
Analysis of the sound quantization

Pure tones don't naturally exist but every sound in the world is the sum multiple pure tones at different amplitudes. A music song is played by multiple instruments and singers. All those instruments produce a combination of sinewaves at multiples frequencies and the overall is an even bigger combination of sinewaves.

A spectrogram is a very detailed, accurate image of your audio, displayed in either 2D or 3D. Audio is shown on a graph according to time and frequency, with brightness or height (3D) indicating amplitude. Whereas a waveform shows how your signal's amplitude changes over time, the spectrogram shows this change for every frequency component in the signal.

As an example, you can see the droplet impact consistently forms large surface bubbles and the standard "bloop" noise on the fig. 4. The color represents the amplitude in dB. In this spectrogram some frequencies are more important than others, so we can build a fingerprinting algorithm.

Analog signals are continuous signals, which means if you take one second of an analog signal, you can divide this second into parts that last a fraction of second. In the digital world, you can't afford to store an infinite amount of information. You need to have a minimum unit, for example, 1 millisecond. During this unit of time the sound cannot change so this unit needs to be short enough so that the digital song sounds like the analog one and big enough to limit the space needed for storing the music.

The Nyquist sampling theorem provides a prescription for the nominal sampling interval required to avoid aliasing. It may be stated simply as follows: the sampling frequency should be at least twice the highest frequency contained in the signal. Or in mathematical terms: $f_s = 2 f_c$ where f_s is the sampling frequency (how often samples are taken per unit of time or space), and f_c is the highest frequency contained in the signal.

A theorem from Nyquist and Shannon states that if you want to digitalize a signal from 0Hz to 20kHz you need at least 40 001 samples per second. The standard sampling rate for digital music in the music industry is 44.1kHz and each sample is assigned 16 bits. Some theorem definitions describe this process as making a perfect recreation of the signal. The main idea is that a sine wave signal at a frequency F needs at least 2 points per cycle to be identified. If the frequency of your sampling is at least twice than the frequency of your signal, you'll end up with at least 2 points per cycle of the original signal.

Need help with the assignment?

Our professionals are ready to assist with any writing!

[GET HELP](#)

Sampling, the process of converting a signal into a numeric sequence is also called analog-to-digital conversion. Quantization is another process of the conversion, which is the accurate measurement of each sample. Analog to digital converters and digital to analog converters encode and decode these signals to record our voices, display pictures on the screen, or to play audio clips through speakers. Since we can digitize media we can handle, recreate, alter, produce, and store text, images, and sounds.

The theorem even though it can be seen as simple has changed the way our modern digital world works. We can uniformly use media to our advantage in multiple numbers of ways. The limitations we have can be addressed through filters and adjusting our sample rates or frequencies. Though it hasn't the same shape nor the same amplitude, the frequency of the sampled signal remains the same.

The analog-to-digital converters perform this type of function to create a series of digital values out of the given analog signal. The following figure represents an analog signal. This signal to get converted into digital has to undergo sampling and quantizing.

Analysis of The Sound Quantization

Quantization is the process of mapping input values from a large set (often a continuous set) to output values in a (countable) smaller set. Rounding and truncation are typical examples of quantization processes. Quantization is involved to some degree in nearly all digital signal processing, as the process of representing a signal in digital form ordinarily involves rounding. Quantization also forms the core of essentially all lossy compression algorithms.

Quantization makes the range of a signal discrete so that the quantized signal takes on only a discrete, usually finite, set of values. Unlike sampling, quantization is generally irreversible and results in loss of information. It, therefore, introduces distortion into the quantized signal that cannot be eliminated.

One of the basic choices in quantization is the number of discrete quantization levels to use. The fundamental tradeoff in this choice is the resulting signal quality versus the amount of data needed to represent each sample. Fig. 6 shows an analog signal and quantized versions for several different numbers of quantization levels. With L levels, we need $N = \log_2 L$ bits to represent the different levels, or conversely, with N bits we can represent $L = 2^N$ levels.

Pulse-Code Modulation of The Sound

Pulse-code modulation (PCM) is a system used to translate analog signals into digital data. It is

Need help with the assignment?

Our professionals are ready to assist with any writing!

GET HELP

used by compact discs and most electronics devices. For example, when you listen to an mp3 file in your computer/phone/tablet, the mp3 is automatically transformed into a PCM signal and then send to your headphones.

