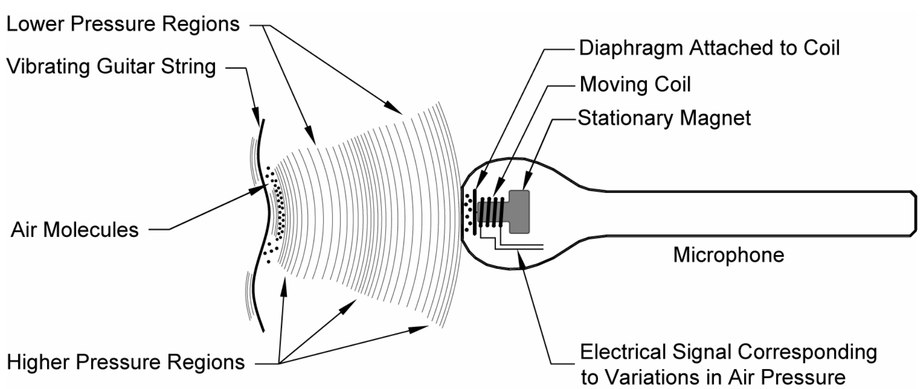
Chapter 11: A Digital Audio Primer

Many people don't care about the technology behind their stereo system. As long as it sounds good and they can press a button and listen to music, everything is fine. However, when you start working with audio on computers and the Internet, it's important to understand a few key principles to achieve good results.

What is Sound?

Sound reaches our ears as waves of rapidly varying air pressure caused by a vibrating object, such as a guitar string. As the string moves in one direction, it pushes on nearby air molecules, causing them to move closer together. This creates a small region of high pressure on one side of the string and low pressure on the opposite side. As the string moves in the opposite direction, the areas of high and low pressure reverse.

Figure 15 - Conversion of Sound Wave to Analog Signal



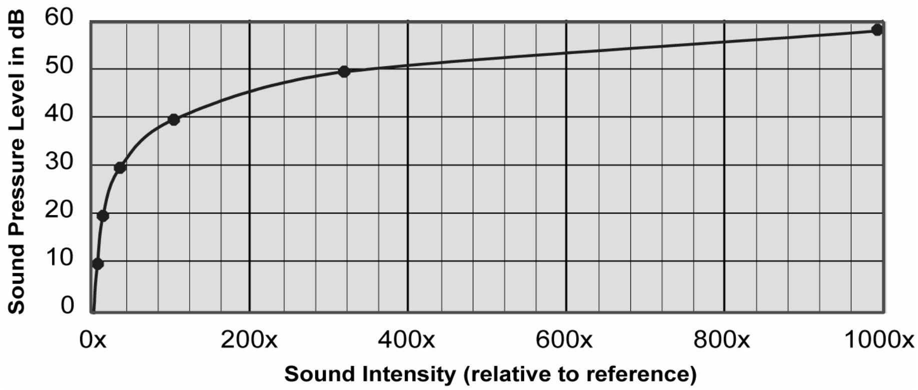
Sound waves occur as these repeating cycles of higher and lower pressure move out and away from the vibrating object. The frequency (pitch) of a sound is the number of times per second that these cycles occur. The amplitude (intensity) of sound is the size of the variations.

Measuring Sound

Our ears respond to sound logarithmically. As a sound gets louder, increasingly larger changes in sound intensity must occur for us to perceive the same amount of change in loudness.

Decibels

Figure 16 - Relationship of Sound Pressure Level to Sound Intensity



The term decibel (dB) means one-tenth of a Bel—named after Alexander Graham Bell. (This is why the B in dB is capitalized). A Bel is the base 10 logarithm of the ratio between the power level of two sounds or signals.

Sound Pressure Level

The intensity of sound is called the sound pressure level (SPL) and is measured in decibels (dB SPL). Decibels are a logarithmic scale that represents how much a sound level or audio signal varies from another signal, or reference level. You might refer to a sound as being 10dB louder than another sound or 3dB softer. A 3dB change is about the minimum change in sound level that most of us can perceive. A 10dB change sounds about twice as loud.

