Josh Schaffer
Mr. Francois
AP Physics I
June 20 2018

A New Frequency-Based Compression Algorithm for Both Lossless and Lossy Audio Encoding

The advancement of computational power within recent years has become exponential. Moore's law claims that utilizing current technologies, the number of transistors within a dense integrated circuit per unit surface area can be expected to double approximately every 2 years. While the law of diminishing returns suggests that innovation is approaching some asymptote, the growing ability for this newfound power to be utilized is undeniable. This power has allowed new technologies to revolutionize entire domains of digital media, such as video, where the h.264 encoding process allows for extremely efficient compression levels. Implementations of the codec are so widespread that there exist kernel patches for low level processor optimization, proving the ability for new algorithms to permeate an entire market.

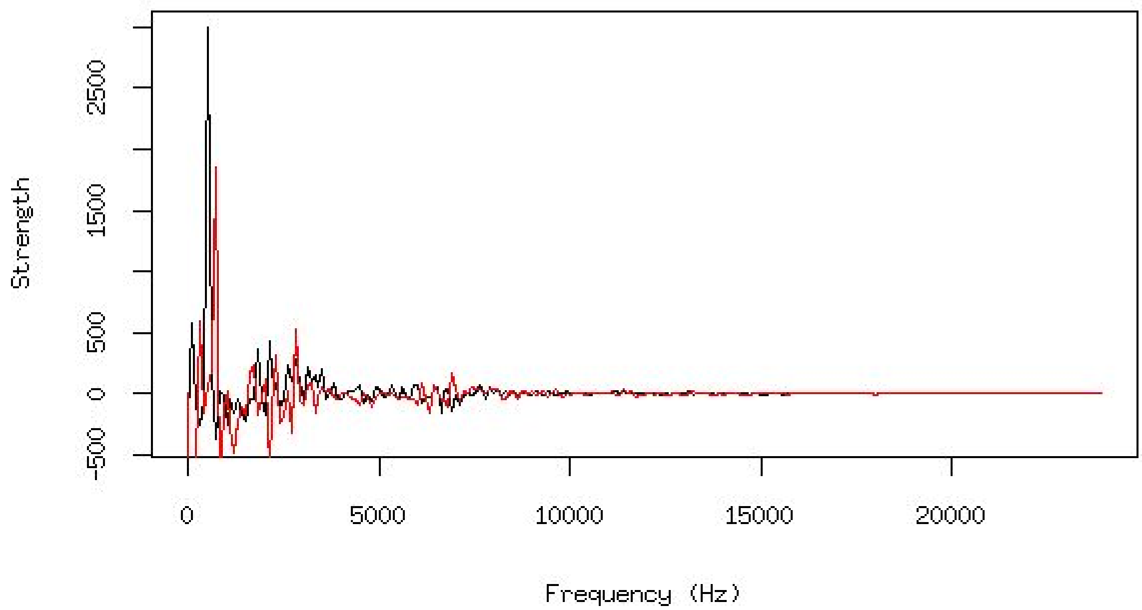
Yet by comparison, the majority of audio, compressed or not, is still stored and transmitted in bitwise encodings (a long string of numbers) within the "time domain". These numbers represent the pressure difference of the sum of audio sources in the recording. These sources interfere constructively and destructively, and an audio file is transcoded by storing the relative pressure difference of this summatted wave in points along said wave over time, taken at a rate known as the acquisition frequency or sampling rate. Harry Nyquist found that a traditional audio file of this form can accurately represent a audio frequency equal to one half of the acquisition frequency, known as the Nyquist frequency. His experimentation showed the limits of human hearing to be around 22,050 Hz, incurring an acquisition frequency of 44,100 Hz. Being a non-round number for computational purposes, the modern day Compact Disc standard has adopted a defacto acquisition frequency of 48000 Hz, incurring a Nyquist frequency of 24,000 Hz. The advantage of these traditional bitwise encodings is their simplicity and computational ease of playback, as the encoded amplitude or pressure difference refers directly to the physical displacement of a speaker.

By representing the same audio data using a non-bitwise encoding in a domain other than the time domain, a new compression algorithm could allow for highly compressed file sizes which maintain a lossless or close to lossless fidelity, at the cost of marginally increased computational power during the time of audio playback. To compare back to the video standard mentioned previously, this is exactly how h.264 encoding works: Instead of storing the exact color and brightness of every pixel during every frame, it relies on an abstract approximation of data which the computer must process at the time of playback.