A PCM stream is a stream of organized bits. It can be composed of multiple channels. For example, a stereo music has 2 channels. In a stream, the amplitude of the signal is divided into samples. The number of samples per second corresponds to the sampling rate of the music. For instance, a 44,1kHz sampled music will have 44100 samples per second. Each sample gives the (quantized) amplitude of the sound of the corresponding fraction of seconds.

There are multiple PCM formats but the most used one in audio is the (linear) PCM 44,1kHz, 16-bit depth stereo format. This format has 44 100 samples for each second of music. Each sample takes 4 bytes (Fig. 7):

- 2 bytes (16 bits) for the intensity (from -32,768 to 32,767) of the left speaker
- 2 bytes (16 bits) for the intensity (from -32,768 to 32,767) of the right speaker

In a PCM 44,1kHz 16-bit depth stereo format, you have 44100 samples like this one for every second of music.

Discrete Fourier Transform Algorithm

The DFT (Discrete Fourier Transform) applies to discrete signals and gives a discrete spectrum (the frequencies inside the signal). The discrete Fourier transform (DFT) is a method for converting a sequence of N complex numbers x_0, x_1, \dots, x_{N-1} to a new sequence of N complex numbers

In this formula:

- N is the size of the window: the number of samples that composed the signal
- $X(n)$ represents the n th bin of frequencies
- $x(k)$ is a k th sample of the audio signal

The DFT is useful in many applications, including the simple signal spectral analysis. Knowing how a signal can be expressed as a combination of waves allows for manipulation of that signal and comparisons of different signals:

- Digital files (jpg, mp3, etc.) can be shrunk by eliminating contributions from the least important waves in the combination.
- Different sound files can be compared by comparing the coefficients $x(k)$ of the DFT.

Need help with the assignment?

Our professionals are ready to assist with any writing!

GET HELP

-
- Radio waves can be filtered to avoid "noise" and listen to the important components of the signal.

Other applications of the DFT arise because it can be computed very efficiently by the fast Fourier transform (FFT) algorithm. For example, the DFT is used in state-of-the-art algorithms for multiplying polynomials and large integers together; instead of working with polynomial multiplication directly, it turns out to be faster to compute the DFT of the polynomial functions and convert the problem of multiplying polynomials to an analogous problem involving their DFTs.

Window Functions

In signal processing, a window function is a mathematical function that is zero-valued outside of some chosen interval. For instance, a function that is constant inside the interval and zeroes elsewhere is called a rectangular window, which describes the shape of its graphical representation. When another function or waveform/data-sequence is multiplied by a window function, the product is also zero-valued outside the interval: all that is left is the part where they overlap, the "view through the window".

In typical applications, the window functions used are non-negative, smooth, "bell-shaped" curves. Rectangle, triangle and other functions can also be used. A more general definition of window functions does not require them to be identically zero outside an interval, as long as the product of the window multiplied by its argument is square integrable, and, more specifically, that the function goes sufficiently rapidly toward zero.

The Fourier transform of the function $\cos \omega t$ is zero, except at frequency $\pm \omega$. However, many other functions and waveforms do not have convenient closed-form transforms. Alternatively, one might be interested in their spectral content only during a certain time period.

In either case, the Fourier transform (or a similar transform) can be applied on one or more finite intervals of the waveform. In general, the transform is applied to the product of the waveform and a window function. Any window (including rectangular) affects the spectral estimate computed by this method.

Windowing of a simple waveform like $\cos \omega t$ causes its Fourier transform to develop non-zero values (commonly called spectral leakage) at frequencies other than ω . The leakage tends to be worst (highest) near ω and least at frequencies farthest from ω .

If the waveform under analysis comprises two sinusoids of different frequencies, leakage can interfere with the ability to distinguish them spectrally. If their frequencies are dissimilar and one

Need help with the assignment?

Our professionals are ready to assist with any writing!

GET HELP

component is weaker, then leakage from the stronger component can obscure the weaker one's presence. But if the frequencies are similar, leakage can render them unresolvable even when the sinusoids are of equal strength. The rectangular window has excellent resolution characteristics for sinusoids of comparable strength, but it is a poor choice for sinusoids of disparate amplitudes. This characteristic is sometimes described as a low dynamic range.

At the other extreme of dynamic range are the windows with the poorest resolution and sensitivity, which is the ability to reveal relatively weak sinusoids in the presence of additive random noise. That is because the noise produces a stronger response with high-dynamic-range windows than with high-resolution windows. Therefore, high-dynamic-range windows are most often justified in wideband applications, where the spectrum being analyzed is expected to contain many different components of various amplitudes.

