15.89
Trial 2 - Red Crossbill - 33.6 seconds				
	Bit-rate (kbps)	Size (kb)	Compression ratio	SNR (dB)
WAV	WAV	3070		10.56
1	96	395	7.8;1	7.03
2	128	526	5.8:1	7.79
3	160	658	4.6:1	8.21
4	256	1010	3:1	8.65
5	320	1280	2.4:1	8.72
Trial 3 - Owl - 3 hours				
	Bit-rate (kbps)	Size (kb)	Compression ratio	SNR (dB)
WAV	WAV	987000		12.74
1	96	123000	8:1	11.52
2	128	164000	6:1	11.67
3	160	205000	4.8:1	11.83
4	256	329000	3:1	11.98
5	320	411000	2.4:1	12.12

Table 2 - Results of Experiment #1 (MP3) * Compression ratio rounded to nearest tenth

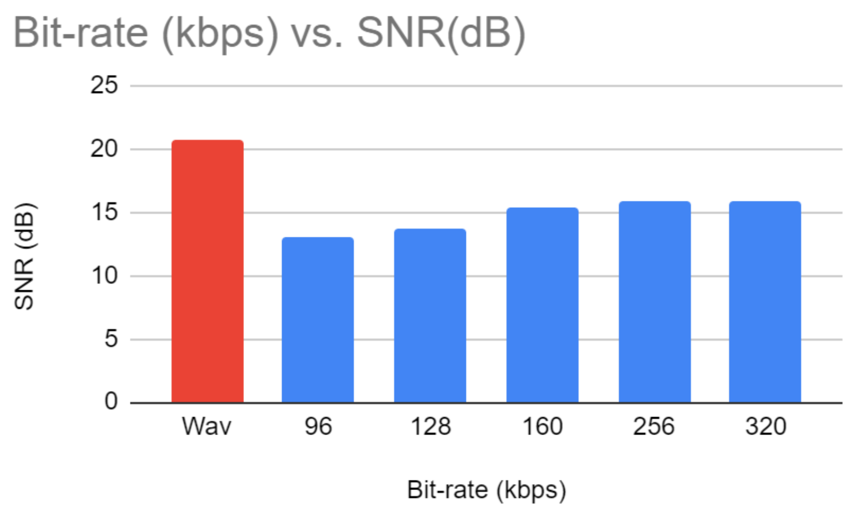


Diagram 1 - LarkSparrow. By author

Table 2 presents the MP3 experiment across 3 trials, each compressed at 5 different bit-rates. Furthermore, *Diagram 1* emphasizes the correlation between the SNR and the bit-rate for the LarkSparrow, a trend seen throughout all of the trials.

FLAC

Trial 1 - LarkSparrow - 0.5 seconds			
	Size (kb)	Compression ratio	SNR (dB)
WAV	434		20.8
FLAC	271	1.6:1	19.9
Trial 2 - Red Crossbill - 33.6 seconds			
	Size (kb)	Compression ratio	SNR (dB)
WAV	3070		10.56
FLAC	1560	2:1	9.76
Trial 3 - Owl - 3 hours			
	Size (kb)	Compression ratio	SNR (dB)
WAV	987000		12.74
FLAC	515000	1.9:1	12.74

Table 3 - Results of Experiment #1 (FLAC) *Compression ratio rounded to nearest tenth

Across 3 trials, *Table 3* demonstrates that FLAC's SNR is similar to WAV's SNR, making FLAC a valid codec for bioacoustics.

Experiment #2

MP3

As seen in *Appendix B (Figure B1, B2, and B3)*, the difference between WAV and MP3 files decreases as the bit-rate increases, as at higher bit-rates, more bits are available for compression. Across all trials, the amount of noise introduced after compression differs; at 96 and 128kbps, the MP3 encoder produces a considerable amount of noise in comparison to the WAV file. However, this quantity does not mean that the noise is audible; in fact, the spectrogram, with the purple shade of color, demonstrates that introduced noise contains low levels of energy (in dB) which does not interfere with louder bird signals.