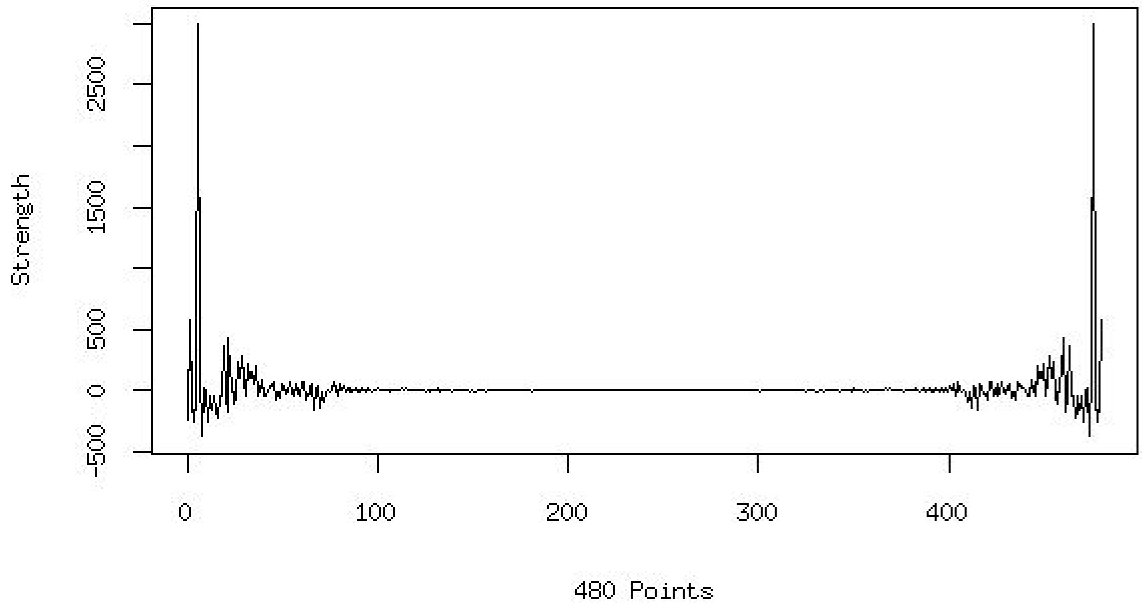
It is universally accepted that any continuous function can be perfectly represented by the infinite sum of sinusoidal graphs of varying frequencies. The Fourier Transformation decomposes a continuous signal into the powers of various frequencies as they contribute to the traditional summated wave signal of the “time domain” by means of constructive and destructive interference. This new “frequency domain” generated by the Fourier Transformation is the basis of the abstract encoding of this new algorithm. This domain is defined by a vertical axis which denotes the power of sinusoids of the frequencies across a horizontal axis.

The actual transformation can best be understood graphically. Assume a time domain signal f(x), the net interference of a continuous audio representation, is wrapped around the center of a cartesian plane by means of rotating a vector about the origin, such that at any given time, the length of the vector is equal to the pressure difference of the signal at that time. Hence, high points of the signal refer to distances farther from the origin, and low points of the signal refer to distances closer to the origin. It is important to note that both of these representations, the original signal and the polar rotation, are time dependent: The first by virtue of its presence in the time domain, and the later by incurring a “winding frequency”. This winding frequency is measured as the portion of one full rotation about the cartesian origin that is achieved by the rotating vector within one second. If this frequency were to be continuously varied, plotting the x-coordinate of the center of mass of the corresponding rotational graph against it, the resulting plot would be the power spectrum of frequencies making up the input signal f(x). The Fourier Transformation yields two of these spectrums, one real and one imaginary, corresponding to the x-coordinate and y-coordinate in complex terms of the center of mass of the rotational graph respectively. The real components refer to even functions of a given frequency, while the imaginary components refer to odd functions of those frequencies (specifically cosine and sin respectively).

The power of a given frequency “s” from the Fourier Transformation “F” can be denoted by the equation F(s) = $\int_{-\infty}^{\infty} f(x)e^{-2\pi xs}dx$, where f(x) is a summated signal in the time domain. For a discrete set of N points, however, as is the case for bitwise-encoded raw audio, the transformation becomes a finite sum equal to $\sum_{n=1}^{N} f(x)e^{-2\pi xn}$. Dividing by the duration of the signal becomes necessary since the nature of the transformation returns larger sums for higher frequencies, yielding $\frac{1}{N}\sum_{n=1}^{N} f(x)e^{-2\pi xn}$. The following is an example of the real and imaginary components, in black and red respectively, of the frequencies returned by a Fourier Transformation of a 1/100 second sound clip encoding at 48,000 Hz.

