The Ghost in the MP3

Ryan Maguire
Virginia Center for Computer Music
ryanmaguire@virginia.edu

ABSTRACT

The MPEG-1 or MPEG-2 Layer III standard, more commonly referred to as MP3, has become a nearly ubiquitous digital audio file format. First published in 1993 [1], this codec implements a lossy compression algorithm based on a perceptual model of human hearing. Listening tests, primarily designed by and for western-european men, and using the music they liked, were used to refine the encoder. These tests determined which sounds were perceptually important and which could be erased or altered, ostensibly without being noticed. What are these lost sounds? Are they sounds which human ears can not hear in their original contexts due to our perceptual limitations, or are they simply encoding detritus? It is commonly accepted that MP3's create audible artifacts such as pre-echo [2], but what does the music which this codec deletes sound like? In the work presented here, techniques are considered and developed to recover these lost sounds, the ghosts in the MP3, and reformulate these sounds as art.

1. TECHNICAL BACKGROUND

The MP3 standard, designed in the early 1990's by the Moving Pictures Experts Group, has become an interesting object of critique in contemporary technology studies [3]. How a standard which subtly reduces the audible quality of sound files has remained in place, despite massively increased bandwidths and storage capacity is impressive, and highlights the foresight (and fortune) of the format's creators. Due to a complex combination of market and social factors, the majority of music listeners today continue to prefer a standard which optimizes the download times and storage capacity of their audio devices [4]. These are often portable machines such as the iPod, on which much listening occurs in noisy environments (gyms, subways, city streets) through (often cheap) ear bud headphones and inexpensive preamplifiers. The loss of fidelity from these external factors, along with the cleverness with which MP3s are coded, a socialization to the sound of MP3 files, and other factors have obviated the need for an upgrade to higher fidelity formats for most end users [5].

Regardless, the MP3 is not always the most appropriate format for a given task, and a critical evaluation of the technology and its limitations is warranted. Many listeners today listen exclusively to MP3 files, even in settings where the gains from a higher fidelity format would be clearly perceptible. This lossy compression codec has thus come to dominate unanticipated listening spaces.

Despite its heralded performance in listening tests, the MP3 compression codec does generate audible artifacts and remove perceptible sonic information. MP3 encoding relies primarily on masking curves, used to calculate frequency and temporal masking [10]. By adjusting masking thresholds, more or less information can be removed from the uncompressed audio depending on the desired target file size. At low bit rates, due to sample rate reductions and low pass filtering, frequencies from the extreme edges of the human hearing range are further attenuated.

For example, white, pink, and brown noise, when compressed to the lowest possible MP3 bit rate [6], sound very different from the original random signal.

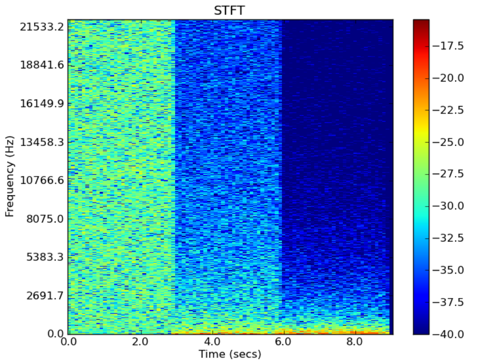

Figure 1. White, Pink, and Brown Noise - Uncompressed WAV.

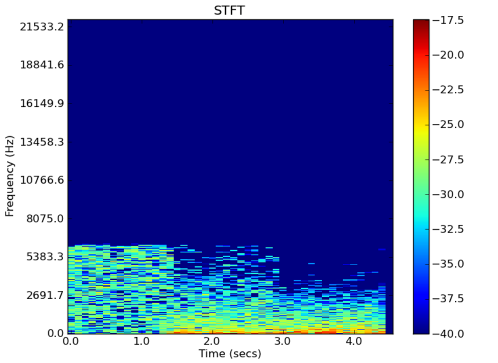

Figure 2. White, Pink, and Brown Noise - 8kbps MP3.

