The Optimal Sample Rate for Quality Audio

By Dan Lavry, Lavry Engineering Inc. May 3, 2012

Imagine that you and a friend were standing at the base of a lone hill in an otherwise flat plain, and you start walking up that hill towards the top. Now imagine that when you reach the top, your friend tries to convince you that you can go higher up the hill if you simply walked twice the distance forward from the base. You would probably pause for a moment, look around, and quickly realize that continuing forward would just take you back down the other side of the hill! This concept of an optimum is not hard to understand, but in digital audio circles more and more people are taking the word of their friend the salesman, who tells them to keep on walking.

In this paper, I will cover some of the myths of higher sampling rate and illustrate how higher sampling rates can actually reduce accuracy in audio conversion. Moreover, I will attempt to elucidate the existence of an optimal sample rate and how conversion at higher or lower rates compromises the accuracy of the audio signal.

The function of audio conversion is to convert, not to alter the signal. Ideally, an AD converter followed by a DA would bring about NO change in the signal (transparency). It has been said that an AD followed by a DA should approach the behavior of a short "wire" at audio frequencies.

Let's talk about converter accuracy. In reality, good audio performance requires extremely low distortion because the ear is very sensitive and perceptive. Personally; I am for accuracy and do not advocate placing limits on accuracy.

In fact, high quality audio converters operating at sample rates no higher than 96 KHz offer results that are very close to the desired theoretical limits. Yet, there are many who subscribe to the false notion that operating above the optimal sample rate can improve the audio. The truth is that there is an optimal sample rate, and that operating above that optimal sample rate compromises the accuracy of audio. To some, this may seem counterintuitive, but is completely proven; whereas most supporters of higher than optimum sample rates offer only subjective results in support.

In my paper "Sampling Theory"[1] I already pointed out that increased speed (bandwidth) reduces accuracy. No one advocates sampling at 10 KHz because that would exclude audio signals above 5 KHz. Clearly, no one knowledgeable in the subject would advocate audio conversion at 100MHz, either. It would bring about poor audio performance due to high distortions, noise and more (which is beyond the scope of this paper). I use these extreme examples to show that sampling can be done too slowly, and it can also be done too fast. There is an OPTIMAL sample rate; fast enough to accommodate everything we hear (the audible range). But exceeding this optimal sample rate will only reduce audio accuracy.

There are many circuit related reasons for the tradeoff between speed and accuracy. In the past I pointed to the example of charging a capacitor – the longer one waits, the closer the voltage to the target[2]. I also mentioned that when feeding OPAMPS (or discrete circuits) with a "voltage step," the longer the wait, the closer the output voltage will be to the target. "Settling time" to an accuracy of .1% is longer than settling time to 1%. I used these examples because they are real issues involving the trade-off between speed and accuracy that one must consider in the design of converters.

In this article I am going to add to my previous arguments.

---

[1] http://lavryengineering.com/pdfs/lavry-sampling-theory.pdf

[2] The "target" is accurate measurement of the signal as versus the signal value skewed by circuit-induced anomalies.

It is always unwise and potentially harmful to include signals that are not needed; it is good practice to keep the signals that are needed, while keeping everything else out. For example, a digital audio clock signal needs to be a clean signal, with minimal radio frequency content, and minimal AC line energy or signals other than what is needed. Similarly, a theoretically perfect DC supply has no spikes and no AC content.

And for audio, it is best to reject the energy outside of the range of hearing. Of course I want to accommodate the most sensitive ears, and with some serious “safety margin”. However, even after adding an extra 10 KHz, we are talking about bandwidth no higher than 40 KHz. That is why 88.2 or 96 KHz are preferred rates for audio quality.

It has been well documented that acoustic musical instruments generate energy at frequencies far above audibility. In the performance space (before any recording takes place), if there is any mechanism that enable ultrasonic[3] frequencies to impact what we hear, it would require energy transfer from ultrasonic frequencies to the audible range. Therefore, using microphones and gear that cover what we hear enables us to capture and keep ALL the energy we need. We can store it, convert it and at some point play back all that we need. There is no good reason for keeping what we don’t hear, because everything we heard in the original performance is already there.