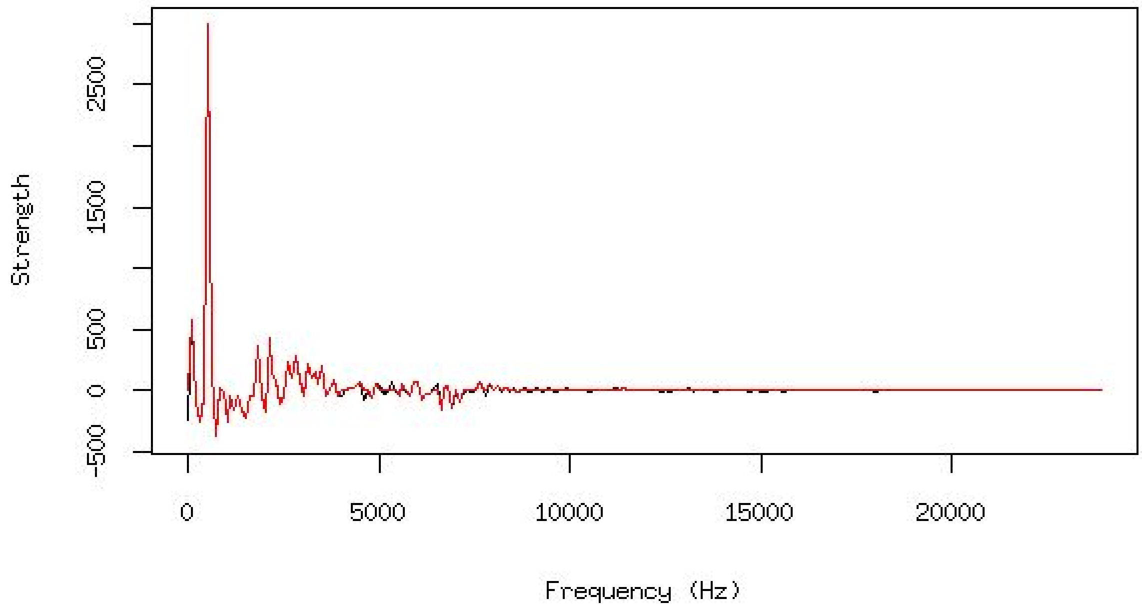


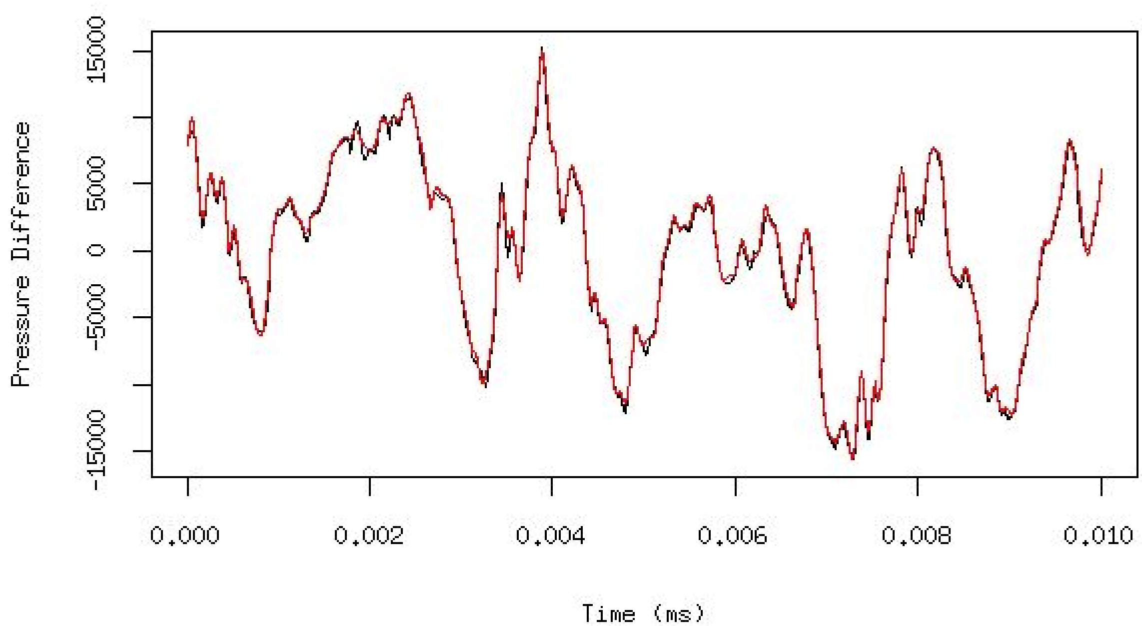
Note that the frequencies span the range of human hearing. Their strength is an arbitrary axis since the amplitude of a final sound wave is largely dependent upon the physical equipment it is played through, such as a digital to analog converter and an amplifier. The above graph is only the first half of the transformation. The real component of the entire graph can be seen below.