Other than the noise introduced, compression (at bit-rates 96, 128, and 160) seems to take away important bird signals from the WAV file because the bird frequency patterns are recognizable in the difference file. Although this seems to be true, the frequency patterns hold small amounts of amplitude. In truth, the MP3 eliminated these signals because they are practically inaudible, hence not important in spectral bioacoustic analysis either. The lack of energy in the frequencies is seen in the faint intensity of color in the spectrogram. To make it more obvious, *Figure 13* shows the similarity between the WAV and MP3 signal, and emphasizes that the difference between files visually and audibly changes little. This is because the MP3's perceptual coding gets rid of signals that are inaudible for the human ear.

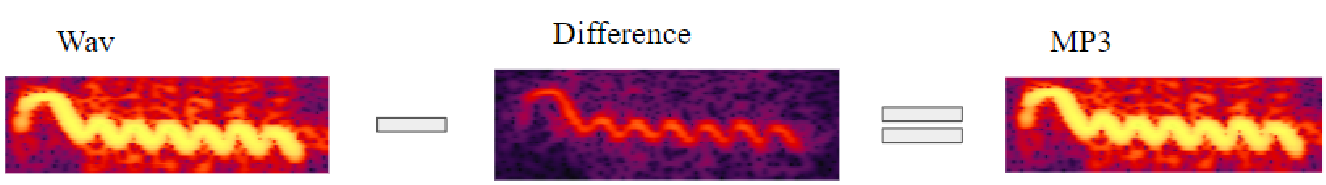


Figure 13 - WAV and MP3 (96 kbps) closer look at LarkSparrow. By author

Finally, *Appendix B* clearly shows that above 256 kbps, MP3 compression scarcely changes the original file because the spectrogram's faint colors illustrate the uniformity of signals. Without a doubt, MP3 compression merely deviates bird signals because bird frequencies (1kHz - 10kHz) reside perfectly in the human hearing bandwidth, which is left rather untouched by the MP3. Also, in general, MP3 compression merely creates noise, proving its applicability in the bioacoustic field; yet, levels of bit-rates should be adapted in accordance to desired audio quality.

FLAC

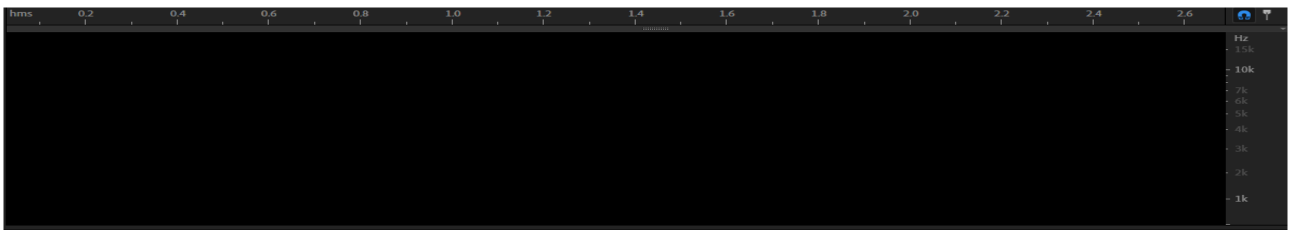


Figure 14 - Difference between WAV and Flac file - LarkSparrow. By author

Figure 15 - Difference between WAV and Flac file - Red Crossbill. By author

Figure 16 - Difference between WAV and Flac file - Owl. By author

As seen in *Figures, 14, 15, and 16*, there is no difference between the WAV and FLAC files, as FLAC is a lossless audio codec. It reproduces the uncompressed format, thus ensuring its usability in a bioacoustic analysis.