Figure 17 - Relative Loudness of Common Sounds

Relative Level	SPL	Sound
10,000,000x	140	Colt 45 Pistol (25 ft)
	130	Fire Engine Siren (100 ft)
		Threshold of Pain
1,000,000x	120	Jet Takeoff (200 ft)
		Rock Concert (10 ft)
	110	
100,000x	100	
		Loud Classical Music
	90	Heavy Street Traffic (5 ft)
10,000x	80	Cabin of Cruising Jet Aircraft
	70	
		Average Conversation (3 ft)
1,000x	60	
	50	Average Suburban Home (night)
100x	40	Quiet Auditorium
	30	
		Quiet Whisper (5 ft)
10x	20	Rustling Leaves
	10	
Reference Level	0 Decibels	Threshold of Hearing

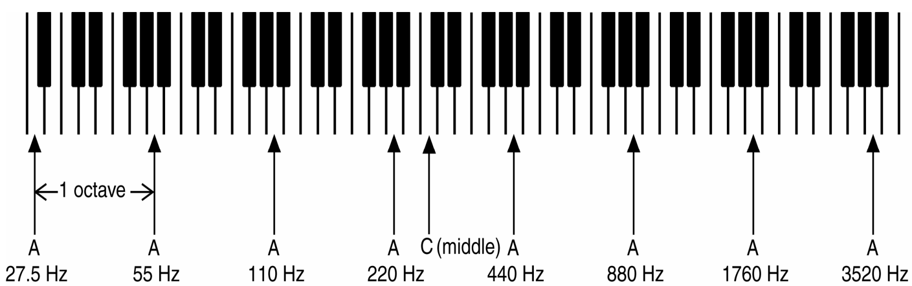
Decibels are always relative. To use decibels to represent a specific quantity, you need to know the reference, or 0 dB level. In the case of sound intensity, 0 dB SPL represents the threshold of hearing of a young undamaged ear (a pressure of about 3 billionths of a pound per square inch). In this case, all sound pressure levels are positive numbers that show how much louder a sound is than the threshold of hearing.

Loudness

Loudness is subjectively how we perceive different sound intensities. The sound intensity of a jet taking off 200 feet away is about 120dB SPL, or a million times more intense than the threshold of hearing. The sound intensity of rustling leaves is about 20dB SPL, or 10 times higher than the threshold of hearing. The sound of the jet is 100,000 times more intense than the rustling leaves (100dB). We actually perceive the jet to be about 1000 times louder than rustling leaves rather than 100,000 times louder.

Frequency

Figure 18 - Octave Intervals and Frequencies for Musical Notes



The frequency of a sound is measured in Hertz (Hz), which means cycles per second. A kilohertz (kHz) is a thousand cycles per second. We perceive pitch exponentially. A unit of pitch all musicians are familiar with is the octave. An octave is the interval between any note and the next higher note with the same name. Notes that are one octave apart sound similar, but one is twice the frequency of the other. For example, the note A below middle C is at a frequency of 220Hz, the note A above middle C is at 440Hz, and the next higher A is at 880Hz.

Analog Audio

The term analog means something that is similar in function or position. The varying voltage produced by a microphone is analogous to the pressure variations of a sound wave. On a cassette tape, variations in magnetic flux in a metal coating on the tape represent pressure variations in the sound wave. On vinyl records, variations in the width of the groove correspond to the pressure variations. The position along the groove or tape corresponds to time.

In an analog audio system, voltages represent sound pressures. These signals are amplified from the millivolt level (1000^{th} of a volt) produced by microphones, playback heads and phono cartridges by about 1000 times (60dB) to the levels found inside stereo preamps. A power amp boosts the voltage level from the preamp to a loudspeaker, which creates sound waves in the air by vibrating rapidly in response to the audio signal.

Digital Audio

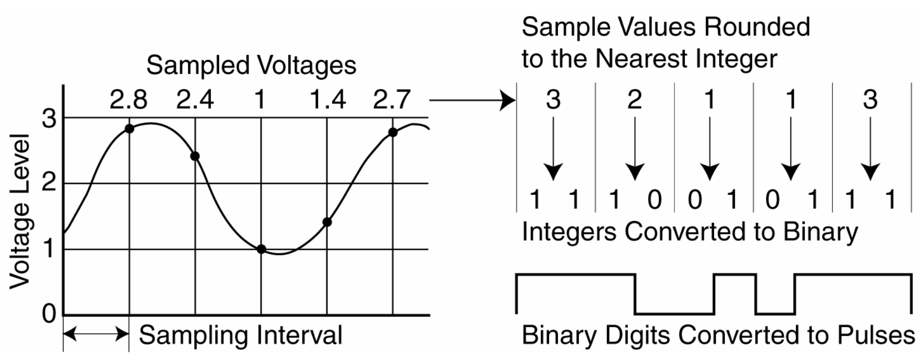
In digital audio, the representation of the audio signal is no longer directly analogous to the sound wave. Instead, the value of the signal is sampled at regular intervals by an analog-to-digital (A/D) converter (or ADC), which produces numbers (digits) that represent the value of each sample. This stream of numbers represents a digital audio signal, which can be stored as a computer file and transmitted across a network.