From the aforementioned discrete equation, a signal represented by a set of N points can be transformed to the powers of N frequencies, so a 1/100 of a second recording at 48,000 Hz is represented by 480 points in the time domain and 480 points in the frequency domain, as shown above. Half of those points, however, are what are known as the complex conjugate of the other 240. This is visible above by the mirrored nature of the graph. This means that for compression purposes, half of the generated points from a Fourier Transformation can be discarded at the time of compression and recalculated at the time of playback, garnering a lossless file size one half of the original. This is better than current lossless audio compression algorithms such as FLAC, which struggle to achieve 50% compression.

By further reducing the number of frequencies that are stored, a new algorithm could also achieve very high fidelity lossless compression. Those 240 complex points in the aforementioned sound clip could be lessened by introducing certain limiting criteria for “valuable” frequencies worth storing. For instance, by discarding all frequencies with either a real or imaginary power less than one third the average of the absolute value of all the real or imaginary powers, those 240 frequencies are further reduced down to just the powers of 63 relevant frequencies. Since these frequencies are not consecutive, storing their index would be required, incurring a total of 126 values to encode. This is would yield a compressed file size equal to just 26.25% of the original, with an extremely high fidelity as seen in both the frequency and time domains below, with the compressed data in red and the raw data in black:


Strength
-500
0
500
1500
2500
0
5000
10000
15000
20000
Frequency (Hz)


Pressure Difference
-15000
-5000
0
5000
15000
0.000
0.002
0.004
0.006
0.008
0.010
Time (ms)

One advantage of this compression method is its frequency independence. While competing algorithms such as MP3 may achieve higher compression rates, they rely on a process known as frequency limiting. While human hearing can go beyond 20 kHz, the hearing of many adults can top out even as low as 10 - 15 kHz. MP3 compression takes advantage of this by chopping off those higher frequencies altogether, and while they might not be heard by some directly, their removal can contribute to a noticeable degradation in the perceived depth, clarity, or warmth of an audio recording. By comparison, this new method judges frequencies not by their pitch but by their quantitative contribution to the final waveform as revealed by the Fourier Transformation. This means it will equally compress high or low frequencies if they are not important to the final sound. The effect of this is observable in the comparison of compressed and uncompressed waveforms in the time domain as seen above, where even on the micro scale of 1/100 of a second, the compressed wave is able to follow even the most acute of changes in pressure difference.

This algorithm for compression is still in its infancy and requires much experimentation to perfect. Arbitrarily chosen constants such as the block size of 1/100 of a second and the limiting criteria for “important” frequencies must be tested over a broad range of audio recordings in order to find the most efficient compression algorithm. For the lossy compression method, different structures for encoding the index of stored frequencies must be tested in order to save space on the bit-level. The format has some inherent advantages such as its high fidelity, streamable composition, lossless support, and lack of frequency limiting. However, for its major adoption, a standard must be decided on so that audio playback software can implement decoders for real time use.

Works Cited

"1702 - Using FFT to Obtain Simple Spectral Analysis Plots." *Polybius at The Clickto Network*, Fox News, web.archive.org/web/20120615002031/http://www.mathworks.com/support/tech-notes/1700/1702.html.

Cutnell, John D., and Kenneth W. Johnson. *Physics*. J. Wiley & Sons, 2005.

"Fourier Transforms." *Double Radio Sources Associated with Active Galactic Nuclei*, www.cv.nrao.edu/course/astr534/FourierTransforms.html.

Moore's Law, www.mooreslaw.org/.

"Nyquist Frequency." *From Wolfram MathWorld*, mathworld.wolfram.com/NyquistFrequency.html.

"Synthetic Soul - Lossless Codec Comparison - 00.Wav." *Polybius at The Clickto Network*, Fox News, web.archive.org/web/20090202063254/http://synthetic-soul.co.uk/comparison/lossless/file.asp?FilePKID=1.