Digital Audio Explained for the Audio Enginner, 2nd Ed.

Nika Aldrich

Chapter 21 – The myths of digital audio, pages 344-350

Chapter Twenty One

The Myths of Digital Audio

Throughout this book we have discussed the finer details of how and why digital audio actually works. At this point we will discuss and discredit, as promised in the preface, seven of the major myths of the audio industry with substantiation borrowed from the last several hundred pages of explanation included herein. To the reader who has not read the entire book, the following points may seem full of holes and incomplete analysis. It really takes this entire book's worth of writing and reading to comprehensively substantiate the following. This is merely a summary combining many of the major topics covered within, with simplified explanations of each.

MYTH NUMBER ONE: HIGHER SAMPLE RATE RECORDINGS INHERENTLY SOUND BETTER THAN LOWER SAMPLE RATE RECORDINGS, ALL THINGS BEING EQUAL

False. This topic probably holds the largest amount of marketing hype, preying on the simple, intuitive conclusion drawn by the majority of people that more equals better.

"HEARING ABOVE 20KHZ"

First, we cannot hear above 20kHz. We established why this is in Chapter Six. We cannot hear above 20kHz. We cannot hear *the effect* of anything above 20kHz. We cannot hear inter-modulation distortion or beat frequencies caused by material above 20kHz being mixed with material in our hearing range. The only way that we can hear the *effect* of material over 20kHz is if it is aliased *back into the hearing range*, or if it is combined with other frequencies in a non-linear environment, creating artificial tones *within our hearing range* that we can hear. In none of these situations are we actually hearing the material above 20kHz. In these situations we are hearing a change in frequency content below 20kHz.

People have suggested that we can "perceive" air pressure changes above 20kHz even if we cannot hear them. It has been pointed out that light, for example, oscillates at a frequency above 20kHz. This, however, is a poor analogy. Light is not comprised of waves of changes in air pressure but is rather comprised of waves of electromagnetic radiation. No documented study has shown that the human body can detect air pressure changes above 20kHz, though one major study has indicated that it cannot. The Oohashi study out of Japan, as presented to the American Physiological Society and the Audio Engineering Society, has demonstrated that under scientifically controlled conditions not a single person in the study was capable of recognizing the presence of harmonic material above 20kHz in a recording.

To date there are, however, three known ways that the body physiologically reacts to sound waves above 20kHz, though in each situation not a single test participant was able to identify that their body was experiencing such a physiological reaction. The first known way again speaks to the Oohashi study out of Japan. In that study, participants yielded different brain wave activity after being hooked up to an alpha-electroencephalogram (Alpha EEG). Certainly it would have been possible for a participant to have looked at his brain wave chart and determined, based on his brain's activity, that his body must have been experiencing frequency content above 20kHz. It would be a stretch to call the looking back at scientific data the "perceiving" of the presence of high frequency information.

The second known physiological change the body can undergo in the presence of sound waves above 20kHz is the transmission of high frequency material through bone conduction. Studies have shown that high frequency material, above the human range, can result in audible detection when a metal rod is applied to the skull and high frequency material is directly applied to it. Specifically, speech recognition amongst the hearing impaired has been shown to improve in these conditions. Unfortunately, however, the signal levels necessary to accomplish this have to be tremendously high and applied directly to the skull. The high frequency material *itself* is not audible, but lower frequency vibrations *in the audible range* are conducted through the bones to the inner ear where they result in fluid vibrations in the audible range that can be heard. For people that are not hearing impaired, however, the material that gets to the ear from bone conduction is so low in relation to the material that gets to the inner ear through the traditional means that it is nearly entirely masked. When the masking effect of this low level transmission is combined with the fact that people do not often listen to music through the conduction of metal rods attached to their heads, the viability that this bone-conduction method of transmitting audio to the brain actually allows the ability to discern material above 20kHz is simply not substantiated. This, in combination with the fact that the Oohashi study mentioned above did not yield a single participant that was capable of discerning the presence of material through this method.

Finally, high frequency sound waves can physically maim the body in extreme situations. This phenomenon is used in the medical industry as a type of non-invasive surgery. Human corneas are currently being reshaped using high frequency sound waves, for example. While this, too, is a physiological change that the body undergoes in the presence of high frequency sound waves, a person should not expect to see any better after listening to music with frequency content above 20kHz. Sonar welding systems currently used in plastics can burn the skin or even a hole right through a person's hand. Not only are the signal levels used in these situations so high as to be laughable when discussing the playback of digital music, but the ability to perceive that one's hand is being burned is likely not the type of "perception" of material above 20kHz that people discuss as a benefit to recording at higher sampling frequencies.

