International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Audio Compression Using Fourier Transform

Perinati Tewari

Abstract: ***All data compression algorithms consist of at least a model and a coder (with optional pre-processing transforms). The coder assigns shorter codes to the more likely symbols. There are efficient and optimal solutions to the coding problem. However, optimal modelling has been proven not computable. Modelling (or equivalently, prediction) is both an artificial intelligence (AI) problem and an art. Lossy compression consists of a transform to separate important from unimportant data, followed by lossless compression of the important part and discarding the rest. The audio compression software crop off the inaudible frequencies, reduce the bit rate of the less sensitive sound signals etc. Thus in this project our aim is to use Fourier Transform to differentiate a frequency from another.***

Keywords: audio compression, lossless, lossy

1. Introduction

1.1 Background

When electronic information processing system (such as computer) first appeared, they were all analog. Analog means that the information 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 dig 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.

1.2 Problem Statement

All data compression algorithms consist of at least a model and a coder (with optional pre-processing transforms). The coder assigns shorter codes to the more likely symbols. There are efficient and optimal solutions to the coding problem. However, optimal modelling has been proven not computable. Modelling (or equivalently, prediction) is both an artificial intelligence (AI) problem and an art. Lossy compression consists of a transform to separate important from unimportant data, followed by lossless compression of the important part and discarding the rest.The audio compression software crop off the inaudible frequencies, reduce the bit rate of the less sensitive sound signals etc. Thus in this project our aim is to use Fourier Transform to differentiate a frequency from another.

1.3 Objective

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. The audio compression software crop off the inaudible frequencies, reduce the bitrate of the less sensitive sound signals etc. Thus in this project we use Fourier Transform to differentiate a frequency from another.

1.4 Future Scope

Although researchers are pursuing many new directions, clear trends that are emerging. AAC and its variations (HE-AAC, Surround, and parametric extensions) will become the foundations for future improvements and will slowly replace legacy MP3 in usage as well as content. Lossless audio coding and its applications will grab a prominent mindshare thanks to the fidelity advantages that they offer. Proprietary codecs will continue to hold market share in certain applications such as streaming media, portable multimedia, and professional audio.

1.5 Signals

1.5.1 Continuous time signal

A continuous signal or a continuous-time signal is a varying quantity (a signal) whose domain, which is often time, is a continuum (e.g., a connected interval of the reals). That is, the function's domain is an uncountable set. The function itself need not be continuous. To contrast, a discrete time signal has a countable domain, like the natural numbers.

A signal of continuous amplitude and time is known as a continuous-time signal or an analogsignal. This (a signal) will have some value at every instant of time. The electrical signals derived in proportion with the physical quantities such as temperature, pressure, sound etc. are generally continuous signals. Other examples of continuous signals are sine wave, cosine wave, triangular wave etc.

The signal is defined over a domain, which may or may not be finite, and there is a functional mapping from the domain to the value of the signal. The continuity of the time variable, in connection with the law of density of real numbers, means that the signal value can be found at any arbitrary point in time.

A typical example of an infinite duration signal is:

$$f(t) = \sin(t), \quad t \in \mathbb{R}$$

A finite duration counterpart of the above signal could be:

$$f(t) = \sin(t), \quad t \in [-\pi, \pi]$$

And

$f(t) = 0$ otherwise.

The value of a finite (or infinite) duration signal may or may not be finite. For example,

$$f(t) = \frac{1}{t}, \quad t \in [0, 1]$$

And

$f(t) = 0$ otherwise,

is a finite duration signal but it takes an infinite value for $t = 0$.

In many disciplines, the convention is that a continuous signal must always have a finite value, which makes more sense in the case of physical signals.

For some purposes, infinite singularities are acceptable as long as the signal is integrable over any finite interval (for example, the t^{-1} signal is not integrable, but t^{-2} is).

Any analog signal is continuous by nature. Discrete-time signals, used in digital signalprocessing, can be obtained by sampling and quantization of continuous signals.

Continuous signal may also be defined over an independent variable other than time. Another very common independent variable is space and is particularly useful in image processing, where two space dimensions are used.