Copyright: © 2014 Ryan Maguire. This is an open-access article distributed under the terms of the Creative Commons Attribution License 3.0 Unported, *which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*

In comparison, low-frequency sine tones sound quite good as an MP3 encoded at 320kbps MP3. Still, some material has been left behind– acoustic information which is mostly unheard in it's original context.

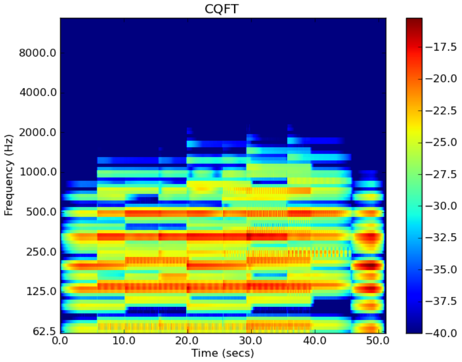

Figure 3. Sine Tone Chords – Uncompressed WAV.

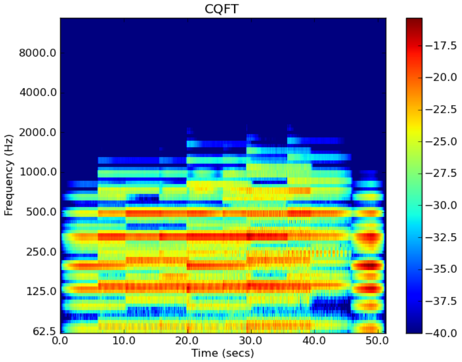

Figure 4. Sine Tone Chords – 320kbps MP3.

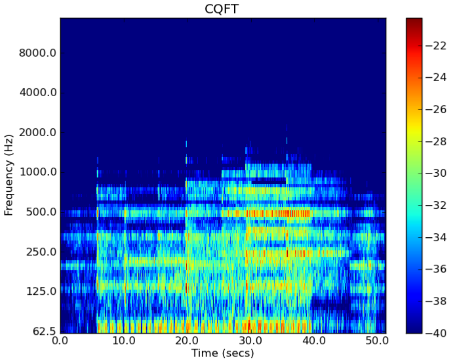

Figure 5. Sine Tone Chords – 320kbps MP3 “Ghost”.

2. GHOST HUNTING

In seeking to capture the sounds lost to MP3 compression there are multiple possible approaches that seem theoretically feasible. From a simple technique such as phase inversion to something as complex as developing a new codec– the inverse mp3, if you will. The approach I will detail here falls somewhere between these two.

I have opted to work in the frequency domain as opposed to the time domain in the experiments outlined below. Attempts at achieving a similar result with the time domain signal have been unsuccessful. I speculate that this is due to quantization and phase estimation differences between the MP3 and the original uncompressed audio, leading a time-domain phase inversion algorithm to be insufficient for the task at hand.

Working in Python and using the Bregman, pyo, and pydub libraries, along with the LAME MP3 encoder, I begin with an uncompressed WAV file and save it as an MP3 file, 128kbps in this example, which sounds quite similar to the original.

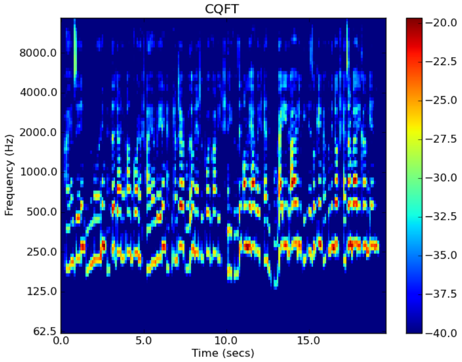

Figure 6. Tom's Diner – Verse 1 – Uncompressed WAV.

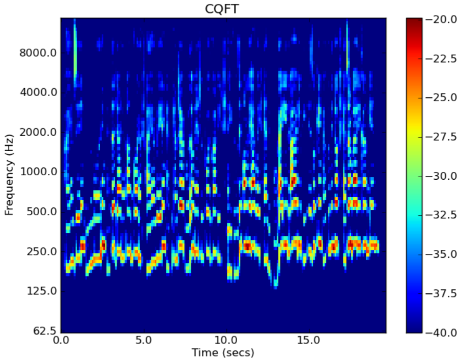

Figure 7. Tom's Diner – Verse 1 – 128kbps MP3.

