

篇名:

ON LOSSY AUDIO COMPRESSION AND FOURIER TRANSFORM

作者:

陳立中。國立基隆高中。214 班

蘇冠綸。國立基隆高中。206 班

游坤明。國立基隆高中。214 班

I. ABSTRACT

There was once that I chanced upon a listening test^[1]. The listening test here refers to a test with the objective of rating various digital audio compression format, by having many to listen to a piece of digital audio recording encoded with various lossy audio compression format, then have them to rank how it sounds comparing to the uncompressed control set.

Since then, I realized that those compressed audio are actually lossy, in the sense that they are may varies slightly from the original audio data. Furthermore, sometime, they may even sound different from the original uncompressed sound track. This invoked my curiosity to look further, I wanted to know how digital audio compression works, how it is done, and how it affects the audio quality.

This thesis is going to cover what is digital audio, how they works, and how are they compressed.

Just a note here on the choice of software for writing this thesis. Originally, I planned to use LaTeX to typeset this document, as LaTeX is the de facto standard for typesetting thesis in almost all world class universities (for mathematics, scientific and engineering at least, that's what I know.), including Cambridge, Oxford, Stanford (Actually LaTeX is written by a Computing Professor in Stanford, and his name is Donald Knuth.), and Harvard. Almost all thesis from the aforementioned universities are typesetted with LaTeX. However, the specified syntax and format by the shs.edu.tw is slightly different from the LaTeX's format. Therefore I decided to stick to word processing softwares like OpenOffice.org Writer or Microsoft Word. I choose OpenOffice.org Writer, for its ability to export its contents directly to PDF file, without any "created with <Insert random software name here> trial version -- <insert random URL here>" nonsense.

II. THESIS

2.1 What is digital audio

Before we can look at any compression method/algorithm, we must first be in possession of the knowledge on what is digital audio. It is certain that every of you readers knows the meaning of audio, so I won't try to explain it. In the next section, we'll try to explain what is digital, and how is it different from the traditional analog audio.

2.1.1 Definition of "digital"

When electronic information processing system (such as computer) first appeared, they were all analog. Analog means that the informations being processed are in analog form, they are represented by a voltage value that is continuous. Meaning that all voltage values within a certain domain are valid representation of information.

However, later on, as the system progressed to be more and more complex, distortion and noise that affected the signals became a significant problem, such that people came up with another method of representing information, that is the "digital" method.

The digital method of representing information involves quantization of information, then represent

them in integers. Normally in base 2, so that only two voltage value is a valid representation of information.

2.1.2 Quantization of Audio signal

As we know, sound is a fluctuation of the medium in which it is being transmitted. A microphone is used to convert this fluctuation into an electrical signal, such that every regular interval in time is assigned a value. This process of assigning value to the time domain is called sampling.

From the definition of digital, the values that are assigned to the time domain is not continuous, such that those values are quantized into integer domain.

The whole process of quantization of audio signal and assigning the value to the time domain is called **Pulse-Code Modulation**.^[2]

2.1.2.1 The Nyquist-Shannon sampling theorem

In the previous section, we mentioned that every regular interval in time is assigned a value, but what is this so-called “regular interval”. To determine this regular interval, let us look at the Nyquist-Shannon sampling theorem. The Nyquist-Shannon sampling theorem states that:

If a function $u(t)$ contains no frequencies higher than W cps, it is completely determined by giving its ordinates at a series of points spaced $1/(2W)$ seconds apart.^[3]

Take note that the “cps” in the theorem is “Cycle per Second”, which is equivalent to the modern unit of Hertz.

The theorem says that a (audio) signal can be completely reconstructed if it is sampled at twice the highest frequency in the signal.

It is known that the frequency range that can be detected by human ear is from 20 Hz to 20 kHz, so, the highest frequency is 20 kHz. Therefore, according to the Nyquist-Shannon sampling theorem, in order to be able to reconstruct a human audible audio signal, we need to sample the audio signal at a minimal sampling frequency of 40 kHz.