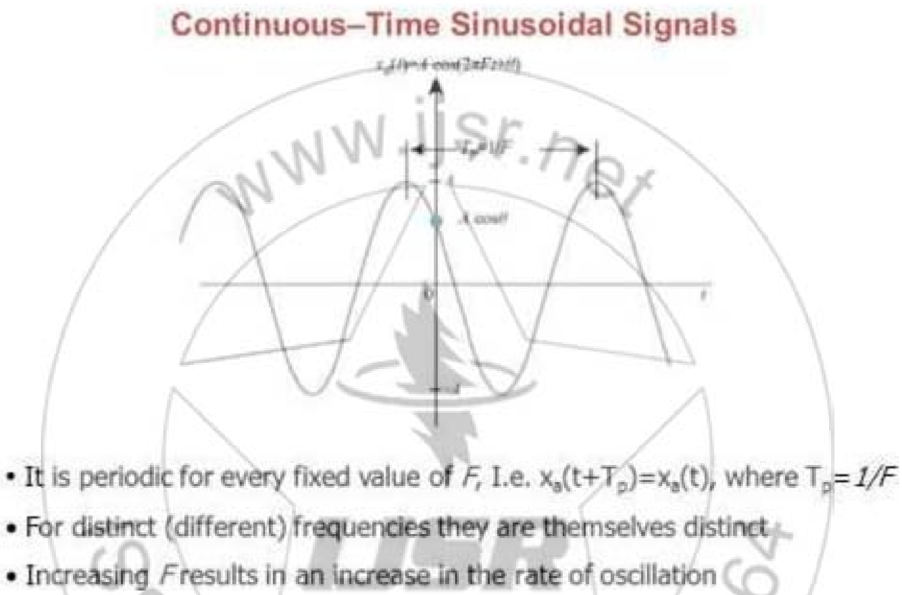

Figure 2.1: Continuous time signal

1.5.2 Discrete Time Signals

A discrete signal or discrete-time signal is a time series consisting of a sequence of quantities. In other words, it is a time series that is a function over a domain of integers.

Unlike a continuous-time signal, a discrete-time signal is not a function of a continuous argument; however, it may have been obtained by sampling from a continuous-time signal, and then each value in the sequence is called a sample. When a discrete-time signal obtained by sampling a sequence corresponds to uniformly spaced times, it has an associated sampling rate; the sampling rate is not apparent in the data sequence, and so needs to be associated as a characteristic unit of the system.