In order to listen to a digital audio signal, it must be converted to analog by a digital-to-analog (D/A) converter (or DAC). In most home stereo systems, the D/A conversion takes place inside the CD player. Computer sound cards, MiniDisc recorders and DATs have both A/D converters (for recording) and D/A converters (for playback). Many home systems have a combination of digital and analog components, but all audio systems end with analog signals at the speakers or headphones.

Sampling

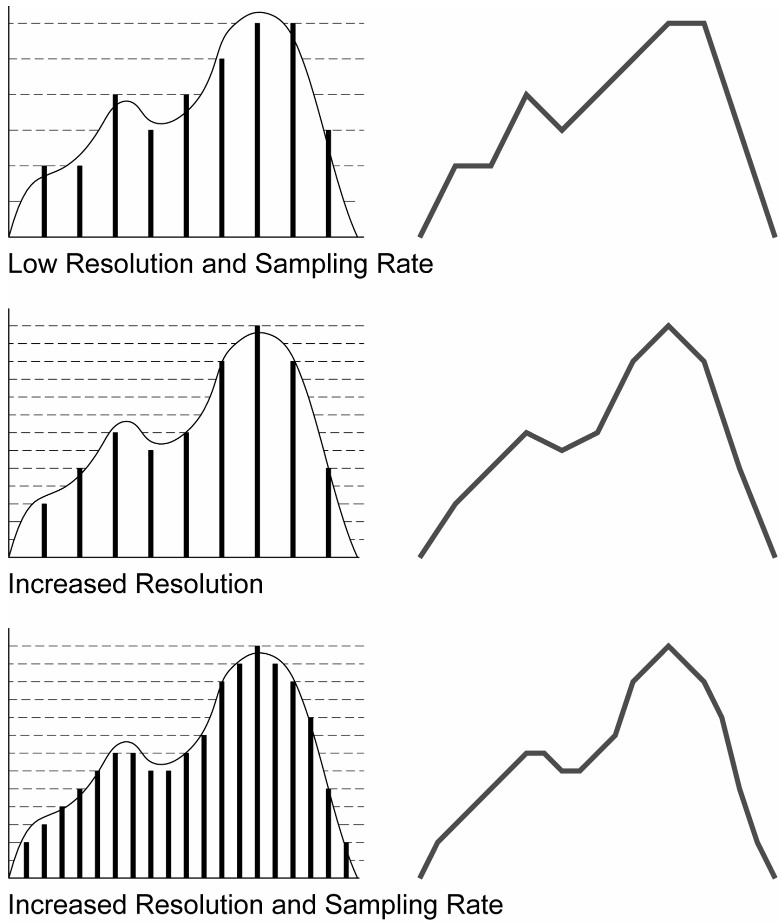
To convert an analog signal to a digital format, the voltage is sampled at regular intervals, thousands of times per second. The value of each sample is rounded to the nearest integer on a scale that varies according to the resolution of the signal. The integers are then converted to binary numbers.

Figure 19 - Sampling and Converting a Waveform to PCM



The sampling rate is how many times per second the voltage of the analog signal is measured. CD audio is sampled at a rate of 44,100 times per second (44.1 kHz). DAT (Digital Audio Tape) supports sampling rates of 32, 44.1 and 48 kHz. Other commonly used sampling rates are 22.05 kHz and 11.025 kHz.

Figure 20 - Effect of Increased Resolution and Sampling Rates



The sampling rate must be at least twice as high as the highest frequency to be reproduced[1][1]. The range of human hearing is roughly from 20 to 20,000 Hz, so a sampling rate of at least 40 kHz is needed to reproduce the full range.

Higher sampling rates allow the use of filters with a more gradual roll-off. This reduces phase shift, which can affect the stereo image at higher frequencies.

The 44.1 kHz sampling rate for CDs was chosen to allow headroom for filters and other types of signal processing. MPEG AAC and DVD Audio support rates up to 96 kHz.