I then take the Fast Fourier Transform of both files, first saving the now compressed MP3 in the same uncompressed format as the original so that the two can be directly compared. In my implementation, I have used both the STFT and constant-q transform for this step, though the discrete cosine transform or other transforms could be used as well. With both sound files now represented as adjacent time-frequency matrices, I compare the two files, bin by bin, taking the differences.

Where the two files are the same or similar, the information in the original audio has been largely preserved in the MP3. However, corresponding time-frequency bins which differ significantly between the two files betray spots where information has been altered or deleted. Where this is the case, we can store the difference between these two files in a new array of equal size to the originals.

Two possible ways forward emerge here: we can either resynthesize the new matrix directly using an inverse transform or, we can zero the corresponding bins in the original uncompressed file where the difference is null or near-null, i.e.- using the MP3 as a mask on the original file. In the work here, I have utilized both techniques to provide a variety of sonic material with which to compose. The sonic differences between these approaches is discussed below.

Each approach returns distinct, though related, sounds. What's more, in both cases, the STFT and inverse STFT involve phase estimation, and different ways of handling this also lead to slightly different results. Other variables include altering the bit and sample rates, and of course the various transform settings such as hop size, window size, etc. Worth noting is that 96 & 128kbps MP3's were used as the "high quality" controls in the original listening tests, while 64kbps were used as "low quality" [4].

In the compositional work detailed below, I have utilized the more recent 320 kbps standard for aesthetic reasons. Higher bit rates generated less information, however the material which is acquired is generally more distinct from the original sound file than the material acquired using low bit rate MP3's. The reason for this is simple– low bit rate MP3's delete more of the original audio file and thus, these artifacts more closely resemble the original sonically. Higher bit rate MP3's erase less content and this material is thus noticeably different from the original file, being only a carefully curated fraction of the original content.

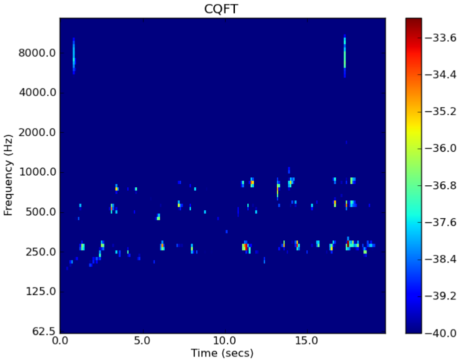

Figure 8. Tom's Diner – Verse 1 – 128kbps MP3 “Ghost”.

3. ARTISTIC BACKGROUND

As previously stated, the MP3 codec was refined using listening tests designed by and for primarily white, male, western-european audio engineers and featuring the music they chose. In a sense, each of these songs acts as a resonant filter for every file encoded in the MP3 format. Tom's Diner by Suzanne Vega, Fast Car by Tracy Chapman, a Haydn Trumpet concerto... these songs carved out the space of sounds that could be successfully encoded as MP3's. To that end, these songs represent a kind of best-case scenario for an MP3 encoding. If anything can be encoded well by this format, it should be these files. And yet these files still leave a residue behind when encoded to MP3. Exploring these sounds helps to define a boundary case for MP3 salvaging.

Further, if the perceptual model on which MP3 encoding is based is to be taken at face value, then these lost sounds are sounds which human ears should not be able to hear in the first place. Thus, by finding the “ghosts” in the MP3, we are finding not only the sounds which are deleted by the encoding process, but also uncovering sounds which previously could not be heard in their original context. The MP3 codec becomes a type of sonic-archeological tool by which we can uncover sounds from recorded history which exist acoustically in recorded media but previously were inaccessible to our perception due to the limitations of our auditory systems.

These concepts exist at the intersection of ideas from both glitch and plunderphonics. Glitch artists focus on digital noise and mechanical error as the substance of their compositions [7]. In contrast, this project examines the negative space of MP3 compression, rather than focussing directly on its sonic artifacts. Further, by salvaging and reworking material from popular culture, this music joins a lineage of sounding art offering cultural commentary, such as John Oswald's famous Plunderphonics project [8], and more recent mash-up culture [9].