There are many ways of quickly running incomplete tests and drawing inaccurate conclusions about the ability to perceive sound waves above 20kHz aided by inaccurate playback equipment that induces beat tones into the audible range. To date, no properly conducted study has indicated that sound waves above 20kHz are perceivable at all to the human.

Because of this, the supposed major benefit of recording material above 44.1kS/s is moot. There is not an audible benefit to capturing the material above 20kHz in a recording in any way. This is actually one of the ways that analog mixers cause degradation to the signal. If two speakers are placed on a stage, one of which plays a 15kHz tone and the other of which plays a 25kHz tone we know that we would only hear the 15kHz tone. The 25kHz tone would not be audible, and the 10kHz "beat frequency" would not be audible on account that the higher tone was inaudible. If the air is allowed to mix the two sounds then the result is that only the 15kHz tone is audible.

If the material is recorded using a microphone in front of each speaker and the recording is done in a digital environment at a sample rate of 44.1kS/s then the 15kHz tone would be recorded and the 25kHz tone would not. If the material is then mixed digitally only the 15kHz tone will be present, just as the ear should hear it.

If the tones are recorded, however, with higher sample frequencies then both the 15kHz and the 25kHz tone will be recorded. If the two channels are then mixed in (or passed through) an analog environment (such as an amplifier) that has some degree of non-linearity then the beat frequency of 10kHz *will* be created, and when the recording is played back the 10kHz tone will be available to be heard. This is an inaccurate reproduction of the original material with respect to the ear. In any natural environment the 10kHz sub-tone would not have been audibly present.

By recording higher frequencies than the ear can hear we create the possibility that the analog equipment in the signal path will create unnatural and inaccurate results. This very effect is often blamed for tests in which square waves at 15kHz are said to sound different than sine waves at 15kHz. We already know that the human ear cannot hear the difference between a square wave and a sine wave, each with a fundamental of 15kHz, as the first overtone of the square wave is at 45kHz, well above the hearing range. If a difference can be heard it can often be identified to be the creation of harmonic material *within* the hearing range because of nonlinearity and distortion in the playback equipment upon attempting to recreate high frequency waveforms. (This, and the fact that a square wave and a sine wave of equal amplitude have different amplitudes of the 15kHz fundamental, resulting in the square wave version sounding louder than the sine wave version by a few decibels.) Since any analog component is non-linear, recording material that should have no effect on audibility only provides the possibility that distortion may be added within the audible range, thereby affecting the accuracy of the playback of the recorded material.

THE CONVERTERS

Several tests have been done that have indicated that recordings done at higher sample frequencies have sounded "different" than recordings done at lower sampling frequencies. In most situations this can be attributed to the analog to digital and digital to analog converters used in the test.

We discussed the importance of the anti-aliasing and reconstruction filters in converter design, and the fact that the proper design of them is imperative to the proper function of digital circuits. If the anti-aliasing filter rolls off at too low of a frequency then the audible spectrum is affected. If the anti-aliasing filter rolls off at a frequency that is too high, or if the transition band reaches above the Nyquist frequency then aliasing will be allowed into the signal within the audible band. Per our discussion on digital filters in Chapter Eighteen, the perfect anti-aliasing filter *can* be designed to not attenuate any frequency content below 20kHz but to attenuate at least 144dB FS at all frequencies at or above 22.05kHz and not affect the phase of the frequencies in the pass band. This filter, however, is very steep and requires a tremendous number of taps and very accurate calculations for each of its respective tap values. Most filters built into the chips of A/D converters are not audibly transparent. As a testament to that fact, several manufacturers design A/D converter circuits that bypass the filters on the converter chip and replace them with very computationally expensive DSP so that the filtering process is more accurate.