Resolution

The resolution of a digital signal is the range of numbers that can be assigned to each sample. CD audio uses 16 bits, which provides a range of binary values from 0 to 65,534 (2^{16}). The binary value of 0000000000000000 (zero) corresponds to -32,768 (the lowest possible level), and the value 1111111111111111 (65,535) corresponds to 32,767 (the highest possible level). Higher resolution increases the dynamic range and reduces quantization distortion and background noise.

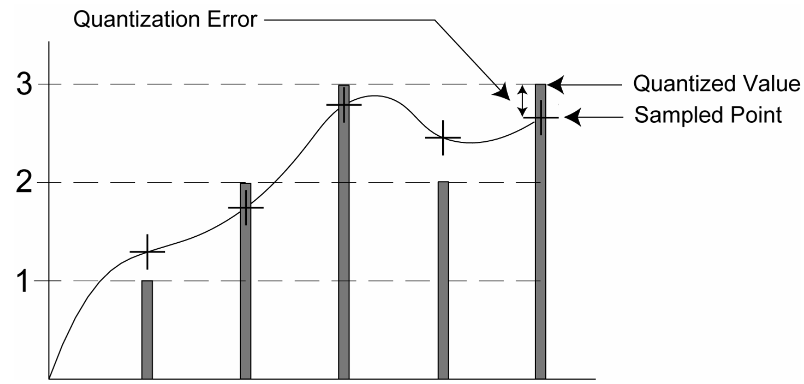
Quantization

Quantization is the process of selecting whole numbers to represent the voltage level of each sample. The A/D converter must select a whole number that is closest to the signal level at the instant it's sampled. This produces small rounding errors that cause distortion.

Quantization distortion increases at lower levels because the signal is using a smaller portion of the available dynamic range, so any errors are a greater percentage of the signal. A key advantage of audio encoding schemes, such as MP3, is that more bits can be allocated to low-level signals to reduce quantization errors.

Dithering

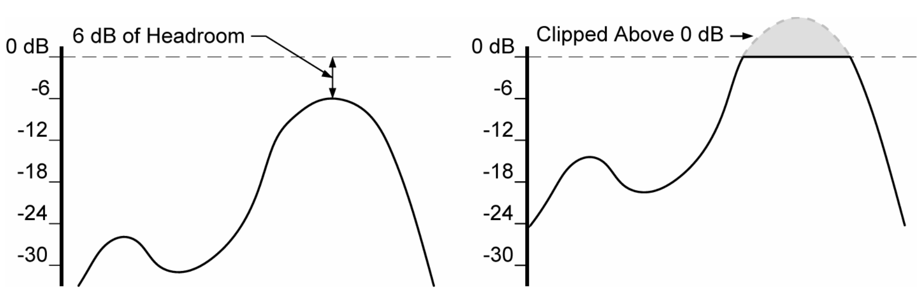
Figure 21 - Quantization Errors



A process called dithering introduces random noise into the signal to spread out the effects of quantization distortion and make it less noticeable. Some audiophiles don't like the notion of noise that is deliberately added to a signal, but the advantages of digital audio are so great that the end result is still better than most analog systems.

Clipping

Figure 22 - Clipping



Levels in a digital audio signal are usually expressed in dB, measured by their relationship to 0 dB, the highest possible level. One of the rules of digital audio is that a signal can never exceed 0 dB. If the level of a signal is raised too much, the peaks will be clipped at the 0 dB level. Clipping causes extreme distortion and should be avoided at all costs.

Bit-rates

The term "bit-rate" refers to how many bits (1s and 0s) are used each second to represent the signal. The bit-rate for digital audio is expressed in thousands of bits per second (kbps) and correlates directly to the file size and sound quality. Lower bit-rates result in smaller file sizes but poorer sound quality, and higher bit-rates result in better quality but larger files.

The bit-rate of uncompressed audio can be calculated by multiplying the sampling rate by the resolution (8-bit, 16-bit, etc.) and the number of channels. For example, CD Audio (or a WAV file extracted from a CD) has a sampling rate of 44,100 times per second, a resolution of 16 bits and two channels. The bit-rate would be approximately 1.4 million bits per second (1,411 kbps).

