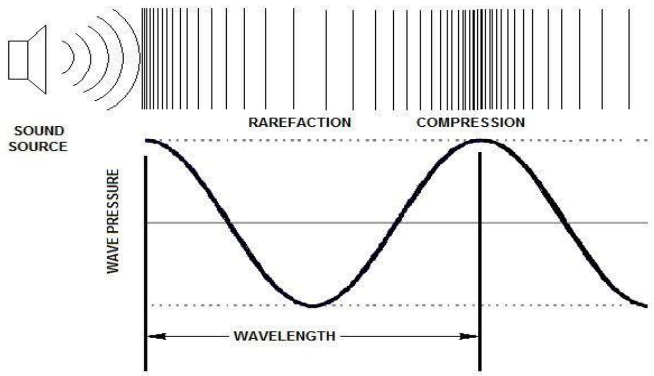
Audio Coding Basics

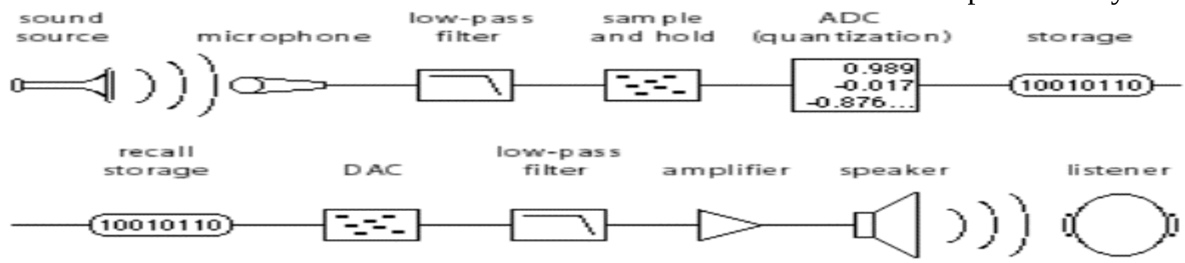
Fundamentals of Digital Audio

Introduction:

Something vibrates or pressurize that makes air Vibrate ,this processs vibrates the sensitive bones in the ear ,these sensations are registered in brain as a SOUND.Fast vibrations are high pitched and slow vibrations are low pitched.



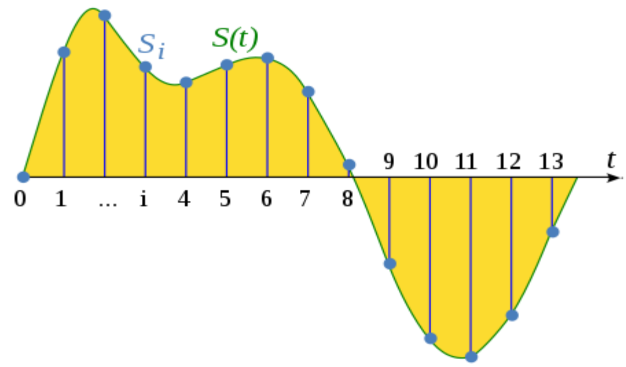
These sounds are anlog in nature. To convert analog sound to the digital realm means taking an analog signal and creating a representation of that signal in the language of computers, which is binary (zeroes and ones). An analog signal is continuous, meaning constantly changing in amplitude and time. Digital conversion requires that it be sampled or measured periodically to make it understandable and editable in a computer system.



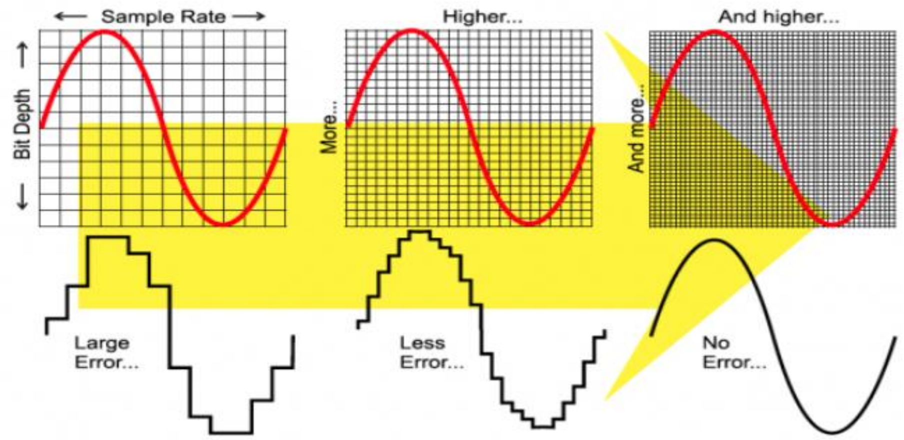
Digital Audio

Digital audio refers to the representation of sound in a digital format using binary code, which consists of 0s and 1s. This digital representation allows for accurate storage, transmission, and manipulation of audio signals. Here are the key concepts and components involved in digital audio:

- **Analog-to-Digital Conversion (ADC)**: The process of converting analog audio signals, which are continuous voltage variations, into digital data is called analog-to-digital conversion (ADC). ADC involves two main steps: sampling and quantization.
- **Sampling**: The continuous analog audio signal is sampled at regular intervals, capturing its amplitude at each sampling point. The rate at which samples are taken is called the sampling rate and is typically measured in Hertz (Hz).