The advantage to higher sample rates in this context is that the filters in the converters do not have to be as steep and can therefore be less expensive to implement. Many inexpensive converters will demonstrate a marked improvement when comparing 44.1kS/s recordings versus 96kS/s recordings, for example. This is often because the filters in the converters are not of audibly transparent quality. On higher quality (and thus more expensive) converters, the difference between 44.1kS/s recording and 96kS/s recording is indeterminable under proper testing conditions and procedures. Not to be fooled, however, even some of the most expensive converters have filters in them that are not audibly transparent. One such manufacturer provides multiple filter settings. Clearly, not all of the filter settings can be accurate. As such, at least most of their multiple settings are "flawed" by the specifications we require.

Because we know that sound waves above 20kHz are not audible, and because we know that digital filters implemented for higher sample rate recording are significantly less affecting of the audible range, a very simple test can be done to determine the quality of the filters in any analog to digital converters: record the same material through the same converters at 44.1kS/s and at 96kS/s. If there is an audible difference then most likely the converters are not audibly transparent in the 44.1kS/s setting. Clearly this is not a completely sufficient test, as the playback equipment after the converters can cause audible-range distortion when trying to reproduce frequencies outside of its specified range. As a general rule, however, it can indicate a lot about the quality of a given converter – especially on good playback equipment.

It is therefore correct that there *can* be audible differences between lower sample rate and higher sample rate recordings of band-limited material, but only insomuch as the quality of the filters in the converters varies. The notion that there is an *inherent* difference is false, as with properly designed, audibly transparent equipment, no difference will be audible.

PROCESSING BENEFITS

We have discussed that the processing of non-linear dynamics should be done at higher sample rates in order to allow the distortion frequencies above the bounds of the digital system to be removed through filtering. This does not, however, necessitate the *recording* of the material at higher sample rates. It only speaks to the *processing* of the material at higher sample rates. By the time the recording is finished, mixed, mastered and ready for delivery, there is no need to present the consumer with recordings at higher sample rates for the sake of processing. There is also not a need to record the material at those rates for the sake of processing that is to be done. During the processing stage, the material should be upsampled, processed and then downsampled. There is not a need to leave it at the higher sample rate throughout the system.

Upsampling and downsampling both require low pass filters as part of the process. The algorithm for the upsampling and downsampling filters are precisely the same as the algorithms needed for anti-aliasing and reconstruction filters. A logical conclusion is that if the converters do not often have audibly transparent filters then the processing stages, such as the implementation of compression, could not be expected to have audibly transparent filters either. It is important when working with base sample rates such as 44.1kS/s to use algorithms that perform the functions accurately and in an audibly accurate fashion including the upsampling and downsampling portions of the algorithms. Again, there is not an *inherent* benefit to recording audio with respect to any necessary processing the signal may undergo, though there can be benefits if processes are used with less than adequate implementation.

ACCURATE REPRESENTATION OF WAVEFORMS

The introduction of the digital audio workstation into the recording industry has allowed a very intuitive (but very incorrect) conclusion to be drawn from looking at waveforms on computer screens. It is easy to correctly analyze that, at 44.1kS/s sample rates, a 20kHz sine wave would only get sampled twice per waveform. A 10kHz sine wave would only get sampled four times per waveform. A 5kHz sine wave would only get sampled 8 times per waveform, etc. This leads to the conclusion that, clearly, the 10kHz waveform, for example, cannot be accurately recreated to be a 10kHz sine wave. The representation on the computer screen clearly shows a triangle wave or a square-ish waveform. Further, the amplitude of the waveform is not accurately represented

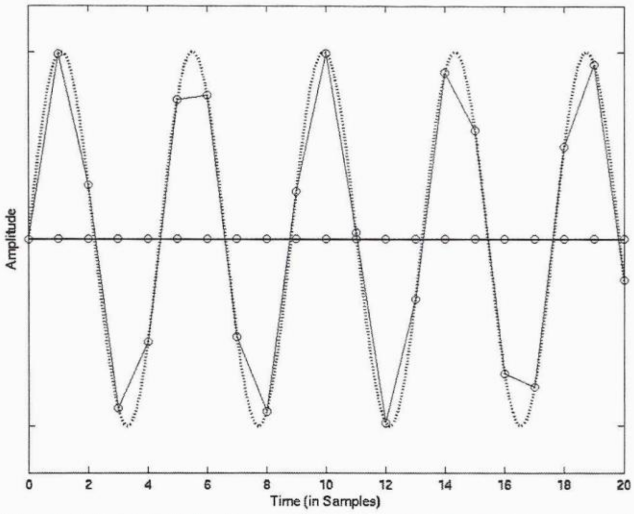

FIGURE 21.1: A Typical Representation Of 10kHz Waveform in a Computer System (at 44.1kS/s)

if the sampling points are near the zero crossing. The typical on-screen representation of sample data is shown in Figure 21.1.

