24/192 Music Downloads

...and why they make no sense

- **Xiph homepage**
- **Planet.Xiph**
- **Monty's demos**

Also see Xiph.Org's new video, Digital Show & Tell, for detailed demonstrations of digital sampling in action on real equipment!

Articles last month revealed that musician Neil Young and Apple's Steve Jobs discussed offering digital music downloads of 'uncompromised studio quality'. Much of the press and user commentary was particularly enthusiastic about the prospect of uncompressed 24 bit 192kHz downloads. 24/192 featured prominently in my own conversations with Mr. Young's group several months ago.

Unfortunately, there is no point to distributing music in 24-bit/192kHz format. Its playback fidelity is slightly inferior to 16/44.1 or 16/48, and it takes up 6 times the space.

There are a few real problems with the audio quality and 'experience' of digitally distributed music today. 24/192 solves none of them. While everyone fixates on 24/192 as a magic bullet, we're not going to see any actual improvement.

First, the bad news

In the past few weeks, I've had conversations with intelligent, scientifically minded individuals who believe in 24/192 downloads and want to know how anyone could possibly disagree. They asked good questions that deserve detailed answers.

I was also interested in what motivated high-rate digital audio advocacy. Responses indicate that few people understand basic signal theory or the sampling theorem, which is hardly surprising. Misunderstandings of the mathematics, technology, and physiology arose in most of the conversations, often asserted by professionals who otherwise possessed significant audio expertise. Some even argued that the sampling theorem doesn't really explain how digital audio actually works [1].

Misinformation and superstition only serve charlatans. So, let's cover some of the basics of why 24/192 distribution makes no sense before suggesting some improvements that actually do.

Gentlemen, meet your ears

The ear hears via hair cells that sit on the resonant basilar membrane in the cochlea. Each hair cell is effectively tuned to a narrow frequency band determined by its position on the membrane. Sensitivity peaks in the middle of the band and falls off to either side in a lopsided cone shape overlapping the bands of other nearby hair cells. A sound is inaudible if there are no hair cells tuned to hear it.
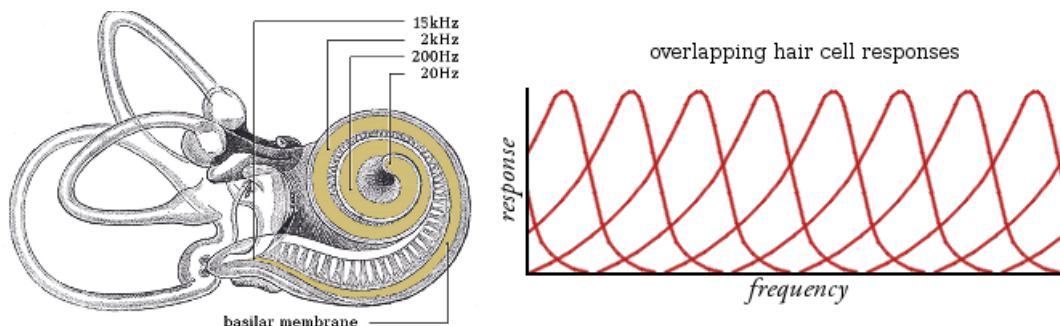
Above left: anatomical cutaway drawing of a human cochlea with the basilar membrane colored in beige. The membrane is tuned to resonate at different frequencies along its length, with higher frequencies near the base and lower frequencies at the apex. Approximate locations of several frequencies are marked.

Above right: schematic diagram representing hair cell response along the basilar membrane as a bank of overlapping filters.


This is similar to an analog radio that picks up the frequency of a strong station near where the tuner is actually set. The farther off the station's frequency is, the weaker and more distorted it gets until it disappears completely, no matter how strong. There is an upper (and lower) audible frequency limit, past which the sensitivity of the last hair cells drops to zero, and hearing ends.

Sampling rate and the audible spectrum

I'm sure you've heard this many, many times: The human hearing range spans 20Hz to 20kHz. It's important to know how researchers arrive at those specific numbers.

First, we measure the 'absolute threshold of hearing' across the entire audio range for a group of listeners. This gives us a curve representing the very quietest sound the human ear can perceive for any given frequency as measured in ideal circumstances on healthy ears. Anechoic surroundings, precision calibrated playback equipment, and rigorous statistical analysis are the easy part. Ears and auditory concentration both fatigue quickly, so testing must be done when a listener is fresh. That means lots of breaks and pauses. Testing takes anywhere from many hours to many days depending on the methodology.

Then we collect data for the opposite extreme, the 'threshold of pain'. This is the point where the audio amplitude is so high that the ear's physical and neural hardware is not only completely overwhelmed by the input, but experiences physical pain. Collecting this data is trickier. You don't want to permanently damage anyone's hearing in the process.
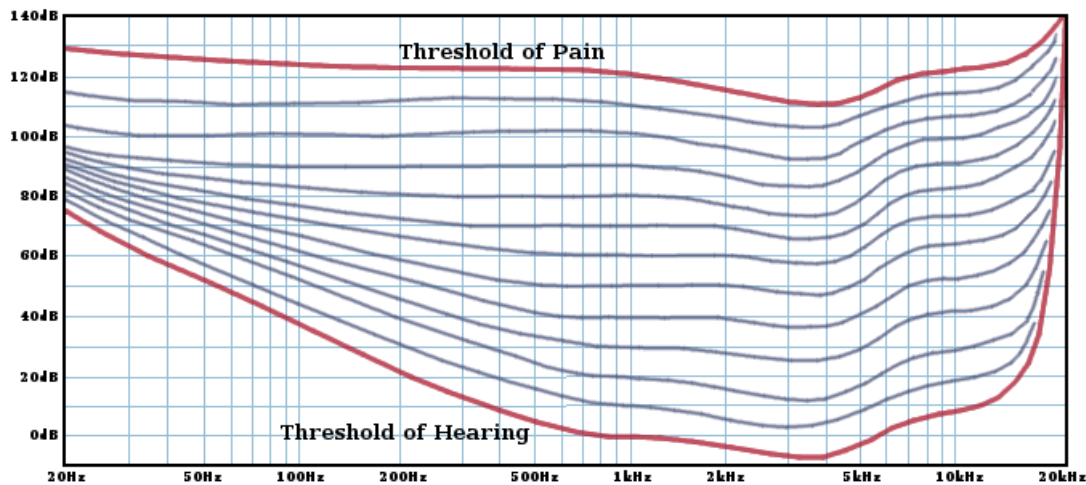
Above: Approximate equal loudness curves derived from Fletcher and Munson (1933) plus modern sources for frequencies > 16kHz. The absolute threshold of hearing and threshold of pain curves are marked in red. Subsequent researchers refined these readings, culminating in the Phon scale and the ISO 226 standard equal loudness curves. Modern data indicates that the ear is significantly less sensitive to low frequencies than Fletcher and Munson's results.