Sampling faster enables recording and keeping the energy that we do not hear. At best it will cause no harm. In reality there is a potential that keeping ultrasonic frequencies will cause unwanted audible alterations. One of the more well-known mechanisms for such alterations is imperfect linearity (non-linearity) in equipment. The type of distortions generated by non-linearity is called intermodulation, and such distortions are rather offensive in nature because the distortion energy is not harmonic. Harmonic distortion tends to alter the timbre, it “colors the sound” by changing the relative harmonic content. Intermodulation is much worse, it is not related to the sound or its harmonics; thus it takes much less intermodulation distortion to become offensive to the ear.

As a practical matter, linearity gets worse and worse as frequency increase. The linearity at the audible range (lower frequencies) is better than the linearity at ultrasonic frequencies. While keeping the unnecessary ultrasonic energy may cause harm, making sure not to include the signals that we don’t hear provides protection against such degradation. With none of the un-needed signals to “spill over” the problem is gone!

Most microphones are designed to match the human hearing frequency range, so they don’t pick up ultrasonic energy; which is a good thing. It prevents the intermodulation I spoke of earlier from happening. Simply speaking, it is a good idea to keep as much as possible of the part of the energy we need, and to get rid of the energy we don’t need (ultrasonic frequencies) as early as possible in the audio chain.

Again, ultrasonic energy will at best cause no harm to the sound you need, but it certainly cannot help; and there is another price to pay. With the current higher sample rates, the file size is doubled or even four times larger, and file transfer rates are proportionally longer. Is this price worth paying when you consider that your audio quality may also be compromised by the “spilling over” of non-musical energy in addition to all of the other inaccuracies due to using a higher than optimal sample rate?

Good conversion requires attention to capturing and reproducing the range we hear while filtering and keeping out energy in the frequency range outside of our hearing. At 44.1 KHz sampling the flatness response may be an issue. If each of the elements (microphone, AD, DA and speaker) limit the audio bandwidth to 20 KHz (each causing a 3dB loss at 20 KHz), the combined impact is -12dB at 20 KHz.

[3] Ultrasonic- Of or involving sound waves with a frequency above the upper limit of human hearing.

At 60 KHz sampling rate, the contribution of AD and DA to any attenuation in the audible range is negligible. Although 60 KHz would be closer to the ideal; given the existing standards, 88.2 KHz and 96 KHz are closest to the optimal sample rate. At 96 KHz sampling rate the theoretical bandwidth is 48 KHz. In designing a real world converter operating at 96 KHz, one ends up with a bandwidth of approximately 40 KHz.

People that go for wide band special mics are actually leaving an opening for distortions. And equipment makers that tell you that it is good to contain a lot more bandwidth, are like someone trying to sell you twice the rental space that you need. The question is: why, when the extra space cost more? Another analogy is a camera capable of reproducing wavelengths that we don't see at extra cost. There is a potential for some ultra-violet and infrared to reduce the quality of what we do see.

Another misconception regarding higher sample rates has to do with the effect of time resolution on stereo location. Sampling does not require a specific time relationship between the sample times and the audio signals. Such a relationship cannot exist; because audio signals are arbitrary (any signal shape within the audio band is possible). There is no such thing as a "correct" sampling time with respect to the signal.

The sampling concept allows perfect capture and reconstruction of the signals as long as the sampling rate is at least twice as fast as the highest frequency contained in the signal to be recorded. There are no restrictions regarding the time at which the samples are taken, as long as the time interval between each adjacent sample is the same.

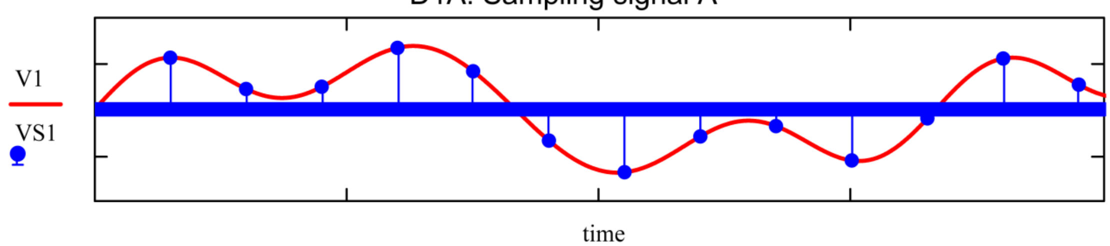
When sampling (and reconstructing) stereo or surround, all the tracks are sampled simultaneously; which ensures that the proper time relationship between channels is maintained. The sample rate does not have ANY impact on the signal timing.

Diagram 1A shows the AD conversion of an audio signal in solid red. The sample values in blue.