Discussion

This section evaluates to which extent MP3 and FLAC compression algorithms are usable in bioacoustic analysis. Overall, the MP3 proved to increase the amount of noise in the compression, although the amount of noise created depended on the bit-rate. At 96kbps, the SNR reached the lowest drop (-7.75dB), yet the reduction of the SNR is not dramatic, taking into account that increasing the bit-depth by 1 bit increases the dynamic range by 6dB (Harris). Above 128kbps, the algorithm stayed in the range of 6dB in the reduction of the SNR, a positive outcome for a lossy compression. Furthermore, Experiment 2 demonstrates that the noise introduced has no real qualitative effect on the recording; at 96 and 128 kbps, substantial noise is introduced, yet the spectrogram illustrates the noise's lack of energy, keeping a relatively high SNR. Above 128kbps, little to no noise is created. Moreover, Experiment 2 suggests that MP3 compression results in a loss of information in bird signals, but a closer comparison between spectrograms proved that deleted information holds hardly any energy (even at 96kbps). Thus, for the MP3, two conclusions can be drawn. First, MP3 compression (at least at 128kbps) does not generate a major deviation in the amount of noise in a recording, an important conclusion to determine whether spectrographic analysis is doable. Second, the MP3 does not degrade important frequencies to the point of being unrecognizable in a bioacoustic analysis.

Since the FLAC algorithm is lossless, the results of the experiments make sense. For Experiment 1, FLAC always compressed around half of the file size and kept a very similar SNR to the original file. This was proven by Experiment 2, where the spectrograms, showing the difference between the WAV and FLAC files, were completely black. Thus, no noise was introduced and the frequencies remained untouched, a positive outcome for the bioacoustic field.

Limitations to the investigation

The quantitative analysis was only based on the SNR parameter in the file. Although this is an important factor to answer the question, the quality of a signal is not excluded by the amount of noise. Measuring other factors such as spectral entropy, which sums the normalized signal spectral power, could have been done with the use of R, a statistical programming language, yet this goes beyond the scope of the Extended Essay (Helakari).

Conclusion

Through the investigation, I concluded that MP3 and FLAC compression algorithms are usable to a notable extent in the field of bioacoustics and bird sound analysis. In both cases, compression becomes an advantage, as it, at least, halves the file size of the original recording, a determining factor in the manipulation and communication of sound files. While FLAC compression should be used because of its lossless nature, MP3 compression can be safely used with a 128kbps bit-rate or higher. As seen in this investigation, the ability to accurately analyze bird signals after compression resides in the ability to recognize frequency patterns in the original file; consequently, the quality of the original recording determines the usability of recordings with both codecs. Thus, biological research should continue to flourish, without having to face the challenge of data storage.

Works Cited

Araya-Salas, Marcelo, et al. *Assessing the effect of sound file compression and background noise on measures of acoustic signal structure*. www.tandfonline.com/doi/abs/10.1080/09524622.2017.1396498?journalCode=tbio20.

Audacity. *Audacity*. Version 3.1.3, Audacity Team. Software.

Bartlett, Bruce, and Jenny Bartlett. *Practical Recording Techniques: The Step-by-Step Approach to Professional Audio Recording*. 5th ed., Boston, Focal Press, 2009.

Brandenburg, K., and H. Popp. *An Introduction to MPEG Layer 3*. www.mp3-tech.org/programmer/docs/trev_283-popp.pdf.

Cornell Lab of Ornithology. *RavenPro*. Version 1.6.3, Ithaca, NY: The Cornell Lab of Ornithology, K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology, 2022. Software.

Delgado, Carlos. "How to convert WAV files to MP3 with the command line using LAME like a boss in Windows 10." *OurCodeWorld*, ourcodeworld.com/articles/read/1402/how-to-convert-wav-files-to-mp3-with-the-command-line-using-lame-like-a-boss-in-windows-10. Accessed 25 Oct. 2022.

Elsea, Peter. *Basics of Digital Recording*. artsites.ucsc.edu/EMS/music/tech_background/TE-16/teces_16.html. Accessed 1996.

Farina, Almo, and Stuart H. Gage. *Ecoacoustics: The Ecological Role of Sounds*. Hoboken, John Wiley & Son, 2017.