The upper limit of the human audio range is defined to be where the absolute threshold of hearing curve crosses the threshold of pain. To even faintly perceive the audio at that point (or beyond), it must simultaneously be unbearably loud.

At low frequencies, the cochlea works like a bass reflex cabinet. The *helicotrema* is an opening at the apex of the basilar membrane that acts as a port tuned to somewhere between 40Hz to 65Hz depending on the individual. Response rolls off steeply below this frequency.

Thus, 20Hz - 20kHz is a generous range. It thoroughly covers the audible spectrum, an assertion backed by nearly a century of experimental data.

Genetic gifts and golden ears

Based on my correspondences, many people believe in individuals with extraordinary gifts of hearing. Do such 'golden ears' really exist?

It depends on what you call a golden ear.

Young, healthy ears hear better than old or damaged ears. Some people are exceptionally well trained to hear nuances in sound and music most people don't even know exist. There was a time in the 1990s when I could identify every major mp3 encoder by sound (back when they were all pretty bad), and could demonstrate this reliably in double-blind testing [2].

When healthy ears combine with highly trained discrimination abilities, I would call that person a golden ear. Even so, below-average hearing can also be trained to notice details that escape untrained listeners. Golden ears are more about training than hearing beyond the physical ability of average mortals.

Auditory researchers would love to find, test, and document individuals with truly exceptional hearing, such as a greatly extended hearing range. Normal people are nice and all, but everyone wants to find a genetic freak for a really juicy paper. We haven't found any such people in the past 100 years of testing, so they probably don't exist. Sorry. We'll keep looking.

Spectrophiles

Perhaps you're skeptical about everything I've just written; it certainly goes against most marketing material. Instead, let's consider a hypothetical Wide Spectrum Video craze that doesn't carry preexisting audiophile baggage.
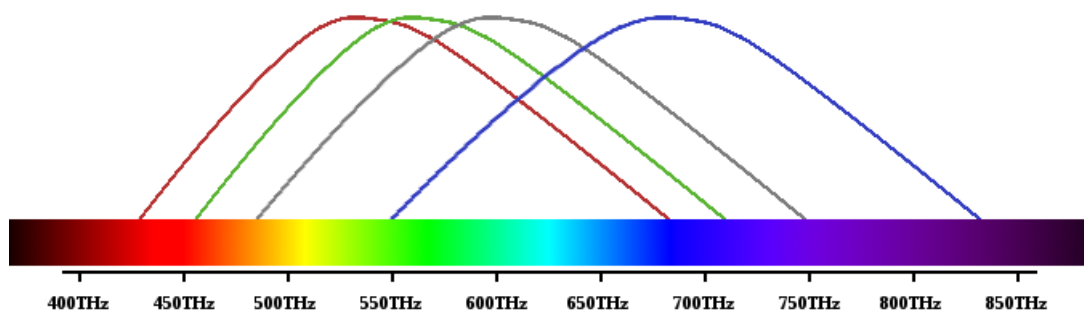
Above: The approximate log scale response of the human eye's rods and cones, superimposed on the visible spectrum. These sensory organs respond to light in overlapping spectral bands, just as the ear's hair cells are tuned to respond to overlapping bands of sound frequencies.

The human eye sees a limited range of frequencies of light, aka, the visible spectrum. This is directly analogous to the audible spectrum of sound waves. Like the ear, the eye has sensory cells (rods and cones) that detect light in

different but overlapping frequency bands.

The visible spectrum extends from about 400THz (deep red) to 850THz (deep violet) [3]. Perception falls off steeply at the edges. Beyond these approximate limits, the light power needed for the slightest perception can fry your retinas. Thus, this is a generous span even for young, healthy, genetically gifted individuals, analogous to the generous limits of the audible spectrum.

In our hypothetical Wide Spectrum Video craze, consider a fervent group of Spectrophiles who believe these limits aren't generous enough. They propose that video represent not only the visible spectrum, but also infrared and ultraviolet. Continuing the comparison, there's an even more hardcore [and proud of it!] faction that insists this expanded range is yet insufficient, and that video feels so much more natural when it also includes microwaves and some of the X-ray spectrum. To a Golden Eye, they insist, the difference is night and day!

Of course this is ludicrous.

No one can see X-rays (or infrared, or ultraviolet, or microwaves). It doesn't matter how much a person believes he can. Retinas simply don't have the sensory hardware.

Here's an experiment anyone can do: Go get your Apple IR remote. The LED emits at 980nm, or about 306THz, in the near-IR spectrum. This is not far outside of the visible range. Take the remote into the basement, or the darkest room in your house, in the middle of the night, with the lights off. Let your eyes adjust to the blackness.
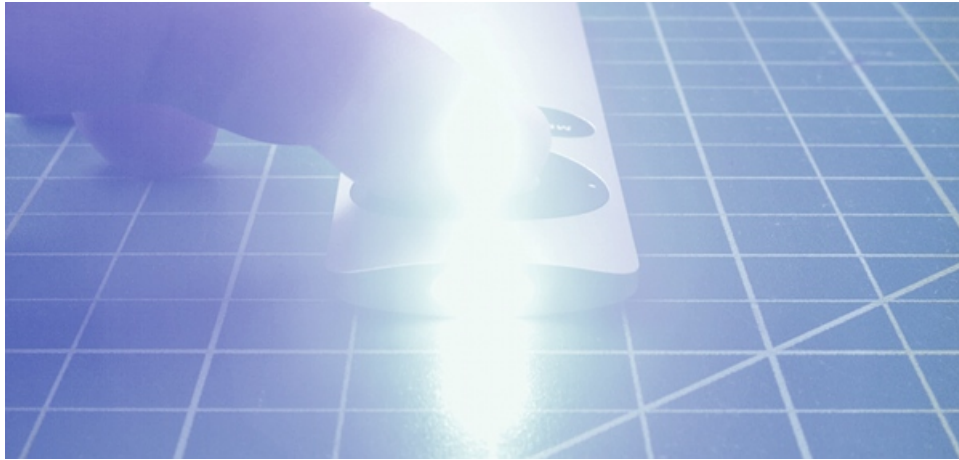
Above: Apple IR remote photographed using a digital camera. Though the emitter is quite bright and the frequency emitted is not far past the red portion of the visible spectrum, it's completely invisible to the eye.