In between the extremes are moderate windows, such as Hamming and Hann. They are commonly used in narrowband applications, such as the spectrum of a telephone channel. In summary, spectral analysis involves a trade-off between resolving comparable strength components with similar frequencies and resolving disparate strength components with dissimilar frequencies. That trade-off occurs when the window function is chosen.

When the input waveform is time-sampled, instead of continuous, the analysis is usually done by applying a window function and then a discrete Fourier transform (DFT). But the DFT provides only a sparse sampling of the actual discrete-time Fourier transform (DTFT) spectrum. Fig. 8 shows a portion of the DTFT for a rectangularly-windowed sinusoid. The actual frequency of the sinusoid is indicated as "0" on the horizontal axis. Everything else is leakage, exaggerated by the use of a logarithmic presentation. The unit of frequency is "DFT bins"; that is, the integer values on the frequency axis correspond to the frequencies sampled by the DFT.

So the figure depicts a case where the actual frequency of the sinusoid coincides with a DFT sample, and the maximum value of the spectrum is accurately measured by that sample. When it misses the maximum value by some amount (up to $\frac{1}{2}$ bin), the measurement error is referred to as scalloping loss (inspired by the shape of the peak). For a known frequency, such as a musical note or a sinusoidal test signal, matching the frequency to a DFT bin can be prearranged by choices of a sampling rate and a window length that results in an integer number of cycles within the window.

In signal processing, operations are chosen to improve some aspect of quality of a signal by exploiting the differences between the signal and the corrupting influences. When the signal is a sinusoid corrupted by additive random noise, spectral analysis distributes the signal and noise components differently, often making it easier to detect the signal's presence or measure certain characteristics, such as amplitude and frequency. Effectively, the signal to noise ratio (SNR) is

Need help with the assignment?

Our professionals are ready to assist with any writing!

[GET HELP](#)

improved by distributing the noise uniformly, while concentrating most of the sinusoid's energy around one frequency.

Processing gain is a term often used to describe an SNR improvement. The processing gain of spectral analysis depends on the window function, both its noise bandwidth and its potential scalloping loss. These effects partially offset, because windows with the least scalloping naturally have the most leakage. The frequencies of the sinusoids are chosen such that one encounters no scalloping and the other encounters maximum scalloping. Both sinusoids suffer less SNR loss under the Hann window than under the Blackman–Harris window. In general (as mentioned earlier), this is a deterrent to using high-dynamic-range windows in low-dynamic-range applications.

The human ear automatically and involuntarily performs a calculation that takes the intellect years of mathematical education to accomplish. The ear formulates a transform by converting sound – the waves of pressure traveling over time and through the atmosphere – into a spectrum, a description of the sound as a series of volumes at distinct pitches. The brain then turns this information into perceived sound.

A similar conversion can be done using mathematical methods on the same sound waves or virtually any other fluctuating signal that varies with respect to time. The Fourier transform is the mathematical tool used to make this conversion. Simply stated, the Fourier transform converts waveform data in the time domain into the frequency domain. The Fourier transform accomplishes this by breaking down the original time-based waveform into a series of sinusoidal terms, each with a unique magnitude, frequency, and phase.

This process, in effect, converts a waveform in the time domain that is difficult to describe mathematically into a more manageable series of sinusoidal functions that when added together, exactly reproduce the original waveform. Plotting the amplitude of each sinusoidal term versus its frequency creates a power spectrum, which is the response of the original waveform in the frequency domain. Fig. 10 illustrates this time to frequency domain conversion concept.

The Fourier transform has become a powerful analytical tool in diverse fields of science. In some cases, the Fourier transform can provide a means of solving unwieldy equations that describe dynamic responses to electricity, heat or light. In other cases, it can identify the regular contributions to a fluctuating signal, thereby helping to make sense of observations in astronomy, medicine, and chemistry. Perhaps because of its usefulness, the Fourier transform has been adapted for use on the personal computer. Algorithms have been developed to link the personal computer and its ability to evaluate large quantities of numbers with the Fourier transform to provide a personal computer-based solution to the representation of waveform data

Need help with the assignment?

Our professionals are ready to assist with any writing!

GET HELP

in the frequency domain.

The fast Fourier transform (FFT) is a computationally efficient method of generating a Fourier transform. The main advantage of an FFT is speed, which it gets by decreasing the number of calculations needed to analyze a waveform. A disadvantage associated with the FFT is the restricted range of waveform data that can be transformed and the need to apply a window weighting function to the waveform to compensate for spectral leakage.