Fraunhofer. "Mp3." *Fraunhofer IIS*, 2017, www.iis.fraunhofer.de/en/ff/amm/consumer-electronics/mp3.html.

Gans, Dave. "Digital Vs. Analog Audio: An Overview." *Klipsch*, www.klipsch.com/blog/digital-vs-analog-audio#:~:text=Like%20images%2C%20audio%20signals%20can,without%20losing%20its%20original%20quality. Accessed 30 Sept. 2022.

Hacker, Scot. *MP3: The Definitive Guide*. 2000.

Hahn, Michael. "Signal-to-noise Ratio: How to Avoid Excess Noise in Your Tracks." *Landr*, blog.landr.com/signal-to-noise-ratio/#:~:text=Signal%2Dto%2Dnoise%20ratio%20is,of%20noise%20in%20the%20system. Accessed 30 Sept. 2022.

Harris, Mark. "What Is Bit Depth?" *Lifewire*, www.lifewire.com/what-is-bit-depth-2438536#:~:text=Bit%20Depth%20and%20Dynamic%20Range&text=Bit%20depth%20also%20determines%20how,96%20dB%20of%20dynamic%20range. Accessed 30 Sept. 2022.

Helakari, H. "Spectral entropy indicates electrophysiological and hemodynamic changes in drug-resistant epilepsy – A multimodal MREG study." *ScienceDirect*, www.sciencedirect.com/science/article/pii/S2213158219301135#:~:text=Spectral%20entropy%20(SE)%20is%20a%20measure%20of%20signal%20irregularity%2C,Faust%20and%20Bairy%2C%202012).

Jared, H. "How Does MP3 Compression Work?" *Ledgernote*, ledgernote.com/blog/q-and-a/how-does-mp3-compression-work/. Accessed 21 Aug. 2021.

Kavitha, P. *A Survey on Lossless and Lossy Data Compression Methods*. Accessed 3 Mar. 2016.

Kleinesfilmröllchen. "File:FLAC-Audiokompression.svg." *Wikimedia Commons*, commons.wikimedia.org/wiki/File:FLAC-Audiokompression.svg. Accessed 25 Oct. 2022.

"LarkSparrow." By Ravenpro. *LarkSparrow*. RavenPro.

Maher, Robert C. *Lossless Compression of Audio Data*. www.montana.edu/rmaher/publications/maher_lossless_chapter_2003.pdf. Accessed 2003.

Maklin, Cory. "Fast Fourier Transform." *Medium*, 27 Dec. 2019, towardsdatascience.com/fast-fourier-transform-937926e591cb#:~:text=In%20layman's%20terms%2C%20the%20Fourier,consisting%20of%20multiple%20pure%20frequencies.

Muin, Fathiah Abdul. *A Review of Lossless Audio Compression Standards and Algorithms*. aip.scitation.org/doi/pdf/10.1063/1.5002024. Accessed 14 Sept. 2017.

Nugent, James. "Lossless Vs. Lossy Audio: FLAC, WAV, MP3, and Other Formats." *HigherHZ*, 22 Oct. 2021, higherhz.com/lossless-vs-lossy-compression-audio-formats/.

"Phase Invert for Subtracting Audio Tracks." *Youtube*, uploaded by Mike Russel, 2 Sept. 2019, www.youtube.com/watch?v=vNvJKBg3yds.

Pohlmann, Ken C. *Principles of Digital Audio (6th Edition)*. 6th ed., New York City, McGraw-Hill Professional Publishing, 2010.

"Quantization - Digital Audio Fundamentals." *Youtube*, 9 Sept. 2019, youtu.be/1KBLguIXL30.

Raissi, Rassol. *The Theory behind Mp3*. Dec. 2002, www.mp3-tech.org/programmer/docs/mp3_theory.pdf.

"Signal to Noise Ratio (SNR) – How It Can Destroy Your Recording." *Magroove*, 2 Nov. 2019, magroove.com/blog/en-us/snr/.