Can you see the Apple Remote's LED flash when you press a button [4]? No? Not even the tiniest amount? Try a few other IR remotes; many use an IR wavelength a bit closer to the visible band, around 310-350THz. You won't be able to see them either. The rest emit right at the edge of visibility from 350-380 THz and may be just barely visible in complete blackness with dark-adjusted eyes [5]. All would be blindingly, painfully bright if they were well inside the visible spectrum.

These near-IR LEDs emit from the visible boundry to at most 20% beyond the visible frequency limit. 192kHz audio extends to 400% of the audible limit. Lest I be accused of comparing apples and oranges, auditory and visual perception drop off similarly toward the edges.

192kHz considered harmful

192kHz digital music files offer no benefits. They're not quite neutral either; practical fidelity is slightly worse. The ultrasonics are a liability during playback.

Neither audio transducers nor power amplifiers are free of distortion, and distortion tends to increase rapidly at the lowest and highest frequencies. If the same transducer reproduces ultrasonics along with audible content, any nonlinearity will shift some of the ultrasonic content down into the audible range as an uncontrolled spray of intermodulation distortion products covering the entire audible spectrum. Nonlinearity in a power amplifier will produce the same effect. The effect is very slight, but listening tests have confirmed that both effects can be audible.
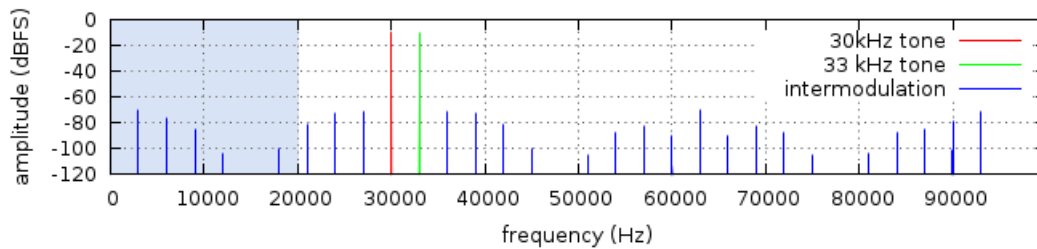
Above: Illustration of distortion products resulting from intermodulation of a 30kHz and a 33kHz tone in a theoretical amplifier with a nonvarying total harmonic distortion (THD) of about .09%. Distortion products appear throughout the spectrum, including at frequencies lower than either tone.

Inaudible ultrasonics contribute to intermodulation distortion in the audible range (light blue area). Systems not designed to reproduce ultrasonics typically have much higher levels of distortion above 20kHz, further contributing to intermodulation. Widening a design's frequency range to account for ultrasonics requires compromises that decrease noise and distortion performance within the audible spectrum. Either way, unneccessary reproduction of ultrasonic content diminishes performance.

There are a few ways to avoid the extra distortion:

1. A dedicated ultrasonic-only speaker, amplifier, and crossover stage to separate and independently reproduce the ultrasonics you can't hear, just so they don't mess up the sounds you can.

2. Amplifiers and transducers designed for wider frequency reproduction, so ultrasonics don't cause audible intermodulation. Given equal expense and complexity, this additional frequency range must come at the cost of some performance reduction in the audible portion of the spectrum.

3. Speakers and amplifiers carefully designed not to reproduce ultrasonics anyway.

4. Not encoding such a wide frequency range to begin with. You can't and won't have ultrasonic intermodulation distortion in the audible band if there's no ultrasonic content.

They all amount to the same thing, but only 4) makes any sense.

If you're curious about the performance of your own system, the following samples contain a 30kHz and a 33kHz tone in a 24/96 WAV file, a longer version in a FLAC, some tri-tone warbles, and a normal song clip shifted up by 24kHz so that it's entirely in the ultrasonic range from 24kHz to 46kHz:

- **Intermod Tests:**

    - 30kHz tone + 33kHz tone (24 bit / 96kHz) [5 second WAV] [30 second FLAC]
    - 26kHz - 48kHz warbling tones (24 bit / 96kHz) [10 second WAV]
    - 26kHz - 96kHz warbling tones (24 bit / 192kHz) [10 second WAV]
    - Song clip shifted up by 24kHz (24 bit / 96kHz WAV) [10 second WAV]
   (original version of above clip) (16 bit / 44.1kHz WAV)

Assuming your system is actually capable of full 96kHz playback [6], the above files should be completely silent with

no audible noises, tones, whistles, clicks, or other sounds. If you hear anything, your system has a nonlinearity causing audible intermodulation of the ultrasonics. Be careful when increasing volume; running into digital or analog clipping, even soft clipping, will suddenly cause loud intermodulation tones.

In summary, it's not certain that intermodulation from ultrasonics will be audible on a given system. The added distortion could be insignificant or it could be noticable. Either way, ultrasonic content is never a benefit, and on plenty of systems it will audibly hurt fidelity. On the systems it doesn't hurt, the cost and complexity of handling ultrasonics could have been saved, or spent on improved audible range performance instead.

Sampling fallacies and misconceptions

Sampling theory is often unintuitive without a signal processing background. It's not surprising most people, even brilliant PhDs in other fields, routinely misunderstand it. It's also not surprising many people don't even realize they have it wrong.
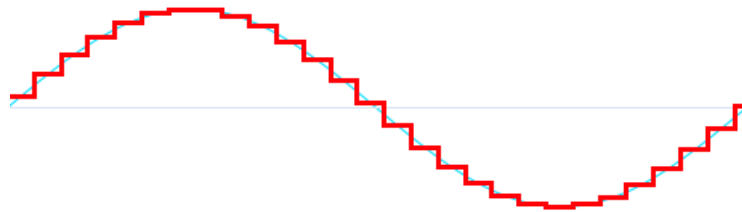
Above: Sampled signals are often depicted as a rough stairstep (red) that seems a poor approximation of the original signal. However, the representation is mathematically exact and the signal recovers the exact smooth shape of the original (blue) when converted back to analog.

The most common misconception is that sampling is fundamentally rough and lossy. A sampled signal is often depicted as a jagged, hard-cornered stair-step facsimile of the original perfectly smooth waveform. If this is how you envision sampling working, you may believe that the faster the sampling rate (and more bits per sample), the finer the stair-step and the closer the approximation will be. The digital signal would sound closer and closer to the original analog signal as sampling rate approaches infinity.