Table 1 - Calculating Bit-rates

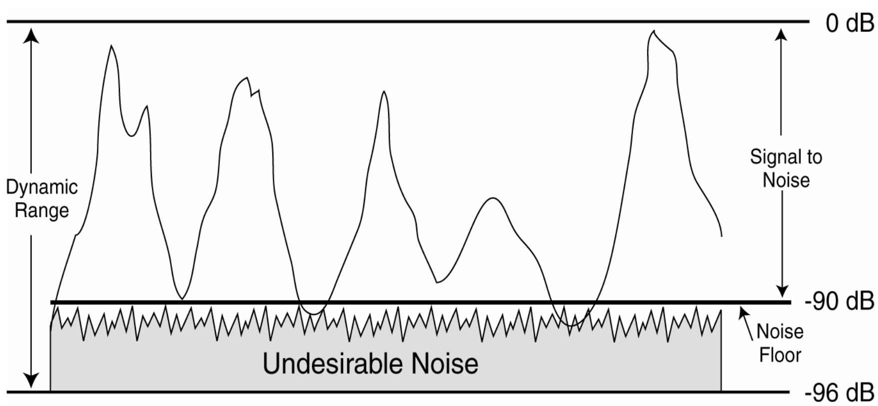
Sampling Rate	x	Resolution	x	# of Channels	=	Bit-rate
44,100	x	16	x	2	=	1,411,200

Dynamic Range

Dynamic range is the range of the lowest to the highest level that can be reproduced by a system. Digital audio at 16-bit resolution has a theoretical dynamic range of 96 dB, but the actual dynamic range is usually lower because of overhead from filters. The dynamic range of vinyl records and cassette tapes is much lower than CDs and varies depending on the quality of the recording and playback equipment. The dynamic range of cassette tapes also varies depending on the type of tape.

Signal-to-noise Ratio

Figure 23 - Dynamic Range and Signal-to-noise Ratio



The signal-to-noise ratio is the ratio of the background noise (hiss, hum and static) level to the highest level that can be reproduced. Each additional bit of resolution corresponds to an increase of 6 dB in signal-to-noise ratio. Audio CDs achieve about a 90 dB signal-to-noise ratio.

Encoding

Encoding is the process of converting uncompressed digital audio to a compressed format such as MP3. The algorithm used in the encoding (and decoding) software is referred to as a codec—as in *cod*ing/*dec*oding. There is often more than one codec for a particular format, and different codecs can vary widely in quality and speed, even for the same format.

Advantages of Digital Audio

For years, audiophiles and engineers have debated the merits of digital audio versus high-end analog systems, and to this day, there are audiophiles who swear by their analog systems. Digital audio has emerged as the winner by most accounts, but it's still useful to understand the advantages of digital versus analog audio, because many audio systems contain a mix of digital and analog components.

The advantages of digital audio can be summed up as follows: wider dynamic range, increased resistance to noise, better copyability and the ability to use error correction to compensate for wear and tear. Many types of digital media, such as CDs and MiniDiscs, are also more durable than common analog media, such as vinyl records and cassette tapes.

Wider Dynamic Range

Digital audio at 16 bits theoretically can achieve a dynamic range of 96 dB, compared to less than 80 dB for the best analog systems. This is especially important for classical music where levels within the same composition can range from the relative quiet of a flute solo to the loudness of dozens of instruments playing simultaneously,.

Increased Resistance to Noise

In analog systems, crackling noise and hum from electromagnetic frequency (EMF) interference is picked up along the way as the signal passes through analog circuits. Background hiss is also generated by thermal noise from analog components. Digital signals are virtually immune to picking up these types of noise, although any noise that enters the signal before it's converted to digital will be reproduced along with the rest of the signal.

Better Copyability

Digital audio can be copied from one digital device to another without any loss of information, unlike analog recording, where information is lost and noise introduced with every copy. Even the best analog systems lose about 3dB of signal-to-noise ratio when a copy is recorded. After several generations of analog copies, the sound quality will deteriorate noticeably. With digital audio, unlimited generations of perfect copies can be made.

This ability to make perfect copies is one reason why the RIAA has gone to so much trouble to introduce the Serial Copy Management System (SCMS) for consumer audio equipment, and why they are so concerned about the proliferation of MP3 files. SCMS prevents multiple generations of copies (copies of copies) from an original and is required by the Audio Home Recording Act of 1992 to be used on all consumer digital audio recording devices sold in the United States. Currently, there is no way to prevent multiple generations of perfect copies from a single MP3 file.