-
- **Quantization**: Each sampled amplitude is then quantized, which means it's assigned a specific digital value. The number of bits used to represent each sample determines the quantization levels and ultimately affects the dynamic range and resolution of the digital audio signal.



Increasing bit depth resolution

-
- **Digital Representation**: After ADC, the analog audio signal has been converted into a stream of digital values. These values are usually represented in binary code, where each value is made up of a series of 0s and 1s.
- **Bit Depth**: Bit depth, often referred to as "resolution," determines the precision of each digital sample. Common bit depths include 16-bit and 24-bit. A higher bit depth allows for more accurate representation of audio levels, resulting in improved dynamic range and reduced quantization noise.
- **Sampling Rate**: The sampling rate defines how many samples are taken per second. Common sampling rates include 44.1 kHz (used in audio CDs), 48 kHz (used in DVDs and most digital audio recordings), and higher rates like 96 kHz or 192 kHz for high-resolution audio.
- **Nyquist Theorem**: According to the Nyquist theorem, the sampling rate must be at least twice the highest frequency present in the audio signal to avoid aliasing. This is why the standard sampling rate for audio CDs is 44.1 kHz, as it can accurately capture frequencies up to around 20 kHz, the upper limit of human hearing.
- **Digital Audio Formats**: Digital audio data is stored in various file formats, each with its own characteristics and compression methods. Some common formats include WAV (uncompressed), FLAC (lossless compression), and MP3 (lossy compression).
- **Digital-to-Analog Conversion (DAC)**: When playing back digital audio, the reverse process occurs. The digital data is converted back into analog audio signals through

digital-to-analog conversion (DAC). The DAC reconstructs the continuous voltage variations that represent the original sound.

- **Audio Interfaces**: Audio interfaces or sound cards connect computers and audio equipment. They facilitate the conversion between analog and digital signals during recording and playback. Some interfaces also offer features like multiple inputs and outputs, MIDI connectivity, and hardware processing.
- **Bitrate**: In compressed audio formats like MP3, bitrate represents the amount of data used to encode a specific amount of audio time. Higher bitrates generally result in better audio quality but require more storage space.

Audio compression

Audio compression refers to the process of reducing the size of audio files while attempting to maintain an acceptable level of audio quality. The goal of audio compression is to save storage space, facilitate faster data transmission, and optimize bandwidth usage for various applications such as streaming, storage, and broadcasting. There are two main types of audio compression: lossless compression and lossy compression.

Lossless Compression:
Lossless compression techniques aim to reduce the size of audio files without any loss of audio quality. This means that when the compressed audio is decompressed, it can be perfectly reconstructed to match the original uncompressed audio.
Lossless compression methods identify and remove redundancy in the audio data. This might include repeating patterns, silence, and other unnecessary information. Common lossless audio compression formats include FLAC (Free Lossless Audio Codec), ALAC (Apple Lossless Audio Codec), and WAV (Waveform Audio File Format).
Lossless compression is often used in scenarios where preserving the exact audio quality is crucial, such as archiving, professional audio production, and audio editing workflows.

Lossy Compression: Lossy compression techniques achieve higher levels of compression by discarding some audio data that is considered less perceptually significant to the human ear. This results in a smaller file size but introduces a certain degree of audio quality loss.
Lossy compression methods take advantage of psychoacoustic models, which analyze the characteristics of human hearing. These models identify frequencies and components that are less noticeable to humans and selectively remove or reduce them. The degree of compression and quality loss is typically controlled by adjusting the compression settings, including bitrate and quality parameters.
Common lossy audio compression formats include MP3 (MPEG Audio Layer III), AAC (Advanced Audio Codec), and OGG Vorbis. Lossy compression is often used for music streaming, online media, and situations where file size reduction is a higher priority than preserving every detail of audio quality.
There are two approches in lossy compression
1.silent compression.
2.companding.

Generic Audio Encoder

Pychacoustic Model:

A psychoacoustic model is a critical component of audio compression techniques, especially in lossy compression algorithms like MP3 and AAC. It involves understanding the characteristics of human auditory perception to selectively remove or reduce audio data that is less perceptually significant. By exploiting the limitations of human hearing, psychoacoustic models enable efficient compression while minimizing perceived audio quality loss.