Similarly, many non-DSP people would look at the following:
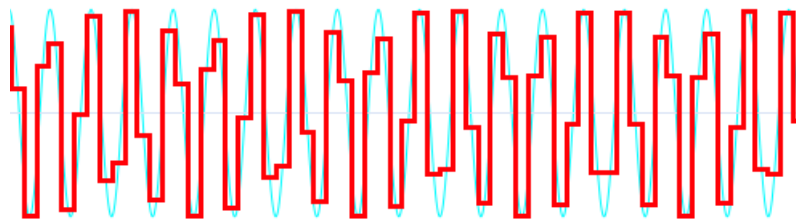
And say, "Ugh!" It might appear that a sampled signal represents higher frequency analog waveforms badly. Or, that as audio frequency increases, the sampled quality falls and frequency response falls off, or becomes sensitive to input phase.

Looks are deceiving. These beliefs are incorrect!

added 2013-04-04:
As a followup to all the mail I got about digital waveforms and stairsteps, I demonstrate actual digital behavior on real equipment in our video Digital Show & Tell so you need not simply take me at my word here!

All signals with content entirely below the Nyquist frequency (half the sampling rate) are captured perfectly and completely by sampling; an infinite sampling rate is not required. Sampling doesn't affect frequency response or phase. The analog signal can be reconstructed losslessly, smoothly, and with the exact timing of the original analog signal.

So the math is ideal, but what of real world complications? The most notorious is the band-limiting requirement. Signals with content over the Nyquist frequency must be lowpassed before sampling to avoid aliasing distortion; this analog lowpass is the infamous antialiasing filter. Antialiasing can't be ideal in practice, but modern techniques bring it very close. ...and with that we come to oversampling.

Oversampling

Sampling rates over 48kHz are irrelevant to high fidelity audio data, but they are internally essential to several modern digital audio techniques. *Oversampling* is the most relevant example [7].

Oversampling is simple and clever. You may recall from my A Digital Media Primer for Geeks that high sampling rates provide a great deal more space between the highest frequency audio we care about (20kHz) and the Nyquist frequency (half the sampling rate). This allows for simpler, smoother, more reliable analog anti-aliasing filters, and thus higher fidelity. This extra space between 20kHz and the Nyquist frequency is essentially just spectral padding for the analog filter.
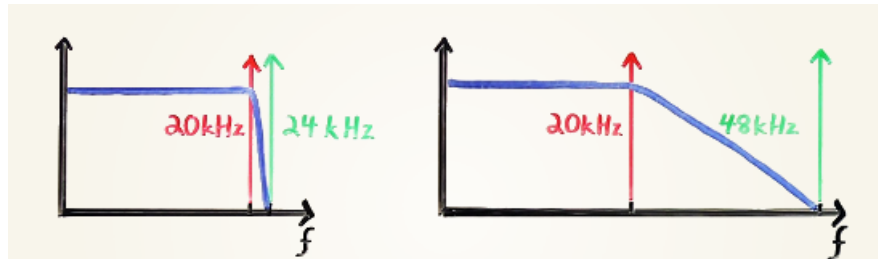
Above: Whiteboard diagram from A Digital Media Primer for Geeks illustrating the transition band width available for a 48kHz ADC/DAC (left) and a 96kHz ADC/DAC (right).

That's only half the story. Because digital filters have few of the practical limitations of an analog filter, we can complete the anti-aliasing process with greater efficiency and precision digitally. The very high rate raw digital signal passes through a digital anti-aliasing filter, which has no trouble fitting a transition band into a tight space. After this further digital anti-aliasing, the extra padding samples are simply thrown away. Oversampled playback approximately works in reverse.

This means we can use low rate 44.1kHz or 48kHz audio with all the fidelity benefits of 192kHz or higher sampling (smooth frequency response, low aliasing) and none of the drawbacks (ultrasonics that cause intermodulation distortion, wasted space). Nearly all of today's analog-to-digital converters (ADCs) and digital-to-analog converters (DACs) oversample at very high rates. Few people realize this is happening because it's completely automatic and hidden.

ADCs and DACs didn't always transparently oversample. Thirty years ago, some recording consoles recorded at high sampling rates using only analog filters, and production and mastering simply used that high rate signal. The digital anti-aliasing and decimation steps (resampling to a lower rate for CDs or DAT) happened in the final stages of mastering. This may well be one of the early reasons 96kHz and 192kHz became associated with professional music production [8].

16 bit vs 24 bit

OK, so 192kHz music files make no sense. Covered, done. What about 16 bit vs. 24 bit audio?

It's true that 16 bit linear PCM audio does not quite cover the entire theoretical dynamic range of the human ear in ideal conditions. Also, there are (and always will be) reasons to use more than 16 bits in recording and production.

None of that is relevant to playback; here 24 bit audio is as useless as 192kHz sampling. The good news is that at least 24 bit depth doesn't harm fidelity. It just doesn't help, and also wastes space.

Revisiting your ears

We've discussed the frequency range of the ear, but what about the dynamic range from the softest possible sound to the loudest possible sound?

One way to define absolute dynamic range would be to look again at the absolute threshold of hearing and threshold of pain curves. The distance between the highest point on the threshold of pain curve and the lowest point on the absolute threshold of hearing curve is about 140 decibels for a young, healthy listener. That wouldn't last long though; +130dB is loud enough to damage hearing permanently in seconds to minutes. For reference purposes, a jackhammer at one meter is only about 100-110dB.

The absolute threshold of hearing increases with age and hearing loss. Interestingly, the threshold of pain decreases with age rather than increasing. The hair cells of the cochlea themselves posses only a fraction of the ear's 140dB range; musculature in the ear continuously adjust the amount of sound reaching the cochlea by shifting the ossicles, much as the iris regulates the amount of light entering the eye [9]. This mechanism stiffens with age, limiting the ear's dynamic range and reducing the effectiveness of its protection mechanisms [10].

Environmental noise

Few people realize how quiet the absolute threshold of hearing really is.

The very quietest perceptible sound is about -8dbSPL [11]. Using an A-weighted scale, the hum from a 100 watt incandescent light bulb one meter away is about 10dBSPL, so about 18dB louder. The bulb will be much louder on a dimmer.

20dBSPL (or 28dB louder than the quietest audible sound) is often quoted for an empty broadcasting/recording studio or sound isolation room. This is the baseline for an exceptionally quiet environment, and one reason you've probably never noticed hearing a light bulb.

The dynamic range of 16 bits

16 bit linear PCM has a dynamic range of 96dB according to the most common definition, which calculates dynamic range as (6*bits)dB. Many believe that 16 bit audio cannot represent arbitrary sounds quieter than -96dB. This is incorrect.