Sripada, Praveen. *MP3 DECODER in Theory and Practice*. www.diva-portal.org/smash/get/diva2:830195/FULLTEXT01.pdf. Accessed 2006.

"Understanding Sample Rate, Bit Depth, and Bit Rate." *HeadPhonesty*, www.headphonesty.com/2019/07/sample-rate-bit-depth-bit-rate/. Accessed 2 Sept. 2021.

Van Beurden, M.Q.C., and Andrew Weaver. *Free Lossless Audio Codec*. datatracker.ietf.org/doc/pdf/draft-ietf-cellar-flac-04. Accessed 21 Aug. 2022.

Welz, Adam. "Listening to Nature: The Emerging Field of Bioacoustics." *YaleEnvironment360*, 5 Nov. 2019, e360.yale.edu/features/listening-to-nature-the-emerging-field-of-bioacoustics.

Wilburn, Thomas. "The AudioFile: Understanding MP3 compression." *ARStechnica*, 3 octo 2007, arstechnica.com/features/2007/10/the-audiofile-understanding-mp3-compression/. Accessed 30 Sept. 2022.

"XC632817." *Xeno-Canto*, uploaded by Stanislas Wroza, xeno-canto.org/632817.

"XC748417." *Xeno-Canto*, uploaded by Varkki Hannu, xeno-canto.org/748417.

"Xeno-Canto." *Xeno-Canto*, xeno-canto.org/. Accessed 30 Sept. 2022.

Xiph Foundation. "Downloads." *FLAC*, edited by Xiph Foundation, xiph.org/flac/download.html. Accessed 25 Oct. 2022.

Yes Pan, David. *Digital Audio Compression*. *Columbia Computer Science*, www.cs.columbia.edu/~hgs/research/projects/echo-detection/Pan93-acomp.pdf.