Out of all the sampling frequency, the 44.1 kHz is the most popular of all. It is used in almost all digital audio applications, in the popular Compact Disc format, most *mp3s*, *ogg vorbis*, *aac*, and *wma*, with the possible exception of high end audio system, which samples at 194 kHz.

Take note that digital audio formats such *mp3*, *ogg vorbis*, *aac*, and *wma*, supports many different sampling frequency, there's no forced standard with these formats.

2.2 Compression algorithms

It is known that there is only a certain range of sound that human auditory system can perceive. Furthermore, the sensitivity of the human auditory system to the audio of various frequency within

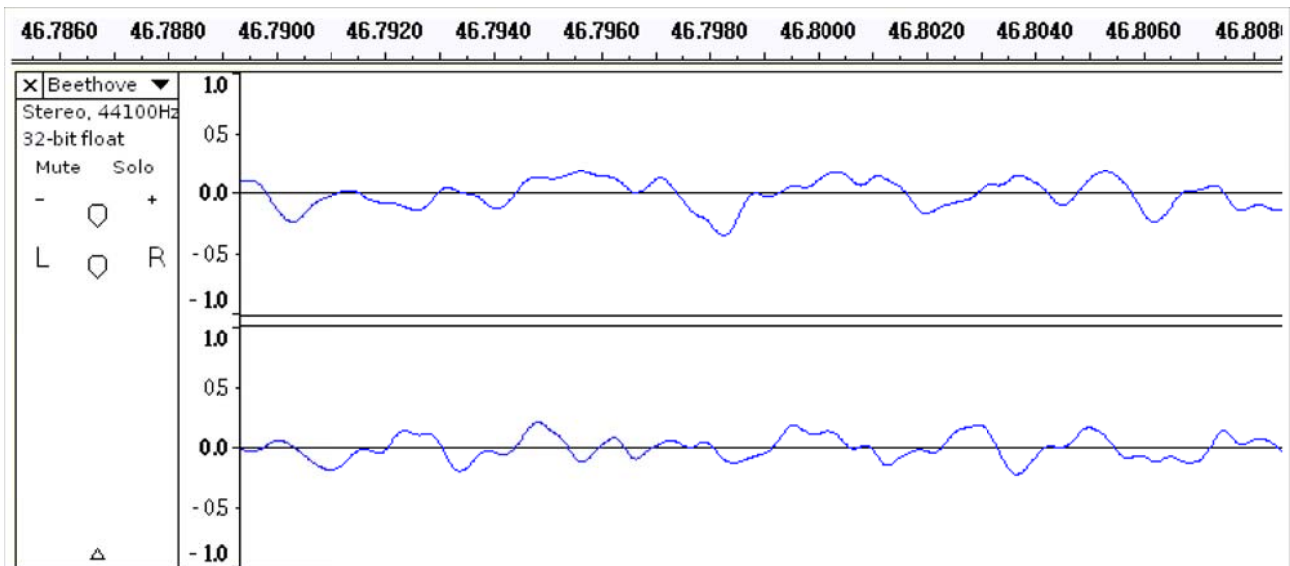
the domain is different. Also, when there is a loud sound, other sounds will be “shadowed” by it, such that we can barely notice it. The study on how the human auditory system perceive sound wave is called **psychoacoustics**.

Most lossy compression algorithms exploits these property of the human auditory system, in order to achieve better compression ratio.

Besides removal of sounds that are barely noticeable, audio compression software also use a method called **Noise Shaping**. Noise shaping refers to using lesser data to encode sound frequencies that are not so sensitive to the human auditory system. Common sense tells us that with lesser data to represent a signal, the signal will be more distorted. However, as those sound frequencies are not so sensitive to the human auditory system, a difference in a few Hertz doesn't really make a difference.