I have linked to two 16 bit audio files here; one contains a 1kHz tone at 0 dB (where 0dB is the loudest possible tone) and the other a 1kHz tone at -105dB.

- **Sample 1:** 1kHz tone at **0 dB** (16 bit / 48kHz WAV)

- **Sample 2:** 1kHz tone at **-105 dB** (16 bit / 48kHz WAV)

63Hz   125Hz   250Hz   500Hz   1kHz   2kHz   4kHz   8kHz
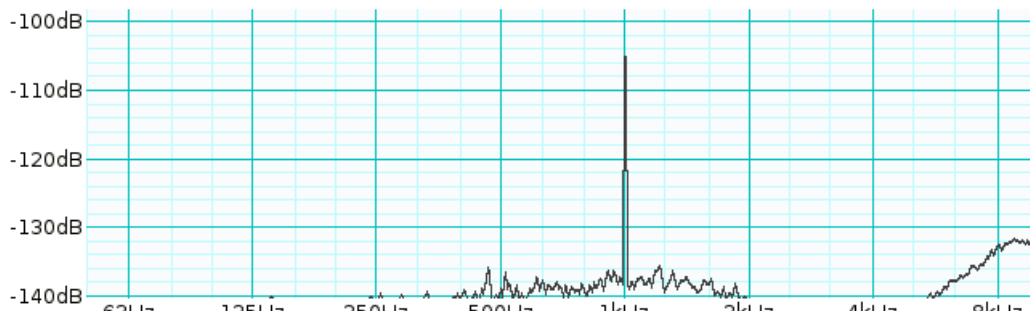
*Above: Spectral analysis of a -105dB tone encoded as 16 bit / 48kHz PCM. 16
bit PCM is clearly deeper than 96dB, else a -105dB tone could not be
represented, nor would it be audible.*

How is it possible to encode this signal, encode it with no distortion, and encode it well above the noise floor, when its peak amplitude is one third of a bit?

Part of this puzzle is solved by proper dither, which renders quantization noise independent of the input signal. By implication, this means that dithered quantization introduces no distortion, just uncorrelated noise. That in turn implies that we can encode signals of arbitrary depth, even those with peak amplitudes much smaller than one bit [12]. However, dither doesn't change the fact that once a signal sinks below the noise floor, it should effectively disappear. How is the -105dB tone still clearly audible above a -96dB noise floor?

The answer: Our -96dB noise floor figure is effectively wrong; we're using an inappropriate definition of dynamic range. (6*bits)dB gives us the RMS noise of the entire broadband signal, but each hair cell in the ear is sensitive to only a narrow fraction of the total bandwidth. As each hair cell hears only a fraction of the total noise floor energy, the noise floor at that hair cell will be much lower than the broadband figure of -96dB.

Thus, 16 bit audio can go considerably deeper than 96dB. With use of shaped dither, which moves quantization noise energy into frequencies where it's harder to hear, the effective dynamic range of 16 bit audio reaches 120dB in practice [13], more than fifteen times deeper than the 96dB claim.

120dB is greater than the difference between a mosquito somewhere in the same room and a jackhammer a foot away.... or the difference between a deserted 'soundproof' room and a sound loud enough to cause hearing damage in seconds.

16 bits is enough to store all we can hear, and will be enough forever.

Signal-to-noise ratio

It's worth mentioning briefly that the ear's S/N ratio is smaller than its absolute dynamic range. Within a given critical band, typical S/N is estimated to only be about 30dB. Relative S/N does not reach the full dynamic range even when considering widely spaced bands. This assures that linear 16 bit PCM offers higher resolution than is actually required.

It is also worth mentioning that increasing the bit depth of the audio representation from 16 to 24 bits does not increase the perceptible resolution or 'fineness' of the audio. It only increases the dynamic range, the range between the softest possible and the loudest possible sound, by lowering the noise floor. However, a 16-bit noise floor is already below what we can hear.

When does 24 bit matter?

Professionals use 24 bit samples in recording and production [14] for headroom, noise floor, and convenience reasons.

16 bits is enough to span the real hearing range with room to spare. It does not span the entire possible signal range of audio equipment. The primary reason to use 24 bits when recording is to prevent mistakes; rather than being careful to center 16 bit recording-- risking clipping if you guess too high and adding noise if you guess too low-- 24 bits allows an operator to set an approximate level and not worry too much about it. Missing the optimal gain setting by a few bits has no consequences, and effects that dynamically compress the recorded range have a deep floor to work with.

An engineer also requires more than 16 bits during mixing and mastering. Modern work flows may involve literally thousands of effects and operations. The quantization noise and noise floor of a 16 bit sample may be undetectable during playback, but multiplying that noise by a few thousand times eventually becomes noticeable. 24 bits keeps the accumulated noise at a very low level. Once the music is ready to distribute, there's no reason to keep more than

16 bits.

Listening tests

Understanding is where theory and reality meet. A matter is settled only when the two agree.

Empirical evidence from listening tests backs up the assertion that 44.1kHz/16 bit provides highest-possible fidelity playback. There are numerous controlled tests confirming this, but I'll plug a recent paper, Audibility of a CD-Standard A/D/A Loop Inserted into High-Resolution Audio Playback, done by local folks here at the Boston Audio Society.

Unfortunately, downloading the full paper requires an AES membership. However it's been discussed widely in articles and on forums, with the authors joining in. Here's a few links:

- The Emperor's New Sampling Rate
- Hydrogen Audio forum discussion thread
- Supplemental information page at the Boston Audio Society, including the equipment and sample lists

This paper presented listeners with a choice between high-rate DVD-A/SACD content, chosen by high-definition audio advocates to show off high-def's superiority, and that same content resampled on the spot down to 16-bit / 44.1kHz Compact Disc rate. The listeners were challenged to identify any difference whatsoever between the two using an ABX methodology. BAS conducted the test using high-end professional equipment in noise-isolated studio listening environments with both amateur and trained professional listeners.

In 554 trials, listeners chose correctly 49.8% of the time. In other words, they were guessing. Not one listener throughout the entire test was able to identify which was 16/44.1 and which was high rate [15], and the 16-bit signal wasn't even dithered!

Another recent study [16] investigated the possibility that ultrasonics were audible, as earlier studies had suggested. The test was constructed to maximize the possibility of detection by placing the intermodulation products where they'd be most audible. It found that the ultrasonic tones were not audible... but the intermodulation distortion products introduced by the loudspeakers could be.