Here's how psychoacoustic models work in audio compression:

Frequency Masking: The human auditory system has a phenomenon called frequency masking. When a loud sound (masker) is present at a certain frequency, it can make nearby quieter sounds (maskees) less audible. Psychoacoustic models identify such masking relationships and prioritize encoding the masker while reducing the bit allocation for the maskees. This results in efficient use of bits for encoding important audio components.

Temporal Masking: Temporal masking refers to the phenomenon where a loud sound can mask a quieter sound that occurs shortly before or after it in time. Psychoacoustic models consider temporal masking effects and allocate fewer bits to encode audio components occurring during temporal masking.

Threshold of Hearing: Human ears have a threshold below which sounds are inaudible. Psychoacoustic models use this threshold to discard or quantize audio components that fall below the threshold, as they are unlikely to be perceived by the listener.

Bit Allocation: Psychoacoustic models determine how many bits to allocate to different frequency bands or time segments of the audio signal. Critical or more audible parts receive more bits, while less critical parts receive fewer bits or are entirely removed.

Masking Curves: Psychoacoustic models create masking curves that represent the extent of masking caused by a loud sound on neighboring frequencies or time intervals. These curves guide the quantization process, helping decide the amount of compression to apply.

Bitrate Control: The psychoacoustic model's analysis assists in controlling the overall bitrate of the compressed audio. This ensures that the most important audio components are encoded with higher precision, while less significant components are encoded with fewer bits.

Perceptual Coding: Lossy audio compression algorithms use the insights from psychoacoustic models to perform perceptual coding. Perceptual coding involves quantizing audio samples in a way that minimizes the perceived quality loss. By removing less audible components, these algorithms achieve high compression ratios while maintaining acceptable audio quality.

MPEG layer3

IN lossy compressioon MPEG has 3 layers in which layer 3 (MP3) is more efficient.

Here are the sequential steps involved in the MP3 audio compression process:

- **Analysis of the Audio Signal**: The audio signal is first divided into smaller blocks or frames, usually consisting of a few hundred to a few thousand audio samples. Each frame typically spans about 26 milliseconds of audio.
- **Windowing and Pre-Processing**: The audio samples within each frame are multiplied by a window function to minimize the artifacts that occur at the frame boundaries due

to the abrupt start and end points. This process reduces distortion caused by the discontinuities.

- **Frequency Transformation**: The pre-processed audio samples in each frame are transformed from the time domain to the frequency domain using the Fast Fourier Transform (FFT) or similar algorithms. This results in a representation of the audio signal in terms of frequency components.
- **Psychoacoustic Model**: A psychoacoustic model analyzes the characteristics of the audio signal to determine which components are perceptually significant and which ones can be discarded or quantized with less precision. This involves considering properties like masking effects, tonal and non-tonal components, and human auditory perception.
- **Quantization**: Based on the psychoacoustic model's analysis, the transformed audio data is quantized, which means reducing the precision of less significant frequency components. Components that are perceptually masked or less audible are assigned fewer bits, allowing for more efficient compression.
- **Entropy Coding**: The quantized audio data is then encoded using entropy coding techniques such as Huffman coding or arithmetic coding. These techniques assign shorter codes to more frequent values and longer codes to less frequent values, further reducing the amount of data to be transmitted.
- **Frame Header and Side Information**: Each encoded frame is preceded by a header that contains essential information about the frame, including the bit rate, sampling rate, and more. Additionally, side information is included that describes the allocation of bits for different audio components within the frame.
- **Bitstream Generation**: The encoded frames, along with their headers and side information, are combined to form the final bitstream. This bitstream represents the compressed audio data that can be transmitted, stored, or played back.
- **Decoding Process**: To play back or decode the MP3 bitstream, the reverse process is applied. The bitstream is parsed, headers and side information are read, and the encoded audio data is decoded using entropy decoding techniques.
- **Inverse Quantization and Synthesis**: The quantized audio data is inverse-quantized, effectively restoring the precision lost during quantization. The resulting data is transformed back into the frequency domain.
- **Synthesis and Windowing**: The inverse transformed data is multiplied by the inverse of the window function applied during the pre-processing stage. This helps reconstruct the time-domain audio samples for each frame.
- **Overlap and Add**: Since adjacent frames overlap due to the windowing process, the reconstructed audio samples from each frame are combined using the overlap and add technique to produce a continuous audio signal.

By following these steps, MPEG-1 Audio Layer III (MP3) achieves high compression efficiency by intelligently discarding or approximating perceptually less significant audio components. This allows for smaller file sizes while maintaining acceptable audio quality, making MP3 a groundbreaking technology in the realm of digital audio distribution and playback.