Digital copies can also be made much faster than analog copies, which usually must be made in real time. For example, with an analog device like a cassette deck, it always takes at least 60 minutes to record 60 minutes of music from a CD. With digital audio, the same 60 minutes of music can be copied to a hard disk in as little as 5 minutes on a system with a fast CD-ROM drive.

Of course, if you are making an original recording with digital equipment, it will take the same amount of time as with analog equipment. (Uncle Jack playing the kazoo for half an hour still takes half an hour to record). But once a digital recording is on your PC, you can make a digital copy in a fraction of the time it would take to record a copy with analog equipment.

Error Correction

Most digital audio media, such as CDs and DATs, have built-in error correction. On an audio CD, approximately 25% of the disc is used for error correction data. If a bad scratch causes an error that can't be corrected, the player will attempt to reconstruct the missing data by interpolation.

Durability

Digital media such as CDs and MiniDiscs are much more durable than any analog media. This improved durability is one of the main reasons people were so eager to migrate from vinyl records to CDs.

Each time you play a record or tape, microscopic bits of vinyl or oxide are scraped away, adding to the cumulative wear. Vinyl records are particularly prone to warping and scratching, and tapes gradually become demagnetized. A CD or MiniDisc can be played hundreds of times, with no loss of quality, as long as there is not excessive physical damage.

Both analog and digital tapes can suffer degradation from magnetic fields, but some popular digital formats like DAT are much more durable than analog tapes (especially cassettes) because the tape is stronger and the oxide coating is thicker.

File Size and Bandwidth

Digital audio can create large files that quickly use up hard disk capacity and require a tremendous amount of bandwidth to transmit over a network. Network bandwidth is like a pipe that carries a stream of bits. The size of the pipe imposes a limit on how many bits can be moved in a given time period. Multiple users competing for the same bandwidth limit the amount of bandwidth available to any one user.

File sizes and bandwidth requirements for uncompressed audio can be calculated by multiplying the sampling rate by the resolution, the number of channels and the time in seconds. The bit-rate has a direct relation to the file size—if you do something that changes the bit-rate, the file size will change proportionally. The bandwidth requirement of a digital audio signal is the same as the bit-rate. This is true whether the signal is compressed or not. Table 2 shows the formula for calculating file sizes for uncompressed audio.

Table 2 - Calculating File Sizes

Sampling Rate	x	Resolution	x	Number of Channels	x	Time in Seconds	/	Bits / Byte	=	File Size (in Bytes)
44,100	x	16	x	2	x	60	/	8	=	10,584,000

You can do several things to control the size of digital audio files, but there will always be a trade-off between file size and sound quality. Lowering the sampling rate will produce a smaller file, but will also lower the maximum frequency response. Lowering the resolution produces a smaller file but reduces the accuracy and allows more noise and distortion to be introduced due to increased quantization errors. A mono signal, used in place of stereo, will cut the size in half (uncompressed audio only).

Table 3 shows how different combinations of sampling rates, resolution and numbers of channels can be used to control file sizes.

Table 3 - File Sizes for a One-minute Audio Clip

Sampling Rate	Resolution	Number of Channels	Bit-rate	File Size (in Bytes)
44,100	16	2	1,411,200	10,584,000
44,100	16	1	705,600	5,292,000
22.050	16	1	352,800	2,646,000
11.025	16	1	176,400	1,323,000
11.025	8	1	88,200	616,000

Compression

Limited network bandwidth and hard disk capacity have been major driving factors behind the development of compressed audio formats. Until recently, only a small number of people used their computers to store CD-quality music. A few people would copy their favorite songs from a music CD and use a CD-Recordable drive to create a compilation CD, similar to the way many people make cassette tapes from prerecorded music.

Audio and electronics engineers have been working to solve the bandwidth bottleneck ever since networks were invented. They work on both sides of the problems by increasing bandwidth (larger pipe) and compressing data (higher pressure). High speed Internet connections such as cable modems and ASDL have been developed to increase the size of the pipe, and compression schemes such as JPEG and MPEG have been developed to squeeze more data through it.

MP3 provides relief by compressing files up to approximately 10=1 without significant loss of quality. Four minutes of CD audio (44.1, kHz 16-bit stereo) requires about 40MB of disk space and would take more than 3-½ hours to download with a 28.8 kbps modem. At this rate, a 2GB hard disk would hold about 50 four-minute songs.