This paper inspired a great deal of further research, much of it with mixed results. Some of the ambiguity is explained by finding that ultrasonics can induce more intermodulation distortion than expected in power amplifiers as well. For example, David Griesinger reproduced this experiment [17] and found that his loudspeaker setup did not introduce audible intermodulation distortion from ultrasonics, but his stereo amplifier did.

Caveat Lector

It's important not to cherry-pick individual papers or 'expert commentary' out of context or from self-interested sources. Not all papers agree completely with these results (and a few disagree in large part), so it's easy to find minority opinions that appear to vindicate every imaginable conclusion. *Regardless, the papers and links above are representative of the vast weight and breadth of the experimental record.* No peer-reviewed paper that has stood the test of time disagrees substantially with these results. Controversy exists only within the consumer and enthusiast audiophile communities.

If anything, the number of ambiguous, inconclusive, and outright invalid experimental results available through Google highlights how tricky it is to construct an accurate, objective test. The differences researchers look for are minute; they require rigorous statistical analysis to spot subconscious choices that escape test subjects' awareness. That we're likely trying to 'prove' something that doesn't exist makes it even more difficult. Proving a null hypothesis is akin to proving the halting problem; you can't. You can only collect evidence that lends overwhelming weight.

Despite this, papers that confirm the null hypothesis are especially strong evidence; confirming inaudibility is far more experimentally difficult than disputing it. Undiscovered mistakes in test methodologies and equipment nearly always produce false positive results (by accidentally introducing audible differences) rather than false negatives.

If professional researchers have such a hard time properly testing for minute, isolated audible differences, you can imagine how hard it is for amateurs.

How to [inadvertently] screw up a listening comparison

The number one comment I heard from believers in super high rate audio was [paraphrasing]: *"I've listened to high rate audio myself and the improvement is obvious. Are you seriously telling me not to trust my own ears?"*

Of course you can trust your ears. It's brains that are gullible. I don't mean that flippantly; as human beings, we're all wired that way.

Confirmation bias, the placebo effect, and double-blind

In any test where a listener can tell two choices apart via any means apart from listening, the results will usually be what the listener expected in advance; this is called confirmation bias and it's similar to the placebo effect. It means people 'hear' differences because of subconscious cues and preferences that have nothing to do with the audio, like preferring a more expensive (or more attractive) amplifier over a cheaper option.

The human brain is designed to notice patterns and differences, even where none exist. This tendency can't just be turned off when a person is asked to make objective decisions; it's completely subconscious. Nor can a bias be defeated by mere skepticism. *Controlled experimentation shows that awareness of confirmation bias can increase rather than decreases the effect!* A test that doesn't carefully eliminate confirmation bias is worthless [18].

In *single-blind* testing, a listener knows nothing in advance about the test choices, and receives no feedback during the course of the test. Single-blind testing is better than casual comparison, but it does not eliminate the experimenter's bias. The test administrator can easily inadvertently influence the test or transfer his own subconscious bias to the listener through inadvertent cues (eg, "Are you sure that's what you're hearing?", body language indicating a 'wrong' choice, hesitating inadvertently, etc). An experimenter's bias has also been experimentally proven to influence a test subject's results.

Double-blind listening tests are the gold standard; in these tests neither the test administrator nor the testee have any knowledge of the test contents or ongoing results. Computer-run ABX tests are the most famous example, and there are freely available tools for performing ABX tests on your own computer[19]. ABX is considered a minimum bar for a listening test to be meaningful; reputable audio forums such as Hydrogen Audio often do not even allow discussion of listening results unless they meet this minimum objectivity requirement [20].

Above: Squishyball, a simple command-line ABX tool, running in an xterm.

I personally don't do any quality comparison tests during development, no matter how casual, without an ABX tool. Science is science, no slacking.

Loudness tricks

The human ear can consciously discriminate amplitude differences of about 1dB, and experiments show subconscious awareness of amplitude differences under .2dB. Humans almost universally consider louder audio to sound better, and .2dB is enough to establish this preference. Any comparison that fails to carefully amplitude-match the choices will see the louder choice preferred, even if the amplitude difference is too small to consciously notice. Stereo salesmen have known this trick for a long time.

The professional testing standard is to match sources to within .1dB or better. This often requires use of an oscilloscope or signal analyzer. Guessing by turning the knobs until two sources sound about the same is not good

enough.

Clipping

Clipping is another easy mistake, sometimes obvious only in retrospect. Even a few clipped samples or their aftereffects are easy to hear compared to an unclipped signal.

The danger of clipping is especially pernicious in tests that create, resample, or otherwise manipulate digital signals on the fly. Suppose we want to compare the fidelity of 48kHz sampling to a 192kHz source sample. A typical way is to downsample from 192kHz to 48kHz, upsample it back to 192kHz, and then compare it to the original 192kHz sample in an ABX test [21]. This arrangement allows us to eliminate any possibility of equipment variation or sample switching influencing the results; we can use the same DAC to play both samples and switch between without any hardware mode changes.

Unfortunately, most samples are mastered to use the full digital range. Naive resampling can and often will clip occasionally. It is necessary to either monitor for clipping (and discard clipped audio) or avoid clipping via some other means such as attenuation.

Different media, different master

I've run across a few articles and blog posts that declare the virtues of 24 bit or 96/192kHz by comparing a CD to an audio DVD (or SACD) of the 'same' recording. This comparison is invalid; the masters are usually different.

Inadvertent cues

Inadvertant audible cues are almost inescapable in older analog and hybrid digital/analog testing setups. Purely digital testing setups can completely eliminate the problem in some forms of testing, but also multiply the potential of complex software bugs. Such limitations and bugs have a long history of causing false-positive results in testing [22].

The Digital Challenge - More on ABX Testing, tells a fascinating story of a specific listening test conducted in 1984 to rebut audiophile authorities of the time who asserted that CDs were inherently inferior to vinyl. The article is not concerned so much with the results of the test (which I suspect you'll be able to guess), but the processes and real-world messiness involved in conducting such a test. For example, an error on the part of the testers inadvertantly revealed that an invited audiophile expert had not been making choices based on audio fidelity, but rather by listening to the slightly different clicks produced by the ABX switch's analog relays!

Anecdotes do not replace data, but this story is instructive of the ease with which undiscovered flaws can bias listening tests. Some of the audiophile beliefs discussed within are also highly entertaining; one hopes that some modern examples are considered just as silly 20 years from now.

Finally, the good news

What actually works to improve the quality of the digital audio to which we're listening?