4. COMPOSITIONAL TECHNIQUES

4.1 moDernisT

As a preliminary foray into codec ghost composition, I am creating a series of pieces based on the songs used in the original MP3 listening tests. As a preliminary, I'd like to briefly discuss my treatment of Tom's Diner. I begin by analyzing the song structure, interpreting the music and text, and I then attempt to arrange the most interesting recovered material via this framework. A case study of the techniques I've used in two verses follows.

4.2 Verse One – The Diner

The first verse finds the narrator in a bustling diner, making observations about her environment. The focus of this text is external to it's author, as opposed to later verses which exist in a more subjective, internal space. By using the lost information as both a mask on the original sound file and by resynthesizing directly, I was able to isolate both clear, pitched content and more ephemeral transient signals. By varying the masking threshold, experimenting with different window and hop sizes, and by either attempting phase estimation or simply discarding all phase information from the original signal, I was able to generate a fairly wide variety of material. Sifting through numerous slight variations on the two basic kinds of material, pitched and transient, I narrowed my focus to a collection of reconstructions which sounded either whisper-like or which offered pointillistically distributed pitches.

Using the python library headspace, and a reverb model of a small diner, I began to construct a virtual 3-d space. Beginning by fragmenting and scrambling the more transient material, I applied head related transfer functions to simulate the background conversation one might hear in a diner. Tracking the amplitude of the original melody in the verse, I applied a loose amplitude envelope to these signals. Thus, a remnant of the original vocal line comes through in its amplitude contour.

Having constructed this background, prominent pitches from the original melody appear and disappear, located variously in this virtual space. These ephemeral sounds hint at a familiar melody, playing with aural memory and imagination, a flickering apparition hovering at the border of consciousness.

4.3 Verse Five – “I am thinking of your voice...”

The fifth verse finds the narrator in a very different psychological state. Instead of buoyantly attending to the activity of the room, she is lost in thought, remembering. Accordingly, I have given this material more space. It is less fragmented, the constant background conversation has receded, the virtual space has drawn closer, it feels more internal than external. Key phrases and snippets of the melody emerge more clearly. When the outro arrives the familiar melody is once again obscured, replaced by mere hints at its former presence. We hear mostly transients, but internally we might fill in the rest.

5. FUTURE DIRECTIONS

Moving forward, I am planning a series of related compositions. First, I plan to explore the songs involved in the listening tests more deeply, both horizontally and vertically, by delving further into the sound world opened up by Tom's Diner, but also crafting new works from the other listening test songs. Following this, I envision working with newly created material to highlight even more explicitly the filtering effect of this codec, and it's relation to the approximations involved in our own auditory perception.

The songs used in developing the MP3 codec are notable for what they are not: they are not music from other cultures, not hip-hop or dance music, nothing with prominent low frequencies, nothing particularly noisy, no outright aggressive sounds, nothing lo-fi. Rather, these sounds have been broadly institutionally accepted and conform to accepted standards of production and recording technique [4]. As MP3's have invaded more and more contemporary listening spaces, the class of privileged sounds which the format inadvertently creates has become more apparent. Originally developed for suboptimal listening environments, MP3's are now heard everywhere, at home, streaming in stores and public spaces, over high-fidelity car stereo systems. This format has become a curator for these spaces: allowing in a great deal of wonderful sound, yes, but at the exclusion of a vast territory in the available sonic terrain. Composing with these sounds and injecting them back into contemporary listening spaces is one possible act of resistance, one available mode of cultural critique.