Appendix A

Compression through the command shell

Figure A1 - MP3 compression of RedCrossbill through the command shell

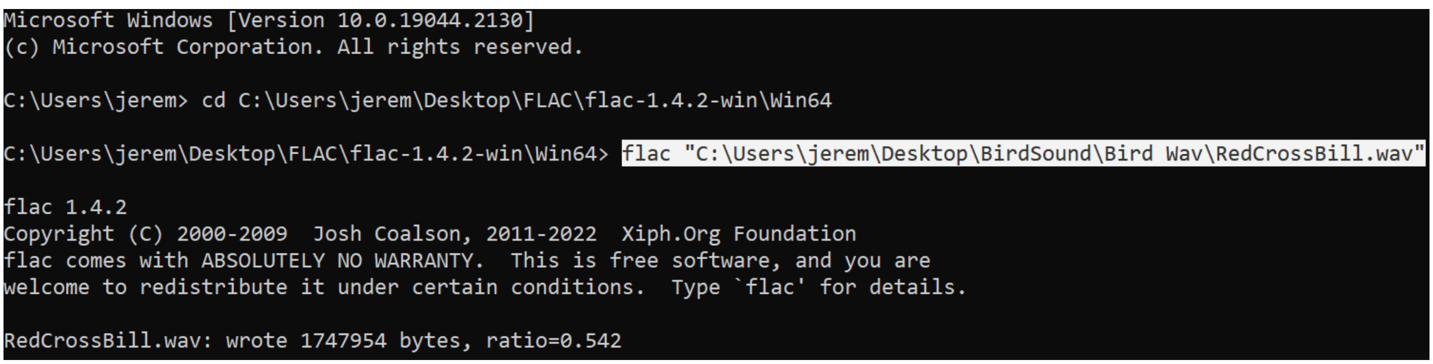


Figure A2 - FLAC compression of RedCrossbill through the command shell