Diagram 1B shows the DA side. The reconstruction samples (blue) are used to "pin down" the signal at sample times. The signal (red dotted line) is reconstructed from those samples.

These plots will be referred to as the "Reference Plot."

D1A: Sampling signal A



D1B: Signal A reconstructed from samples

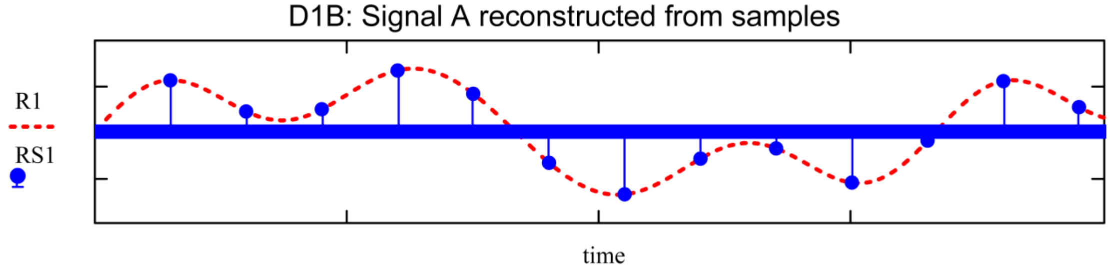
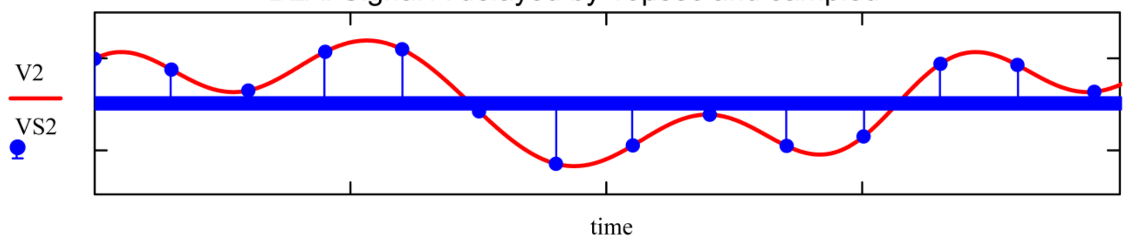


Diagram 2A illustrates feeding the AD with a delayed audio signal (relative to the Reference Plots). The sample times are kept the same as in the Reference Plots. The delay results in different sample values for the same audio waveform.

Diagram 2B shows that the DA output is timed correctly. The reconstruction is made with sample values that reflect the time delay. The samples pin down the signal time properly. The timing of the reconstructed audio has nothing to do with the position of the samples relative to the waveform. In theory, sampling and reconstruction enable zero timing error between channels. Sample rate has nothing to do with accuracy of stereo time location (audio delay between channels).

D2A: Signal A delayed by 10μsec and sampled



D2B: Delayed signal A reconstructed from samples

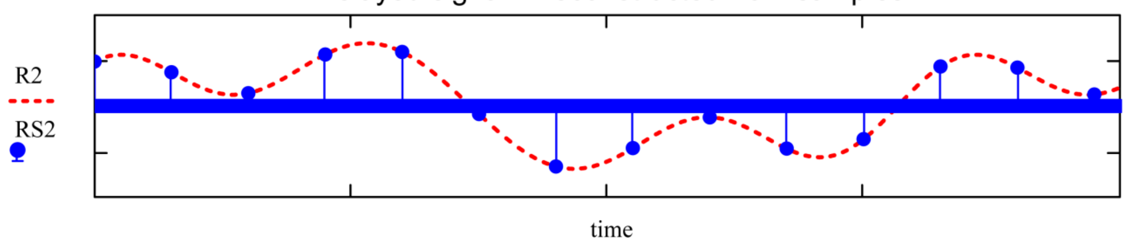
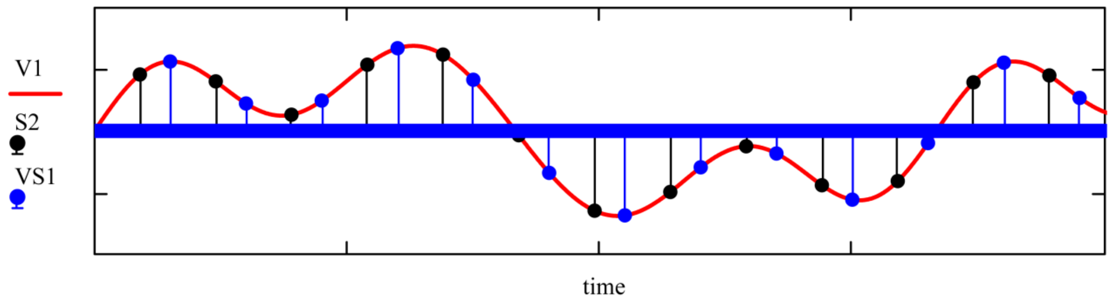