While the format is indeed based on a perceptual model of human hearing, this model is only an approximation. Many of the sounds deleted by MP3 encoding are indeed sounds that we would not hear in context. I submit that these sounds are nonetheless interesting for precisely that reason. This process reveals to us sounds that we would not otherwise hear due to the limitations of our perception. This is not the whole story though. We perceive sound not only through our ears but also through our bodies, especially when experiencing low frequencies as the mechanical vibrations that they are. The perceptual model is thus incomplete to begin with in that it equates the experience of sound with hearing only, when we know it a priori to be more than that. Further, it would be folly to assume that the implementation of the perceptual model used in MP3 encoding is in all cases perfect even given its limited premises, and a simple examination of the literature on listening tests, or even a simple experiment with ones own perception, reveals these shortcomings. Thus, the ghosts we find here are also sounds that are taken from us in the current sonic culture which values MP3's above all other formats and modes of musical dissemination. We thus draw attention to the limitations of this ubiquitous format and hope to point towards a day when we will not have to sacrifice sonic information in the interests of limited bandwidths and storage capacities.

There are various technical areas related to this work which are open territory for exploration: from developing an MP3 negative-space codec (an "anti-encoder") to exploring the idea of anti-filtering and anti-processing more generally, we might develop new tools to explicitly explore the sounds which our most commonly used techniques preclude. Further, we might develop real-time implementations of these effects. I have recently begun development of both real-time MP3 and Anti-MP3 filters. Numerous approaches are possible here and could lead to interesting new audio plug-ins and compositions based on these sounds. Finally, one might consider using a similar approach as this to uncover previously inaccessible sounds in the history of recorded music– sounds which, due to our perceptual limitations, might only be accessible now.

6. CONCLUSIONS

In conclusion, composing with MP3 files is an attempt to derive interesting material from lost sounds– sounds that we either can not hear otherwise or which have been filtered out of our contemporary listening spaces by technologically imposed perceptual models. As a composer, MP3 ghosts are difficult to predict and provide exciting, externally generated material to react to and work with, while not limiting the freedom of the artist to arrange, alter, and interpret these sounds. With the entire history of recorded music at our digital fingertips, the possibilities for exploration are immense.

Investigating a particular format for its aesthetic possibilities is inspired by musics built around previous technologies- "tape music", for example. I see "format music" as a contemporary analogue of these practices. Through questioning and exploring the limitations of the technologies with which I find myself intertwined, I hope to gain new insights into the limitations of my own perception and into what it means to be a composer, music enthusiast, and participant in sonic culture.

Audio Examples can be found online at: `http://theghostinthemp3.com`.

Acknowledgments

Special thanks goes to Tara Rodgers, Aden Evens, Larry, Polansky, and Michael Casey for their help formulating and articulating these ideas. Further thanks to Judith Shatin, Matthew Burtner, and the many friends and colleagues who have helped me refine and continue this work.

7. REFERENCES

[1] Brandenburg, Karlheinz, and Gerhard Stoll. “ISO/MPEG-1 Audio: A Generic Standard for Coding of High-quality Digital Audio.” Journal of the Audio Engineering Society 42, no. 10 (1994): 780–792.

[2] Pras, Amandine, Rachel Zimmerman, Daniel Levitin, and Catherine Guastavino. “Subjective Evaluation of Mp3 Compression for Different Musical Genres.” In Audio Engineering Society Convention 127, 2009.

[3] Sterne, Jonathan. “The Mp3 as Cultural Artifact.” New Media & Society 8, no. 5 (2006): 825–842.

[4] Sterne, Jonathan. MP3: The Meaning of a Format. Duke University Press Books, 2012.

[5] Evens, Aden. Sound Ideas: Music, Machines and Experiences. Minneapolis: University of Minnesota Press, 2005.

[6] Brandenburg, Karlheinz. “MP3 and AAC Explained.” In Audio Engineering Society Conference: 17th International Conference: High-Quality Audio Coding, 1999.

[7] Cascone, Kim. “The Aesthetics of Failure:‘Post-digital’ Tendencies in Contemporary Computer Music.” Computer Music Journal 24, no. 4 (2000): 12–18.

[8] Oswald, John. “Plunderphonics, or Audio Piracy as a Compositional Prerogative.” In Wired Society Electro-Acoustic Conference, 1985.

[9] Miller, Vincent. Understanding digital culture. Sage Publications, 2011.

[10] Bosi, Marina, and Richard E. Goldberg. “Introduction to digital audio coding and standards”. Springer, 2003.