2.2.1 Distinguishing Frequencies

If we take a look at some audio signal, such as those that we see in Figure 1^[4]:



Δ Figure 1: Audio signal (From Piano Sonata No 14. in C# minor “Quasi una fantasia”, Opus 27, No. 2, by Ludwig van Beethoven, also known as Moonlight Sonata)

High school physics tells us that sound waves are sine waves. However, what we see above, a segment of sound signal from Moonlight Sonata, is not sine wave.

Actually, sound can be made up of many sine waves, each with their own frequency, amplitude, and phase. Even such simple piece of sound signal, such as the one shown in Figure 1, in which, there is only one instrument -- Piano

As stated in the previous sections, many audio compression software make use of psychoacoustic property of the human auditory system. They crop off the inaudible frequencies, reduce the bitrate of the less sensitive sound signals... etc. However, from what we see in Figure 1, all the frequencies are mixed up! So, how do they differentiate a frequency from another?

The answer is a mathematical device called the Fourier Transform.

2.2.2 The Fourier Transform

The Fourier Transform, as its name suggests, is named after a French mathematician – Jean Baptiste Joseph Fourier (March 21, 1768 – May 16, 1830). The word “transform” tells us that the method maps a function into another function.

Fourier Transform is actually a method to map a time domain function into a frequency domain function. For example, let's say there's a function:

$$f(t) = \text{The value that is assigned to time } t$$

Then, with Fourier Transform, we can map this function into another function:

$$F(x) = \text{The amplitude and phase of the signal at frequency } x$$

This way, we can separate a sound signal into various frequency, so that the audio compression software can identify those frequency and make use of the psychoacoustic property

The Fourier Transform is defined as:

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi f t} dt$$

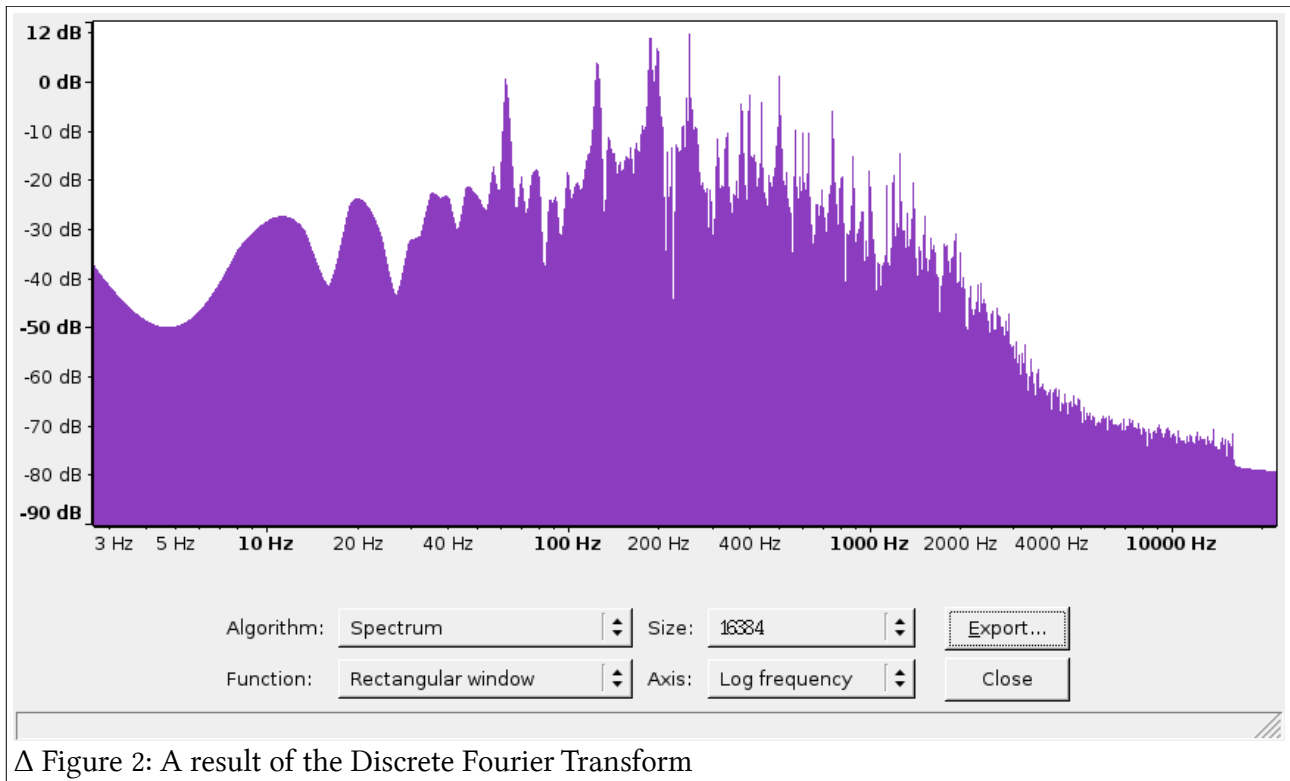
However, in Senior High School mathematics lesson, we know that only continuous function is integrable. From the section on digital audio, we know that digital audio signal is *not* continuous, so there's a problem.