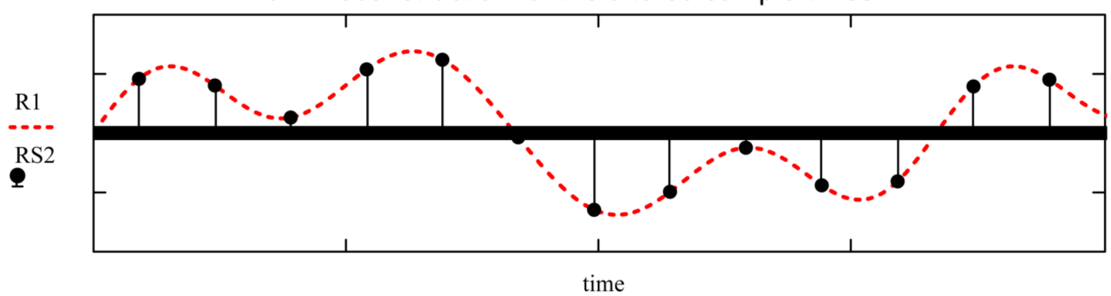
Diagram 3A illustrates the same AD input signal as in the Reference Plot (without a delay). The “new” sample times (shown in black) are shifted with respect to the sample times in diagram 1A. The Reference Plot sample times are shown in blue for comparison.

Diagram 3B show how the reconstruction samples (black) "pin down" the signal at the new sample times. The signal (red dotted line) is reconstructed at the correct time. Sampling the signal at different times does not alter the delay of the audio signal at all.

D3A: Signal A with altered sample times (previous times in blue)

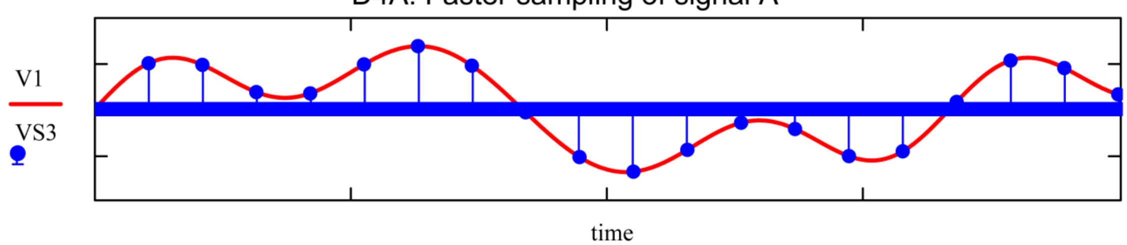


D3B: Reconstruction for the altered sample times

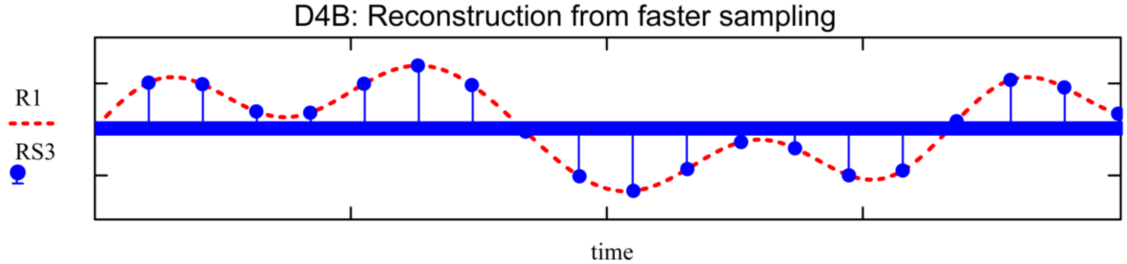


Now, let’s visit the myth that faster sample rates changes the timing of an audio signal. Plots 4A and 4B show the Reference Plot signal sampled and reconstructed at a higher rate. The only effect that having more "dots on the curve" has is to enable the capture and reconstruction of ultrasonic frequencies. There is NO IMPACT on the timing of the signal.