The visual demonstration on the screen shows a "connect the dots" display and does not demonstrate what waveform will be created after the digital values have been put through the reconstruction filter in the digital to analog converter. A 10kHz sine wave of a given amplitude, digitally sampled at 44.1kS/s, will only be sampled four times in a single cycle. When those four samples are fed through a reconstruction filter and properly converted to analog, however, the resulting waveform will be a 10kHz waveform of precisely the same frequency, phase, and amplitude that went in to the system.

FROM WHENCE THE MYTH ORIGINATES

In the case of high sample rate recording, the myth likely stems from the fact that audible differences have been able to be heard because of poor filtering in converters. At the time that the industry started to murmur about high sample rate recording, converters used analog anti-aliasing filters instead of the process of oversampling and the incorporation of digital anti-aliasing filters. Analog filters have much worse performance than digital (linear phase) filters do because they are not phase linear. Until digital filters were used, the industry was clamoring for a way to improve the sound quality and the notion of increasing the sample frequency seemed an easy remedy.

This myth has been furthered by the intuitive (yet incorrect) deduction that "more must be better" and simple misunderstandings of many of the issues covered in this book.

Finally, manufacturers of audio equipment and industry media have aided and abetted the myth by either presenting completely incorrect information to consumers or by allowing consumers to draw incorrect conclusions.

While this myth is incorrect in its wording, it is valid to recall that many pieces of equipment are designed to fit particular budgets and the result can yield lesser quality implementation. This can lead to an audible improvement when using higher sample frequencies. There is not, however, an *inherent* difference, and when all is equal and the systems are properly designed for audible transparency there are no audible difference between different sample rates for recording.

MYTH NUMBER TWO: MORE BITS INHERENTLY SOUND BETTER THAN FEWER BITS, ALL THINGS BEING EQUAL

False. This statement is again an example of intuitive deductions based on incomplete information providing the opportunity for incorrect conclusions.

There are only four characteristics of audio waveforms: amplitude, frequency, phase, and dynamic range. In digital sampling technology there are two variables: the bit depth and the sample rate. The sample rate affects the maximum frequency that can be represented and the bit depth affects only one property of recordings: the dynamic range.

Per our discussion in Chapter Nine, the higher the bit depth the more quantization steps are available. The more quantization steps the less the quantization error. The less the quantization error the lower the quantization noise. The lower the quantization noise the lower the noise floor. The lower the noise floor the greater the dynamic range. Therefore, the more bits used, the more dynamic range, but that is the *only* property of a waveform that is affected by bit depth.

Per the formula that we derived, each "bit" provides 6dB of dynamic range in the recording system. This means that 16 bits provides a maximum capability of 96dB of dynamic range. 24 bits provides a maximum capability of 144dB of dynamic range. The number of bits needed to record a signal is completely based on the dynamic range of the material being recorded.

Further, if the material is to be played back, the number of bits needed is limited to the amplitude with which the material is to be played. If the material is to be played at a level of peak=85dB SPL then the noise level of an optimal 16-bit system will be –11dB SPL, well below the threshold of hearing. If the material is to be played at a level of peak=85dB SPL, but the white noise floor in the room it is to be played in is 45dB SPL then the maximum dynamic range that will be heard is around 40dB SPL. Even recognizing the human's ability to hear up to 25dB or so below a given white noise floor, the maximum dynamic range that will be heard (including the range below the noise floor) is only 65dB. 65dB Of dynamic range only requires 11 bits. Therefore, an 11-bit recording will sound the same in that situation as a 24-bit recording.

The only situation in which a recording that utilizes 24 bits will sound different than a recording that uses 16 bits (using the same, linear converters) is when the difference between the amplitude of the playback (in dB SPL) and the amplitude of the lowest level discernable sound that can be heard in the room (in dB SPL) is greater than the dynamic range capability of the 16-bit converter (in dB).

The only situation where using 24 bits to record with will sound different than using 16 bits to record with, assuming linear converters, is if the difference between the peak amplitude of the signal and the lowest level discernable sound in the signal (the dynamic range of the signal) is greater than the dynamic range of the 16-bit converters. Recording such signals is actually not very common, but a further benefit exists in the recording part of the signal chain in that an increase of bit depth allows for more headroom while recording in order to prevent clipping from running the signal too hot to ensure an accurate capture.