Luckily, there is another version of the Fourier Transform, which is called the Discrete Fourier Transform. The “discrete” in its name tells us that it doesn't act on continuous function, instead, it acts on discrete sets of values, just like our digital audio signals. The Discrete Fourier Transform is defined as:^[5]

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N} k n} \quad k = 0 \dots\dots N - 1$$

As seen from above, the output is a discrete sets of values as well, which is advantageous, as the digital information processing system is suited to process this type of information, as opposed to the continuous function.

Let us look at the DFT results of the sound signal shown in Figure 1:



2.3 Storing Audio Information

Most audio compression formats use a method to store the audio data. The method is to divide all data into **frames**. For example, all data frames in MP3 audio format holds exactly 26 ms (0.026 second) of audio information. Each frame have header that hold information regarding what is stored within that frame, such as size of the frame, bitrate, type of data within it, or data version.

For some formats, a data frame can hold not only audio information, but also other informations such as the title, artist, album for the particular track.

2.3.1 Bitrate

Bitrate refers how many storage is used to encode a particular segment of audio data. For example, is 128 kilobytes are used to encode a second of audio data, then the bitrate for the segment of compressed audio data is 128 kbps (KiloByte Per Second). Usually, the bitrate is constant through out a frame.

Sometime, bitrate varies from frame to frame, as some audio information requires more data to encode, while some requires lesser to encode. This method of varying bitrate throughout an audio file is called VBR – Variable Bitrate.

III. CONCLUSION

From this thesis, we learn that lossy data compression is actually a method that discard audio signals

that are irrelevant to human hearing, by using psychoacoustic methods to determine those signals. Furthermore, we learn a useful mathematical device, called The Fourier Transform, which maps a time domain function into a frequency domain function. It is applicable in many fields including audio compression, signal processing ... etc.

IV. REFERENCES

- [1] – Listening Test on HydrogenAudio Forum (MPC vs. Vorbis vs. MP3 vs. AAC) - <http://www.hydrogenaudio.org/forums/index.php?showtopic=36465> – Retrieved on 9th of March, 2008
- [2] - Introduction on Pulse Code Modulation - <http://cbdd.wsu.edu/kewlcontent/cdoutput/TR502/page13.htm> – Accessed on 9th of March, 2008
- [3] – English Wikipedia Article on Nyquist-Shannon Sampling Theorem - http://en.wikipedia.org/wiki/Sampling_theorem – Accessed on 9th of March, 2008
- [4] – Figure 1 is made by myself with Audacity - <http://audacity.sourceforge.net/> - Accessed on 9th of March, 2008
- [5] – Mathematics of The Discrete Fourier Transform with Audio Applications (Second Edition) by Julius O. Smith III – Published by Stanford University – Chapter 1-1