The FFT is just a faster implementation of the DFT. The FFT algorithm reduces an n-point Fourier transform to about $(n/2) \log_2(n)$ complex multiplications. For example, calculated directly, a DFT on 1,024 (i.e., 2¹⁰) data points would require $n^2 = 1,024 \times 1,024 = 1,048,576$ multiplications. The FFT algorithm reduces this to about $(n/2) \log_2(n) = 512 \times 10 = 5,120$ multiplications, for a factor-of-200 improvement.

But the increase in speed comes at the cost of versatility. The FFT function automatically places some restrictions on the time series to be evaluated in order to generate a meaningful, accurate frequency response. Because the FFT function uses a base 2 logarithm by definition, it requires that the range or length of the time series to be evaluated contains a total number of data points precisely equal to a 2-to-the-nth-power number (e.g., 512, 1024, 2048, etc.). Therefore, with an FFT you can only evaluate a fixed length waveform containing 512 points, or 1024 points, or 2048 points, etc. For example, if your time series contains 1096 data points, you would only be able to evaluate 1024 of them at a time using an FFT since 1024 is the highest 2-to-the-nth-power that is less than 1096.

Because of this 2-to-the-nth-power limitation, an additional problem materializes. When a waveform is evaluated by an FFT, a section of the waveform becomes bounded to enclose 512 points, or 1024 points, etc. One of these boundaries also establishes a starting or reference point on the waveform that repeats after a definite interval, thus defining one complete cycle or period of the waveform. Any number of waveform periods and more importantly, partial waveform periods can exist between these boundaries. This is where the problem develops.

The FFT function also requires that the time series to be evaluated is a commensurate periodic function, or in other words, the time series must contain a whole number of periods as shown in Figure 2a to generate an accurate frequency response. Obviously, the chances of a waveform containing a number of points equal to a 2-to-the-nth-power number and ending on a whole number of periods are slim at best, so something must be done to ensure an accurate representation in the frequency domain.

The FFT is a computationally fast way to generate a power spectrum based on a 2-to-the-nth-power data point section of the waveform. This means that the number of points plotted in the

Need help with the assignment?

Our professionals are ready to assist with any writing!

[GET HELP](#)

power spectrum is not necessarily as many as was originally intended. The FFT also uses a window to minimize power spectrum distortion due to the end-point discontinuity. However, this window may attenuate important information appearing on the edges of the time series to be evaluated.

Acoustic Fingerprint

An acoustic fingerprint is a condensed digital summary, a fingerprint, deterministically generated from an audio signal, that can be used to identify an audio sample or quickly locate similar items in an audio database.

Practical uses of acoustic fingerprinting include identifying songs, melodies, tunes, or advertisements; sound effect library management; and video file identification. Media identification using acoustic fingerprints can be used to monitor the use of specific musical works and performances on the radio broadcast, records, CDs and peer-to-peer networks. This identification has been used in copyright compliance, licensing, and other monetization schemes.

A robust acoustic fingerprint algorithm must take into account the perceptual characteristics of the audio. If two files sound alike to the human ear, their acoustic fingerprints should match, even if their binary representations are quite different. Acoustic fingerprints are not hash functions, which must be sensitive to any small changes in the data. Acoustic fingerprints are more analogous to human fingerprints where small variations that are insignificant to the features the fingerprint uses are tolerated. One can imagine the case of a smeared human fingerprint impression which can accurately be matched to another fingerprint sample in a reference database; acoustic fingerprints work in a similar way.

Perceptual characteristics often exploited by audio fingerprints include average zero crossing rate, estimated tempo, average spectrum, spectral flatness, prominent tones across a set of frequency bands, and bandwidth.

Most audio compression techniques will make radical changes to the binary encoding of an audio file, without radically affecting the way it is perceived by the human ear. A robust acoustic fingerprint will allow a recording to be identified after it has gone through such compression, even if the audio quality has been reduced significantly. For use in radio broadcast monitoring, acoustic fingerprints should also be insensitive to analog transmission artifacts.

Generating a signature from the audio is essential for searching by sound. One common technique is creating a time-frequency graph called spectrogram.

Need help with the assignment?

Our professionals are ready to assist with any writing!

GET HELP

Any piece of audio can be translated to a spectrogram. Each piece of audio is split into some segments over time. In some cases adjacent segments share a common time boundary, in other cases, adjacent segments might overlap. The result is a graph that plots three dimensions of audio: frequency vs amplitude (intensity) vs time.

gradesfixer.com

Need help with the assignment?

Our professionals are ready to assist with any writing!

GET HELP