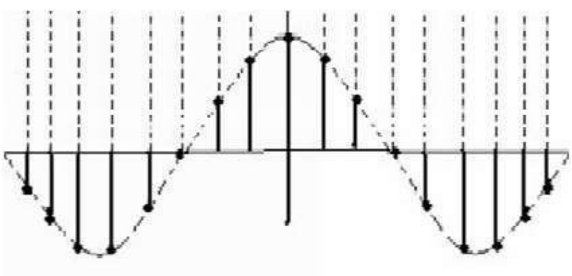

Figure 2.2: Discrete time signal

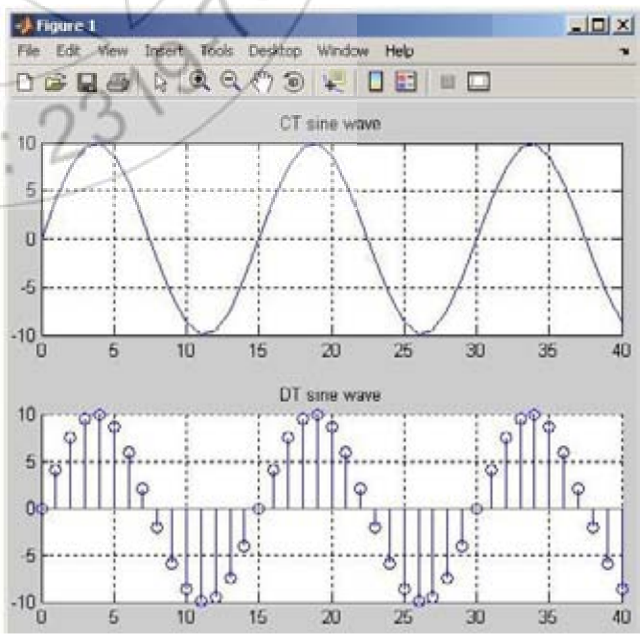

Figure 2.3: Continuous and Discrete time signals represented in MATLAB

2. Audio Compression Techniques

2.1 Useful auditory properties

2.1.1 Non-linear frequency response of the hear

Humans are able to hear frequencies in the range approximately from 20 Hz to 20 kHz. However, this does not mean that all frequencies are heard in the same way. One could make the assumption that a human would hear frequencies that make up speech better than others, and that is in fact a good guess. Furthermore, one could also hypothesize that hearing a tone becomes more difficult close to the extremes frequencies (i.e. close to 20 Hz and 20kHz).After many cochlear studies, scientists have found that the frequency range from 20 Hz to 20 kHz can be broken up into critical bandwidths, which are non-uniform, non-linear, and dependent on the level of the incoming sound. Signals within one critical bandwidth are hard to separate for a human observer. A detailed description of this behaviour is described in the Bark scale and Fletcher curves.

2.1.2 Masking property of the auditory system

Auditory masking is a perceptual property of the human auditory system that occurs whenever the presence of a strong audio signal makes a temporal or spectral neighbourhood of weaker audio signal imperceptible. This means that the masking effect can be observed in time and frequency domain. Normally they are studied separately and known as simultaneous masking and temporal masking. If two sounds occur simultaneously and one is masked by the other, this is referred to as simultaneous masking. A sound close in frequency to a louder sound is more easily masked than if it is far apart in frequency. For this reason, simultaneous masking is also sometimes called frequency masking. It is important to differentiate between tone and noise maskers, because tonality of a sound also determines its ability to mask other sounds.

A sinusoidal masker, for example, requires a higher intensity to mask a noise like masker than a loud noise-like masker does to mask a sinusoid. Similarly, a weak sound emitted soon after the end of a louder sound is masked by the louder sound. Infact, even a weak sound just before a louder sound can be masked by the louder sound. These two effects are called forward and backward temporal masking, respectively. Temporal masking effectiveness attenuates exponentially from the onset and offset of the masker, with the onset attenuation lasting approximately 10 ms and the off-set attenuation lasting approximately 50 ms. It is of special interest for perceptual audio coding to have a precise description of all masking phenomena to compute a masking threshold that can be used to compress a digital signal. Using this, it is possible to reduce the SNR and therefore the number of bits. A complete masking threshold should be calculated using the principles of simultaneous masking and temporal masking and the frequency response of the ear. In the perceptual audio coding schemes, these masking models are often called psychoacoustic models.

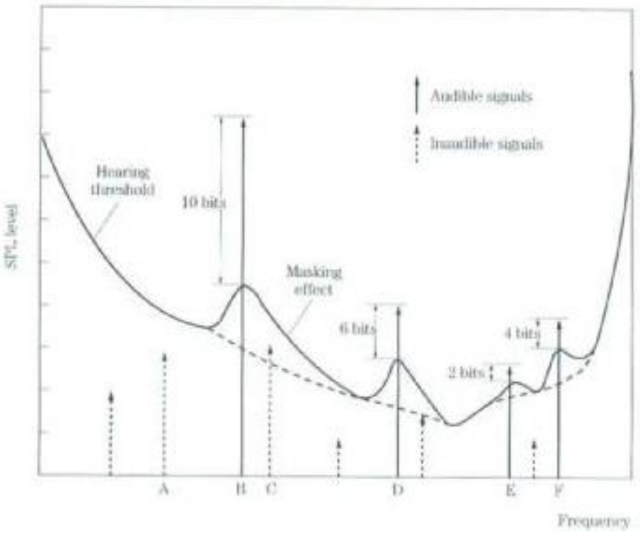

Figure 3.1: An example that shows how the auditory properties can be used to compress and digital audio signal

2.2 Audio Compression

The idea of audio compression is to encode audio data to take up less storage space and less bandwidth for transmission. To meet this goal different methods for compression have been designed. Just like every other digital data compression, it is possible to classify them into two categories: lossless compression and lossy compression.

