

TECH TALK

Mixing in the Box

A detailed look at some of the myths and legends surrounding Pro Tools' mix bus.

By Stan Cotey

Introduction

Welcome to an article that's all about mixing with Pro Tools. This article will cover the basic signal path in Pro Tools, how the mixer plug-in works, and how signals get from place to place and in what form, thus providing enough information to make you better engineers and mixers. We'll be exploring the math of Pro Tools in some detail (but don't worry, we'll keep it pretty high-level), some of the myths having to do with the summing bus, and some general information about how digital audio works.

That Pro Tools is a comprehensive, all-in-one recording, editing, processing, and mixing environment goes almost without saying — it has become nearly ubiquitous in our industry. But along the road to Pro Tools' acceptance, have we lost sight of what should be an underlying principal of any audio device, that of preserving high intrinsic audio quality?

The performance of an audio interface is easy to measure and even easier to qualify through listening tests. Similarly, the quality of the basic input-to-output audio path in a workstation is also easy to measure or to judge through carefully controlled listening tests. These tend to be the easier parts of a workstation to trust.

It is harder to quantify what happens sonically when a workstation must mix anywhere from a few to over a hundred audio channels. We feel it's again easy to perform listening tests to validate the final results and have several times set up such tests in facilities with well-known producers and engineers. The results have been very clear. In a well-executed test with modern analog and digital consoles (see Note 1 at end of article), no one has so far been able to reliably determine whether they were listening to the summing output from the external mixer or from Pro Tools itself.

There is a "standard" practice of using Pro Tools as a multi-track recorder connected to an external console to mix the audio down to the final output. While this practice in itself is fine and allows the console's own EQ and dynamics to be brought into the mix, it leaves behind some of the more powerful aspects of working within Pro Tools — that of integrating the audio data tightly with the mixing automation data, as well as having a unified session file that contains the mixer configuration, channel names, effects processing and settings, and automation all in one portable file format along with the audio data.

But do you lose sound quality when you decide to work entirely in Pro Tools' domain or "in the box?" This article will examine the workings of the Pro Tools mixer, as well as provide some general technical information about analog and digital recording and mostly, mixing, with the hope of clearing up some of the prevalent myths.

Pro Tools Mixing and Quality Audio: Strange Bedfellows?

If any serious work is to be accomplished, the single most important aspect of this multi-faceted system must be the sound quality. The quality of the recorded audio on a per-track basis is due mostly to the sonic quality of the analog-to-digital converters. The quality of the word clock, measured directly at the converter, has an impact as well, but is beyond the scope of this paper. That said, due to the complexity of the topic, we'll be examining this in another paper in the near future.

As far as playback of the audio is concerned, there is more to consider. First, the digital-to-analog converter and clock source play a large part in sound quality, but any processing and mixing will have an impact as well. Let's assume that you're happy with the sound of your converters and proceed to looking at what happens once the audio is inside Pro Tools.

Let's begin by examining the capabilities of a system with only a 24-bit audio path from input to final output. A well-implemented 24-bit audio system will be able to deliver digital audio performance with a signal-to-noise ratio of 140 dB or better, which is significantly better, performance-wise, than possibly every converter currently available.

In order to briefly examine the relationship between bits and signal-to-noise ratio, you need only to know that in a non-dithered audio system, the signal-to-noise ratio is equal to the number of bits times 6.02, plus 1.76 dB. For example, a 24-bit, non-dithered system could have a theoretical signal-to-noise ratio of 146 dB. For practical purposes, it's not uncommon to think of 6 dB of signal equaling one bit.

As the signal level becomes lower, the amount of distortion increases. At very low signal levels, the distortion can become obvious. In order to reduce the distortion, a process called "dithering" is sometimes used. It is possible to inject a small amount of noise into the signal, called "dither," which forces the lower bits to become more active than they normally would be. The effect of this is to allow more audio information to be preserved or encoded into the noise floor. We'll look a bit into how this works soon.

With a dithered system using triangular (also called TPDF for *triangular probability distribution function*) dither, the signal-to-noise ratio equals the number of bits times 6.02, minus 3 dB. A 24-bit, TPDF-dithered system is theoretically capable of yielding a signal-to-noise ratio of 141.5 dB. Note the 4.5 dB difference between using and not using dither. We'll talk more about this ahead.

It's worth talking about distortion in a digital system as well. With a non-dithered system, the audio becomes more and more distorted as the signal level decreases. This means that the low-level detail that is so important to transparent audio can become lost, or worse, overrun by distortion products. This problem plagued early CD releases and contributed to digital audio's early "harsh" reputation, especially with wide dynamic range material where the average signal level is much lower.

With a properly dithered system, the distortion is *entirely* eliminated by the dither, with an increased, regular noise floor as the sole byproduct. In the case of using triangular dither, this is what the 4.5 dB noise floor increase mentioned earlier buys you. Not a bad trade when the noise we're talking about is more than 140 dB down!