D4A: Faster sampling of signal A

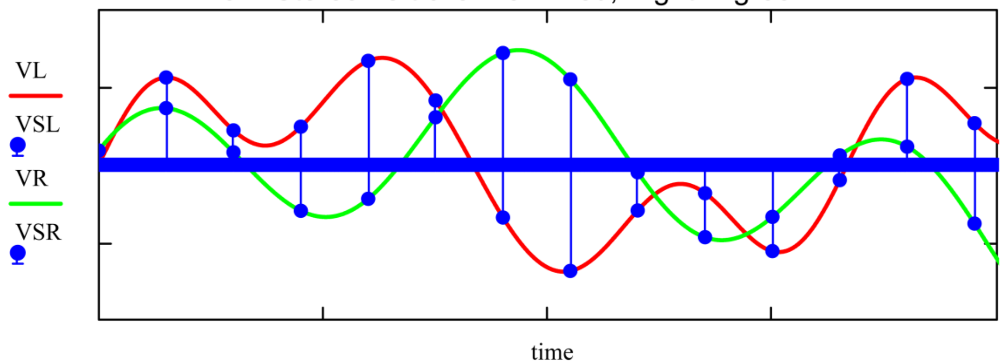


D4B: Reconstruction from faster sampling

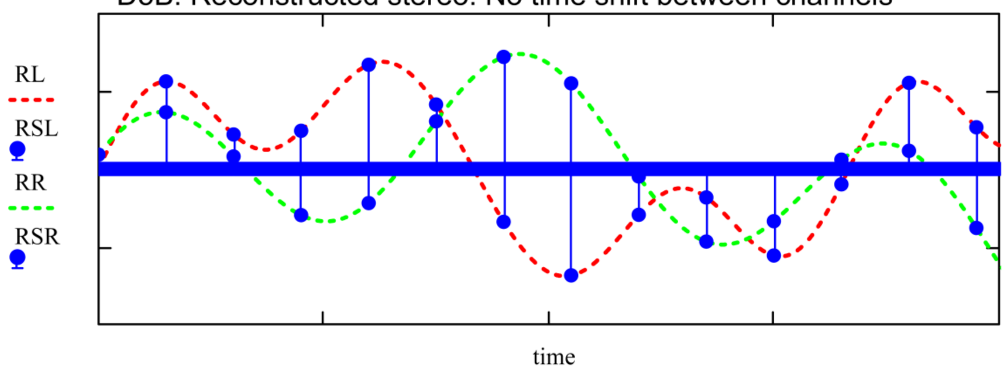


The plot below shows STEREO sampling and reconstruction. As I explain in my paper "Sampling Theory," connecting the dots to generate the "rest of the curve" (the signal between the sample points) can be done perfectly (the correct curvatures); as long as the sample rate exceeds the signal bandwidth by at least a factor of 2. The reconstruction (between the sample dots) is based on filtering. But the reconstructed signal voltage at each sample time is defined directly by the sample value at that time.

D5A: Stereo Left channel in red, Right in green



D5B: Reconstructed stereo: No time shift between channels



Given that stereo audio tracks are sampled simultaneously and reconstructed simultaneously, there is no relative time shift between channels. Clearly, the time between samples for a 44.1 KHz system is 22.7µsec. For 384 KHz the time is 2.6µsec. But it does not matter. Decreasing the time between samples does nothing to alter the accuracy of the timing relationship between channels!

Proper time location for audio does demand close enough matching of delay between channels. Some people say they can hear µsec differences, others think it is tens of µsec (for the most part, under 22.7µSec). This is not in question. The ERROR is to assume that decreasing the time between samples increases the accuracy of the timing of the audio signal, yielding better matching between channels. The notion that faster sampling yields better stereo location is FALSE, because it is based on an incorrect assumption that timing accuracy increases with sample frequency!

Some people claim in error that for double the storage and double the time required for file transfer or download, they will increase converter accuracy and as a result, provide a whole new aspect of audibility that you cannot achieve otherwise. The fact is that they are providing the opposite of what they claim- faster sampling only *compromises* audio accuracy. Proper use of technology begins with understanding the specific goals. Technology makes it possible for factories to manufacture pills with 10 times the dosage required by humans. Needless to say it is a bad idea, until the day that humans are 10 times larger. Similarly, technology makes it possible to convert audio at 10 times the speed. That too is a bad idea. Marketing will always be looking for ways to sell new gear, but false notions regarding sample rates will remain false until the range of human hearing increases beyond where it has been for thousands of years.