Better headphones

The easiest fix isn't digital. The most dramatic possible fidelity improvement for the cost comes from a good pair of headphones. Over-ear, in ear, open or closed, it doesn't much matter. They don't even need to be expensive, though expensive headphones can be worth the money.

Keep in mind that some headphones are expensive because they're well made, durable and sound great. Others are expensive because they're $20 headphones under a several hundred dollar layer of styling, brand name, and marketing. I won't make specfic recommendations here, but I will say you're not likely to find good headphones in a big box store, even if it specializes in electronics or music. As in all other aspects of consumer hi-fi, do your research (and caveat emptor).

Lossless formats

It's true enough that a properly encoded Ogg file (or MP3, or AAC file) will be indistinguishable from the original at a moderate bitrate.

But what of badly encoded files?

Twenty years ago, all mp3 encoders were really bad by today's standards. Plenty of these old, bad encoders are still in use, presumably because the licenses are cheaper and most people can't tell or don't care about the difference anyway. Why would any company spend money to fix what it's completely unaware is broken?

Moving to a newer format like Vorbis or AAC doesn't necessarily help. For example, many companies and individuals used (and still use) FFmpeg's very-low-quality built-in Vorbis encoder because it was the default in FFmpeg and they were unaware how bad it was. AAC has an even longer history of widely-deployed, low-quality encoders; all mainstream lossy formats do.

Lossless formats like FLAC avoid any possibility of damaging audio fidelity [23] with a poor quality lossy encoder, or even by a good lossy encoder used incorrectly.

A second reason to distribute lossless formats is to avoid generational loss. Each reencode or transcode loses more data; even if the first encoding is transparent, it's very possible the second will have audible artifacts. This matters to anyone who might want to remix or sample from downloads. It especially matters to us codec researchers; we need clean audio to work with.

Better masters

The BAS test I linked earlier mentions as an aside that the SACD version of a recording *can* sound substantially better than the CD release. It's not because of increased sample rate or depth but because the SACD used a higher-quality master. When bounced to a CD-R, the SACD version still sounds as good as the original SACD and better than the CD release because the original audio used to make the SACD was better. Good production and mastering obviously contribute to the final quality of the music [24].

The recent coverage of 'Mastered for iTunes' and similar initiatives from other industry labels is somewhat encouraging. What remains to be seen is whether or not Apple and the others actually 'get it' or if this is merely a hook for selling consumers yet another, more expensive copy of music they already own.

Surround

Another possible 'sales hook', one I'd enthusiastically buy into myself, is surround recordings. Unfortunately, there's some technical peril here.

Old-style discrete surround with many channels (5.1, 7.1, etc) is a technical relic dating back to the theaters of the 1960s. It is inefficient, using more channels than competing systems. The surround image is limited, and tends to collapse toward the nearer speakers when a listener sits or shifts out of position.

We can represent and encode excellent and robust localization with systems like Ambisonics. The problems are the cost of equipment for reproduction and the fact that something encoded for a natural soundfield both sounds bad when mixed down to stereo, and can't be created artificially in a convincing way. It's hard to fake ambisonics or holographic audio, sort of like how 3D video always seems to degenerate into a gaudy gimmick that reliably makes 5% of the population motion sick.

Binaural audio is similarly difficult. You can't simulate it because it works slightly differently in every person. It's a learned skill tuned to the self-assembling system of the pinnae, ear canals, and neural processing, and it never assembles exactly the same way in any two individuals. People also subconsciously shift their heads to enhance localization, and can't localize well unless they do. That's something that can't be captured in a binaural recording, though it can to an extent in fixed surround.

These are hardly impossible technical hurdles. Discrete surround has a proven following in the marketplace, and I'm personally especially excited by the possibilities offered by Ambisonics.

Outro

"I never did care for music much.
It's the high fidelity!"
　　—Flanders & Swann, A Song of Reproduction

The point is enjoying the music, right? Modern playback fidelity is incomprehensibly better than the already excellent analog systems available a generation ago. Is the logical extreme any more than just another first world problem? Perhaps, but bad mixes and encodings *do* bother me; they distract me from the music, and I'm probably not alone.

Why push back against 24/192? Because it's a solution to a problem that doesn't exist, a business model based on willful ignorance and scamming people. The more that pseudoscience goes unchecked in the world at large, the harder it is for truth to overcome truthiness... even if this is a small and relatively insignificant example.

"For me, it is far better to grasp the Universe as it really is than to persist in delusion, however satisfying and reassuring."
　　—Carl Sagan

Further reading

Readers have alerted me to a pair of excellent papers of which I wasn't aware before beginning my own article. They tackle many of the same points I do in greater detail.

- Coding High Quality Digital Audio by Bob Stuart of Meridian Audio is beautifully concise despite its greater length. Our conclusions differ somewhat (he takes as given the need for a slightly wider frequency range and bit depth without much justification), but the presentation is clear and easy to follow. *[Edit: I may not agree with many of Mr. Stuart's other articles, but I like this one a lot.]*

- Sampling Theory For Digital Audio [Updated link 2012-10-04] by Dan Lavry of Lavry Engineering is another article that several readers pointed out. It expands my two pages or so about sampling, oversampling, and filtering into a more detailed 27 page treatment. Worry not, there are plenty of graphs, examples and references.

Stephane Pigeon of audiocheck.net wrote to plug the browser-based listening tests featured on his web site. The set of tests is relatively small as yet, but several were directly relevant in the context of this article. They worked well and I found the quality to be quite good.

Footnotes

1. As one frustrated poster wrote,

   "[The Sampling Theorem] hasn't been invented to explain how digital audio works, it's the other way around. Digital Audio was invented from the theorem, if you don't believe the theorem then you can't believe in digital audio either!!"

   http://www.head-fi.org/t/415361/24bit-vs-16bit-the-myth-exploded

2. If it wasn't the most boring party trick ever, it was pretty close.

3. It's more typical to speak of visible light as wavelengths measured in nanometers or angstroms. I'm using frequency to be consistent with sound. They're equivalent, as frequency is just the inverse of wavelength.

4. The LED experiment doesn't work with 'ultraviolet' LEDs, mainly because they're not really ultraviolet. They're deep enough violet to cause a little bit of fluorescence, but they're still well within the visible range. Real ultraviolet LEDs cost anywhere from $100-$1000 apiece and would cause eye damage if used for this test. Consumer grade not-really-UV LEDs also emit some faint white light in order to appear brighter, so you'd be able to see them even if the emission peak really was in the ultraviolet.