2.2.1 Lossless compression

Lossless compression in audio is usually performed by waveform coding techniques. These coders attempt to copy the actual shape of the analog signal, quantizing each sample using different types of quantization. These techniques attempt to approximate the waveform, and, if a large enough bit rate is available they get arbitrary close to it. A popular waveform coding technique, that is considered uncompressed audio format, is the pulse code modulation (PCM), which is used by the Compact Disc Digital Audio (or simply CD). The quality of CD audio signals is referred to as a standard for hi-fidelity. CD audio signals are sampled at 44.1 kHz and quantized using 16 bits/sample Pulse Code Modulation (PCM) resulting in a very high bit rate of 705 kbps. As mentioned before, human perception of sound is affected by SNR, because adding noise to a signal is not as noticeable if the signal energy is large enough. When digitalize an audio signal, ideally SNR could to be constant for al quantization levels, which requires a step size proportional to the signal value. This kind of quantization can be done using a logarithmic compander (compressor-expander). Using this technique it is possible to reduce the dynamic range of the signal, thus increasing the coding efficiency, by using fewer bits. The two most common standards are the μ-law and the A-law, widely used in telephony. Other lossless techniques have been used to compress audio signals, mainly by finding redundancy and removing it or by optimizing the quantization process. Among those techniques it is possible to find Adaptative PCM and Differential quantization. Other lossless techniques such as Huffman coding and LZW have been directly applied to audio compression without obtaining significant compression ratio.

2.2.2 Lossy compression

Opposed to lossless compression, lossy compression reduces perceptual redundancy; i.e. sounds which are considered perceptually irrelevant are coded with decreased accuracy or not coded at all.

3. Fourier Transform

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 So, how do they differentiate a frequency from another?

The answer is a mathematical device called the Fourier Transform.

The Fourier Transform is defined as:

$$X(\omega) = \int_{-\infty}^{+\infty} x(t) e^{-j\omega t} dt \qquad \text{Fourier Transform}$$

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} X(\omega) e^{j\omega t} dt \qquad \text{Inverse Fourier Transform}$$

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, in acts on discrete sets of values, just like our digital audio signals.

The Discrete Fourier Transform is defined as

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{\frac{-i2\pi nk}{N}}, k=0 \ldots N-1$$