Appendix B

MP3 Results of Experiment #2

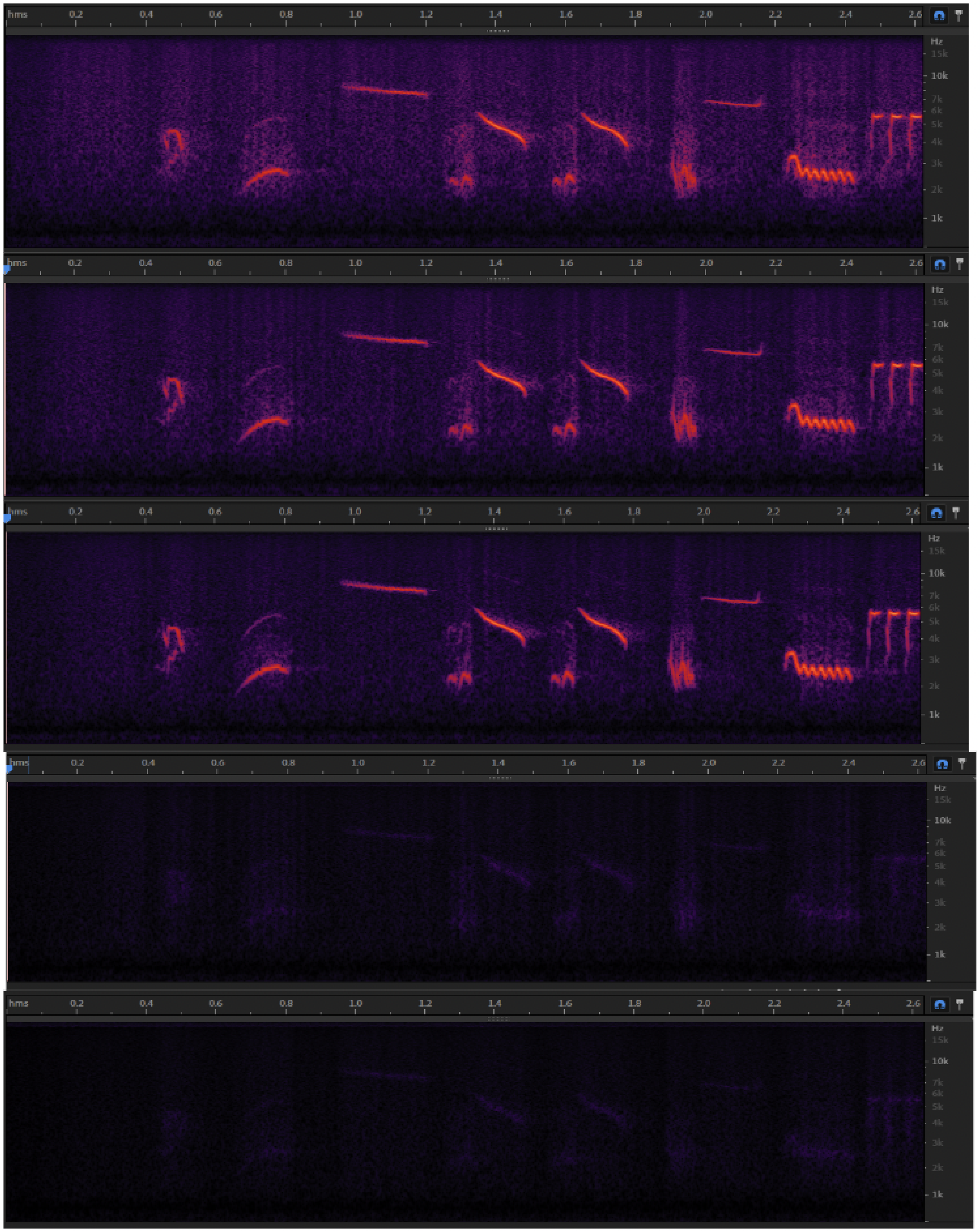


Figure B1 - Difference between Wav and MP3 files - LarkSparrow. By author

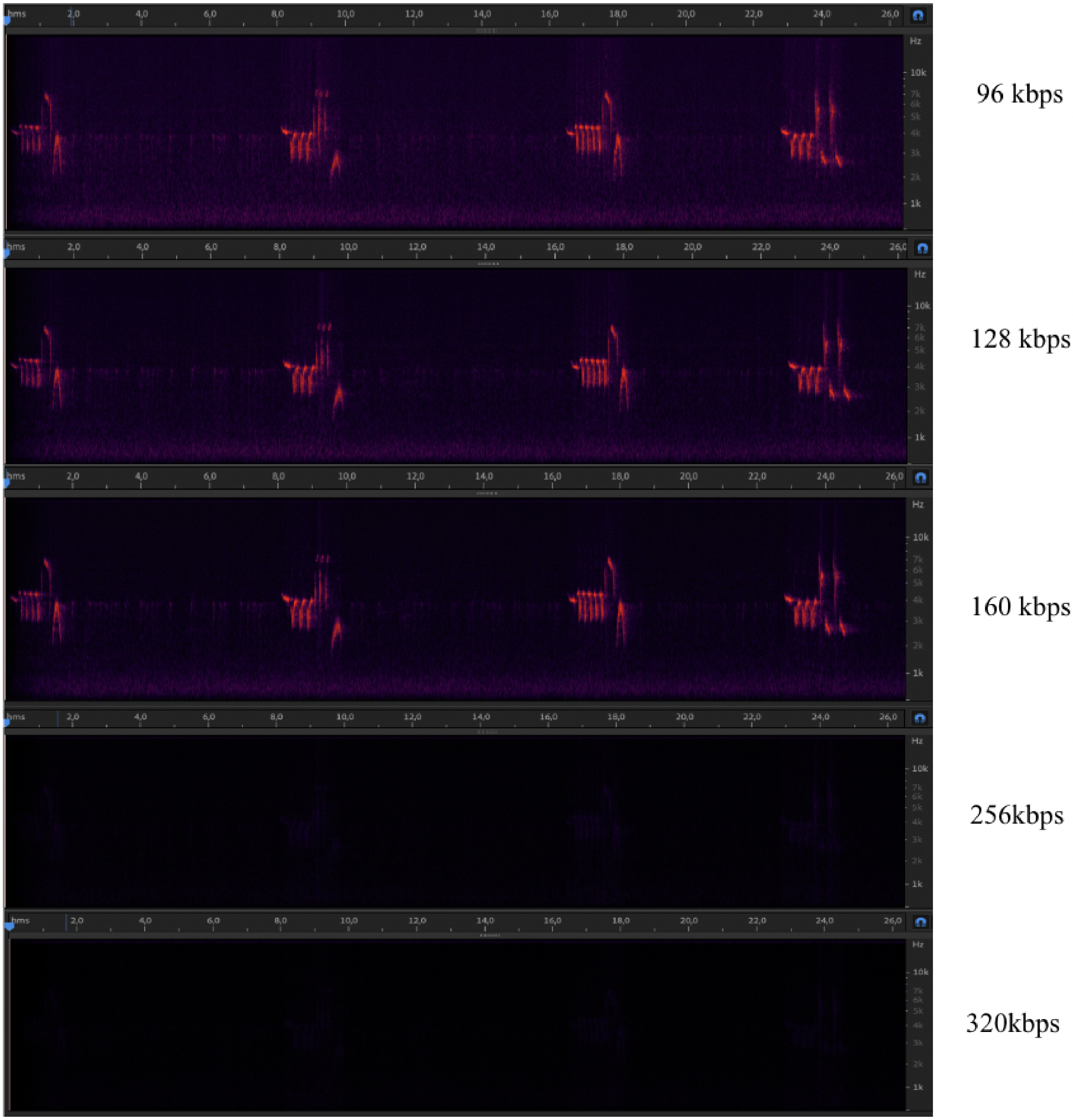