5. The original version of this article stated that IR LEDs operate from 300-325THz (about 920-980nm), wavelengths that are invisible. Quite a few readers wrote to say that they could in fact just barely see the LEDs in some (or all) of their remotes. Several were kind enough to let me know which remotes these were, and I was able to test several on a spectrometer. Lo and behold, these remotes were using higher-frequency LEDs operating from 350-380THz (800-850nm), just overlapping the extreme edge of the visible range.

6. Many systems that cannot play back 96kHz samples will silently downsample to 48kHz, rather than refuse to play the file. In this case, the tones will not be played at all and playback would be silent no matter how nonlinear the system is.

7. Oversampling is not the only application for high sampling rates in signal processing. There are a few theoretical advantages to producing band-limited audio at a high sampling rate eschewing decimation, even if it is to be downsampled for distribution. It's not clear what if any are used in practice, as the workings of most professional consoles are trade secrets.

8. Historical reasoning or not, there's no question that many professionals today use high rates because they mistakenly assume that retaining content beyond 20kHz sounds better, just as consumers do.

9. The sensation of eardrums 'uncringing' after turning off loud music is quite real!

10. Some nice diagrams can be found at the HyperPhysics site: http://hyperphysics.phy-astr.gsu.edu/hbase/sound/protect.html#c1

11. 20µPa is commonly defined to be 0dB for auditory measurement purposes; it is approximately equal to the threshold of hearing at 1kHz. The ear is as much as 8dB more sensitive between 2 and 4kHz however.

12. The following paper has the best explanation of dither that I've run across. Although it's about image dither, the first half covers the theory and practice of dither in audio before extending its use into images:

    Cameron Nicklaus Christou, Optimal Dither and Noise Shaping in Image Processing

13. DSP engineers may point out, as one of my own smart-alec compatriots did, that 16 bit audio has a theoretically infinite dynamic range for a pure tone if you're allowed to use an infinite Fourier transform to extract it; this concept is very important to radio astronomy.

    Although the ear works not entirely unlike a Fourier transform, its resolution is relatively limited. This places a limit on the maximum practical dynamic depth of 16 bit audio signals.

14. Production increasingly uses 32 bit float, both because it's very convenient on modern processors, and because it completely eliminates the possibility of accidental clipping at any point going undiscovered and ruining a mix.

15. Several readers have wanted to know how, if ultrasonics can cause audible intermodulation distortion, the Meyer and Moran 2007 test could have produced a null result.

    It should be obvious that 'can' and 'sometimes' are not the same as 'will' and 'always'. Intermodulation distortion from ultrasonics is a possibility, not a certainty, in any given system for a given set of material. The Meyer and Moran null result indicates that intermodulation distortion was inaudible on the systems used during the course of their testing.

    Readers are invited to try the simple ultrasonic intermodulation distortion test above for a quick check of the intermodulation potential of their own equipment.

16. Karou and Shogo, Detection of Threshold for tones above 22kHz (2001). Convention paper 5401 presented at the 110th Convention, May 12-15 2001, Amsterdam.

17. Griesinger, Perception of mid-frequency and high-frequency intermodulation distortion in loudspeakers, and its relationship to high definition audio

18. Since publication, several commentators wrote to me with similar versions of the same anecdote [paraphrased]: "I once listened to some headphones / amps / recordings expecting result [A] but was totally

surprised to find [B] instead! Confirmation bias is hooey!"

I offer two thoughts.

First, confirmation bias does not replace all correct results with incorrect results. It skews the results in some uncontrolled direction by an unknown amount. How can you tell right or wrong *for sure* if the test is rigged by your own subconscious? Let's say you expected to hear a large difference but were shocked to hear a small difference. What if there was actually no difference at all? Or, maybe there *was* a difference and, being aware of a potential bias, your well meaning skepticism overcompensated? Or maybe you were completely right? Objective testing, such as ABX, eliminates all this uncertainty.

Second, "So you think you're not biased? Great! Prove it!" The value of an objective test lies not only in its ability to inform one's own understanding, but also to convince others. Claims require proof. Extraordinary claims require extraordinary proof.

19. The easiest tools to use for ABX testing are probably:

    - [Foobar2000](#) with the [ABX plug-in](#)

    - [Squishyball](#), a Linux command-line tool we use within Xiph

20. At Hydrogen Audio, the objective testing requirement is abbreviated *TOS8* as it's the eighth item in the Terms Of Service.

21. It is commonly assumed that resampling irreparably damages a signal; this isn't the case. Unless one makes an obvious mistake, such as causing clipping, the downsampled and then upsampled signal will be audibly indistinguishable from the original. This is the usual test used to establish that higher sampling rates are unneccessary.

22. It may not be strictly audio related, but... faster-than-light neutrinos, anyone?

23. [Wired magazine implies that lossless formats like FLAC are not always completely lossless](#):

    "Some purists will tell you to skip FLACs altogether and just buy WAVs. [...] By buying WAVs, you can avoid the potential data loss incurred when the file is compressed into a FLAC. This data loss is rare, but it happens."

    This is false. A lossless compression process never alters the original data in any way, and FLAC is no exception.

    In the event that Wired was referring to hardware corruption of data files (disk failure, memory failure, sunspots), FLAC and WAV would both be affected. A FLAC file, however, is checksummed and would detect the corruption. The FLAC file is also smaller than the WAV, and so a random corruption would be less likely because there's less data that could be affected.

24. The ['Loudness War'](#) is a commonly cited example of bad mastering practices in the industry today, though it's not the only one. Loudness is also an older phenomenon than the Wikipedia article leads the reader to believe; as early as the 1950s, artists and producers pushed for the loudest possible recordings. Equipment vendors increasingly researched and marketed new technology to allow hotter and hotter masters. Advanced vinyl mastering equipment in the 1970s and 1980s, for example, tracked and nested groove envelopes when possible in order to allow higher amplitudes than the groove spacing would normally permit.

    Today's digital technology has allowed loudness to be pumped up to an absurd level. It's also provided a plethora of automatic, highly complex, proprietary DAW plugins that are deployed en-masse without a wide understanding of how they work or what they're really doing.

—Monty (monty@xiph.org) March 1, 2012
last revised March 25, 2012 to add improvements suggested by readers.
Edits and corrections made after this date are marked inline, except for spelling errors
spotted on Dec 30, 2012 and March 15, 2014, and an extra 'is' removed on April 1, 2013]

Monty's articles and demo work are sponsored by Red Hat Emerging Technologies.
(C) Copyright 2012 Red Hat Inc. and Xiph.Org
Special thanks to Gregory Maxwell for technical contributions to this article