Finally, with the bit reduction algorithms (such as colored dither and noise shaping) available on the market, 16-bit delivery formats can provide up to 150dB of dynamic range in the most sensitive frequency ranges of human hearing, keeping all of the quantization error below the threshold of human hearing at all frequencies for any viable amplitude with which the signal might be played back. While there can still be viable recording situations wherein 24 bits has advantages over 16 bits (if, for no other reason, the headroom gained when recording live) only in extreme situations should a mastered 24-bit signal sound any different from a properly prepared and mastered and bit-reduced 16-bit signal.

LOSS OF "RESOLUTION"

Until now the term "resolution" has not been used in conjunction with bit depth of digital recordings in this book. Certainly, more quantization steps provides for more "statistical resolution" by mere definition. The term "resolution," however, also implies a subjective measure of quality. The phrase, "that recording didn't have a lot of 'resolution'" implies that the signal was audibly distorted. The connection between the phrase "bit depth" and the word "resolution" has led to the confusion that fewer bits leads to a reduction in sound quality. There are, however, only four characteristics of a waveform, as described above. The "resolution" of the waveform is not one of them. A change in "statistical resolution" *causes* a change in one of the characteristics in a waveform, and per our reading that area of change is the dynamic range of the waveform. If the statistical resolution of the recording provides enough dynamic range to represent a signal accurately then more quantization steps does not improve the audible quality of a recording.

FROM WHENCE THE MYTH ORIGINATES

The myth of the number of bits in a recording grows, again, out of incorrect conclusions drawn by "intuitive" but incorrect analysis. The widely held philosophy that "more is always better" assists in this deduction. This is further compounded by the observation that every increase in bit depth yields a *doubling* of the amount of quantization steps available, such that the difference between 16 bits (65,536 quantization steps) and 17 bits (131,072 quantization steps) is double. It should be clear after reading Chapter Nine, however, that the number of bits used in a recording does not need to exceed the dynamic range requirements of a signal.

This does not undermine the benefits of higher bit depth *processing* of signals. Processing should almost always be done at bit depths that are greater than the bit depth of the inbound data in order to maintain accuracy regardless of the dynamic range needed for the recording or the playing back of signals. This also does not undermine the value of 24-bit recorders. 24-bit analog to digital converters can currently achieve a dynamic range of approximately 120dB (while theoretically 144dB, approximately 24dB is lost to the inherent noise in the conversion process). There are several potential audio sources that have very wide dynamic ranges. Further, we want the dynamic range of the recording system to be larger than the largest dynamic range in the loudest source we intend to record in order to provide adequate "headroom."

This does, however, undermine the notion that greater than 16 bits is required for playback. When a recording is ready to be heard it does not need any headroom. If all 16 bits in a 16-bit recording are used then the result is only inadequate if, in an anechoic chamber with 0dB SPL of room noise, the recording is to be played back at greater than 96dB SPL, peak amplitude. In more realistic environments where the threshold of audibility is above 20dB SPL, a 16-bit recording is only inadequate if the material is played back at amplitudes that border on the threshold of pain, louder than concert levels, and at 116dB SPL. As this is entirely an unrealistic need, a playback format that uses more than 16 bits is entirely unnecessary. If the converter has a linear "transfer function" in so much as the ear can discern, then there will be no audible difference between 16-bit and 24-bit recordings when played back at low amplitudes in environments with a noise floor high enough that the dynamic range of the playback environment is less than the dynamic range of the recording.

Further, 16-bit material playing back can exceed the audible dynamic range of 16-bit material being recorded because on playback one can use noise shaping such as POW-r to effectively get 24-bit performance out of only 16 bits. Although the 16-bit playback signal never has more than 96dB of dynamic range, the noise is spread unevenly across the frequency spectrum to accommodate the contour of the threshold of hearing with respect to frequency. At the frequencies at which the ear is most sensitive, the dynamic range can be up to 150dB while at frequencies at which the ear has a very high threshold of hearing (perhaps 60dB SPL) the dynamic range can be lower. Over the entire frequency range it may provide 96dB of dynamic range, but this can be spread to provide seemingly 24-bit quality as far as the ear is concerned using only 16 bits.