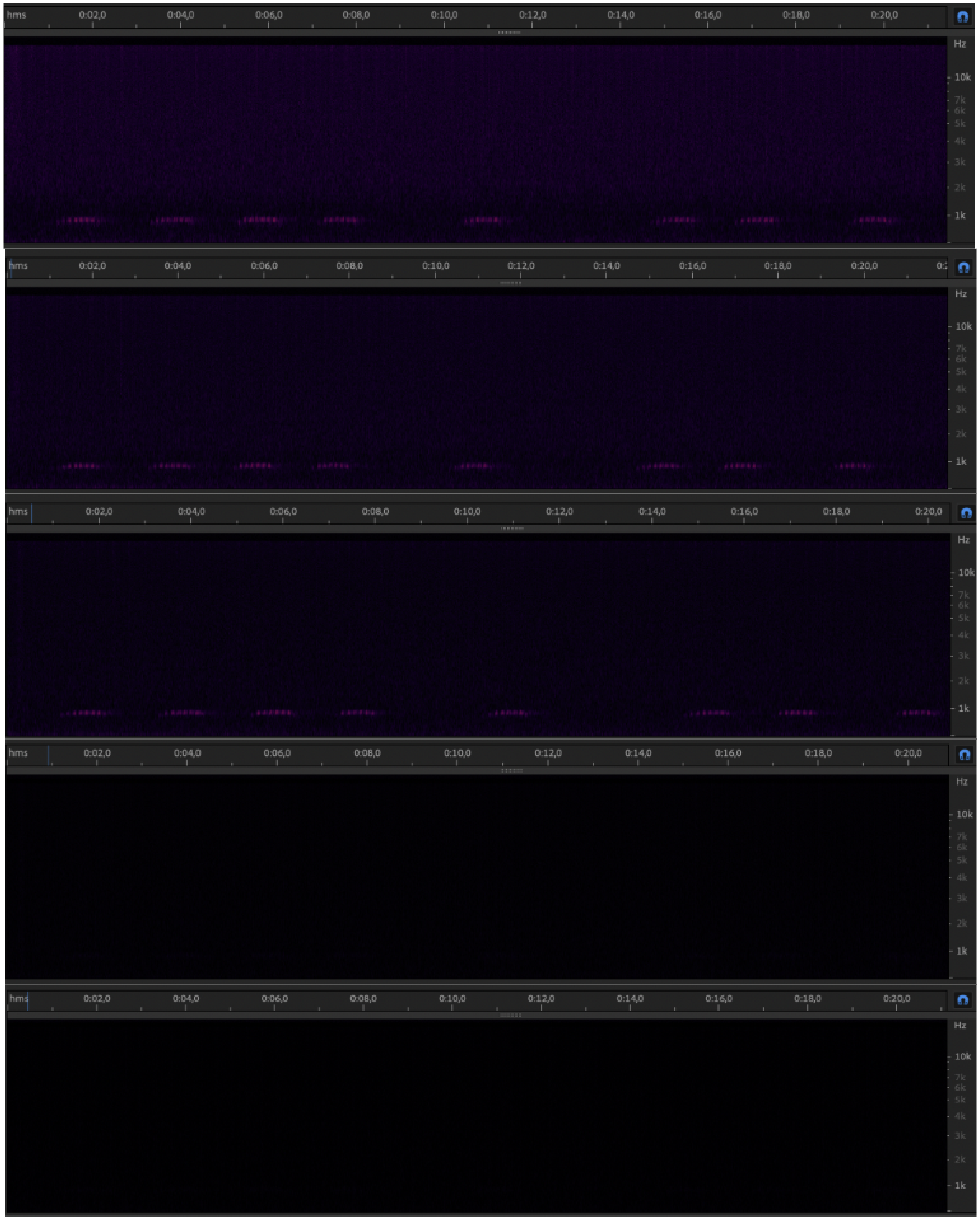


Figure B2 - Difference between Wav and MP3 files - Red Crossbill. By author



96 kbps

128 kbps

160 kbps

256kbps

320kbps

Figure B3 - Difference between Wav and MP3 files - Boreal Owl. By author