With MP3 encoded at 128 kbps, each four-minute song would take up less than 4MB of space and could be downloaded in less than 20 minutes with a 28.8 kbps modem. A 2GB hard disk could now hold more than 500 songs. This much compression, coupled with the larger and cheaper hard disks that are now available, makes it possible to use a PC as a high-capacity, CD-quality jukebox in place of tape decks, turntables and CD players.

Table 4 - Typical Download Times* for Four-minute Songs

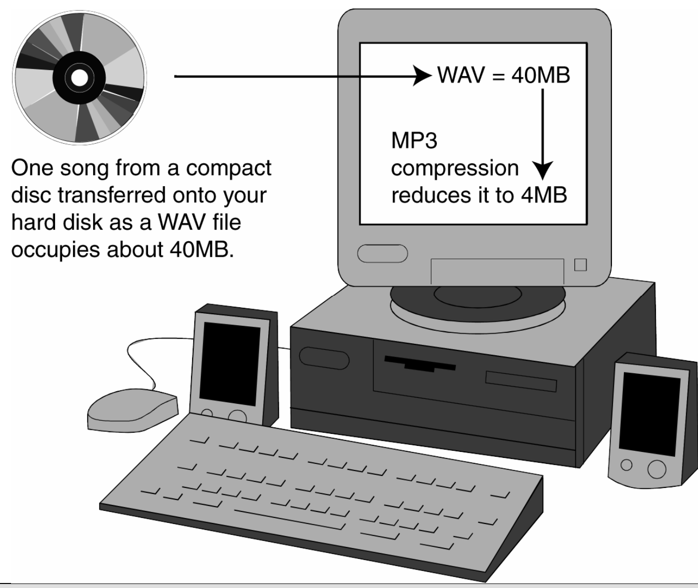
Format	28.8 k Modem	56 k Modem	Dual ISDN 128 kbps	Cable 1.5 Mbps	T1 Line 1.5 Mbps	ADSL 500 kbps+
CD Audio	3.6 hrs	2 hrs	44 min	4 min	4 min	7 min
MP3 at 128 kbps	19.7 min	9 min	4 min	20 sec	20 sec	39 sec

* Actual speed will usually be less.

Newer generations of MPEG Audio, such as AAC (Advanced Audio Coding), offer even higher levels of compression and better sound quality but have not yet reached the consumer market because of high licensing costs.

Lossy vs. Lossless Compression

Figure 24 - Typical MP3 Compression



> **Dynamic Range Compression**
>
> Dynamic range compression reduces the range in dB between the lowest and highest levels of a signal, but does not affect the file size or bandwidth requirement. Dynamic range compression is often used by recording engineers to make songs sound louder without clipping

There are two basic categories of compression: lossless and lossy. Lossless compression works by encoding repetitive pieces of information with symbols and equations that take up less space but provide all the information needed to reconstruct an exact copy of the original. Lossy compression works by discarding unnecessary and redundant information (sounds that most people can't hear) and then applying lossless compression techniques for further size reduction.

There is an ongoing debate among audiophiles about the merits of lossless versus lossy compression. With lossless compression, there is never a loss of fidelity (unless an error gets introduced during the process)—there is no debate about that. With lossy compression (such as MPEG Audio), there is always some loss of fidelity that becomes more noticeable as the compression ratio is increased. The goal then becomes producing sound where the losses are not noticeable, or noticeable but not annoying.

The highest compression ration for lossless audio is about 2 to 1, but the quality will always be indistinguishable from the original. With lossy compression, the quality will vary according to factors such as the bit-rate, the complexity of the music and the quality of the encoding software. Some forms of lossy compression, such as MPEG AAC, can achieve compression ratios of up to 11 to 1, with quality indistinguishable from the original. Numerous controlled tests with trained listeners have verified this.

Even with the best lossy formats, a few people with very sensitive ears may be able to tell the difference between the original and encoded file when listening to critical material (complex music) on expensive hi-fi systems. Most people will not be able to detect any differences at the higher bit-rates, but a few people will always feel like they are being cheated when they know something has been taken away (even if they can't tell the difference).

---

[1] According to the Nyquist Theorem

Sourced from..

The MP3 and Internet Audio Handbook:
Your Guide to the Digital Music Revolution!

March 2000, TeamCom Books
ISBN# 1-928791-10-7

(Out of Print)