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

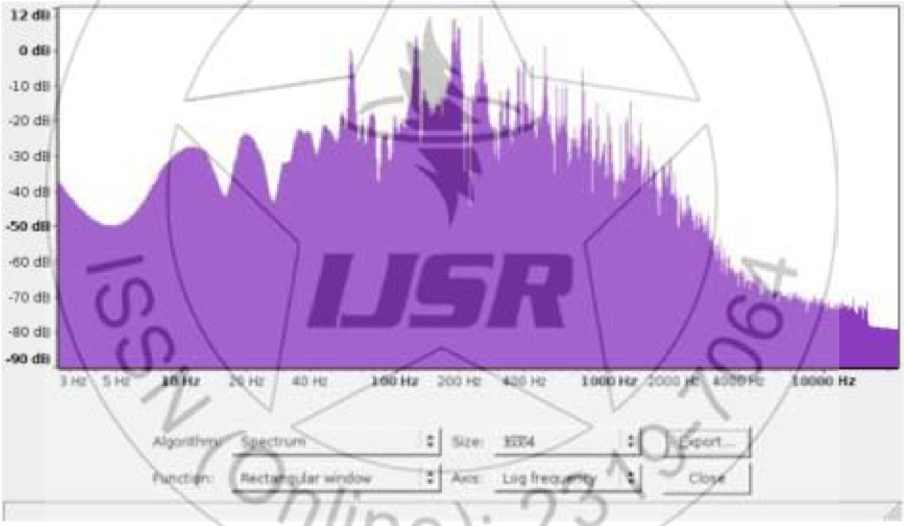

Figure 5.1: Sound Signal

4. Results and Discussion

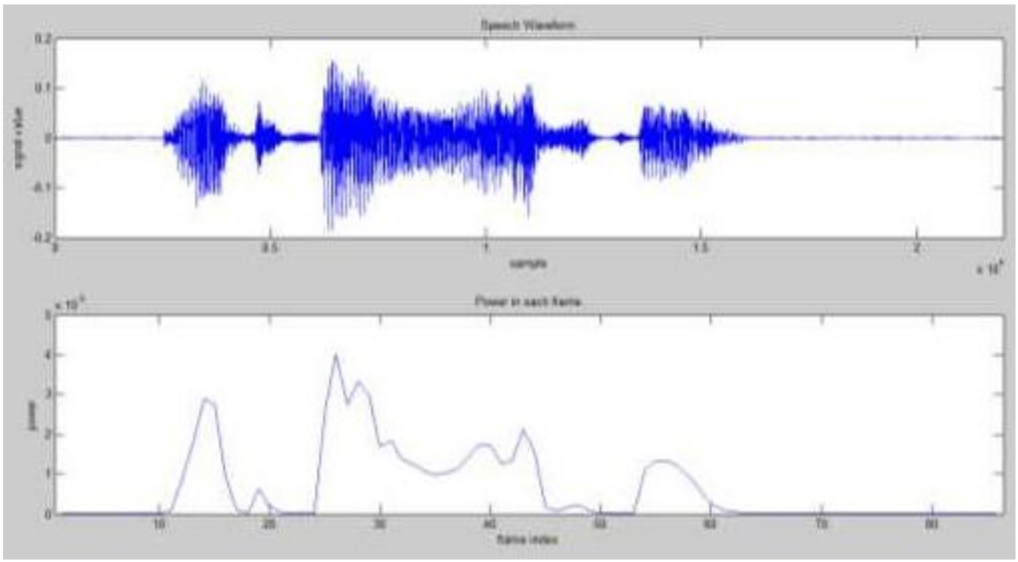

Figure 8.1: Speech waveform and power in each frame

Level Detection - The start of an input audio signal is identified by a stored threshold value.

Framing - The continuous signal is blocked into frames of N samples. (N=256) Each frame consists of 256 samples of audio signals. Each frame overlaps with subsequent frames. This technique is called framing.

Windowing- After framing, windowing is applied to prevent leakage.

Fast Fourier Transform – The FFT converts the time domain signal into frequency domain to yield a complex signal. The signal's FFT has both real and imaginary components.

Power spectrum calculation- The power of the frequency domain is calculated by summing the square of real and imaginary components of the signal to yield a real signal. The second half of the samples in the frame are ignored since they are symmetric to the first half (audio signal being real)

Approximate Signal – The signal in RED is the approximate signal drawn by taking the maximum frequency components of the maximum power frame. The signal in BLUE is the original signal.

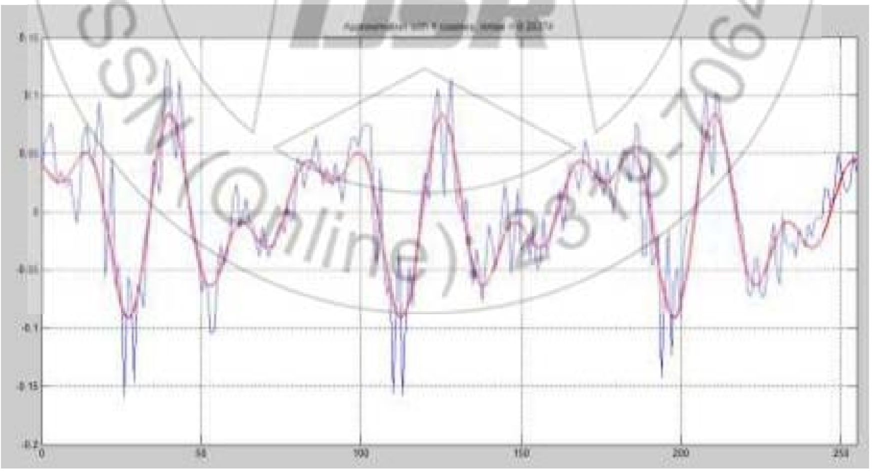

Figure 8.2: Comparison of original and approximate signal

References

[1] Listening Test on Hydrogen Audio 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] Mathematics of The Discrete Fourier Transform with Audio Applications (Second Edition) by Julius O. Smith III – Published by Stanford University – Chapter 1-17