Of course, modern converter technology still has a way to go to be able to keep up with these numbers, with the highest quality converters delivering somewhere around a 120 dB dynamic range. This is important to keep in mind because many of the noise and distortion artifacts produced in a 24-bit audio system are significantly below the noise floor of even the highest-quality converters.

So, is 24 bits enough? It depends. If you're talking about capturing and reproducing one channel of audio and that channel requires less than 140 dB of dynamic range, then yes, it probably is. But what about mixing several channels together, where some if not most of the channels will be attenuated quite a bit before they hit the mix bus?

This is the case with most modern mixer topologies where the summing bus operates at unity gain. If all of the channels were perfectly non-coherent and you added many channels together, the result would be that

every time you doubled the number of channels (from one to two, two to four, etc.) the audio output level would increase by 3 dB. If the channels were perfectly coherent (i.e., the same signal), the level would increase by 6 dB each time you doubled the number of inputs. Real-world signals from multi-miked orchestras, drum kits, or stereo pianos tend to fall somewhere in between.

Let's take an example with a 64-channel mixer. Sixty-four non-coherent inputs summed at unity gain would yield 18 dB more output level. In order to not clip the 24-bit mix bus, you would have to pull each fader down by 18 dB to compensate. Since we already know that a 6.02 dB level increase or decrease equals one more or less bit respectively, we can see that the result of our gain change exercise leaves us with a whole bunch of 21-bit signals.

Do 21-bit signals sound bad? That depends, too. With a high-quality converter having a noise floor of -120 dB, and a 21-bit non-dithered signal having a dynamic range of 127.76 dB, we can see that some of the artifacts produced will be below the range of the converter. Of course, it's not quite as simple as this. For one thing, higher-order distortion products can usually be quite easily heard. Also, we're doing this to a bunch of channels and the results are summed, again adding to audibility. Again, the material will make a difference with higher dynamic ranges faring worse.

So 24-bit mixing is probably not enough. What is enough, you ask? More! Lots more!

Dither and Double Precision Math

You can probably argue that each time you double the number of audio channels, you should probably have another bit's worth of math lying around to preserve the resolution and allow enough headroom for the combined result. A 128-channel mixer could theoretically require seven additional bits to preserve all of the low-level information while not clipping the output. If the faders have any additional gain above 0 (which is very common), this had better be accommodated as well. This means that the 24-bit bus should be expanded to have at least 32 bits to prevent signal loss or clipping.

What this doesn't take into account is the notion of pulling the channel faders down from unity gain or 0. While we're expanding the mix bus in the upward direction, we should also add bits at the bottom to preserve the signal quality when we pull the faders down.

How far down you want to be able to pull a fader and not lose information will dictate the number of bits you should add to the bottom. It's certainly not unheard of to have some faders at -30 or lower in a mix, so we could add five additional bits to keep the low-level resolution, leaving us with a 37-bit mixer.

Is this enough? Probably, but it's worth noting that Pro Tools uses a 48-bit mixing bus, allowing 128 channels to be summed at maximum gain with no clipping at one extreme and the ability to pull any or all of the channels' faders to lower than -80 dB and still keep the full 24-bit signal in the mix.

The DSPs used in a Pro Tools TDM system normally have 24-bit inputs. So how do we operate with these 48-bit signals? Simple. We often use a mode of operation called "Double Precision," which means that a 48-bit signal is chopped into two 24-bit words, which are operated on separately. Since the 24-bit DSP used has a 56-bit accumulator, keeping a 48-bit result around is not a problem. The hard thing is doing the math and piping the signals around.

Double-precision math requires more instruction cycles to process, so you could conclude that having an entire system built with double-precision math could be wasteful and overly expensive. It's probably best to keep double-precision math around only where it's needed and rely otherwise on the normal 24-bit, 140-plus dB dynamic range of the system.

Here are a few diagrams that describe Pro Tools' use of the 48-bit word in detail.

Diagram A ([click to view](#)) shows the distribution of the bits in a 48-bit word. Note the serious dynamic range available below the normal 24-bit word to preserve low-level detail. Also note the nearly 288 dB dynamic range of the mixer. There is no audible difference with the output signal if you pull a channel fader down and increase the master fader or decrease the master fader and increase the channel fader. This is because of the extreme dynamic range of mix bus, which preserves all of the audio data from each channel. Remember, the mixer has enough headroom to accommodate the signal from 128 full-level, fully coherent channels without clipping (as long as you pull the master fader down far enough to avoid clipping the output DAC or AES/EBU port, etc.).

Diagram B ([click to view](#)) shows roughly how the master fader works by effectively scaling up and down the 48-bit range. If a master fader is not used in the session, Pro Tools behaves as if there is a master fader behind the scenes set to unity gain. With the dithered mixer plug-in, high-quality dithering is used to deliver the final 24-bit output signal.

More on Dither

We mentioned before that the two most common methods of reducing the number of bits in an audio signal are simple truncation and the use of dither followed by truncation. Truncation means, in our 48-bit result example, that the system would "chop off" the lowest 24 bits, leaving the highest, or loudest 24 bits untouched.

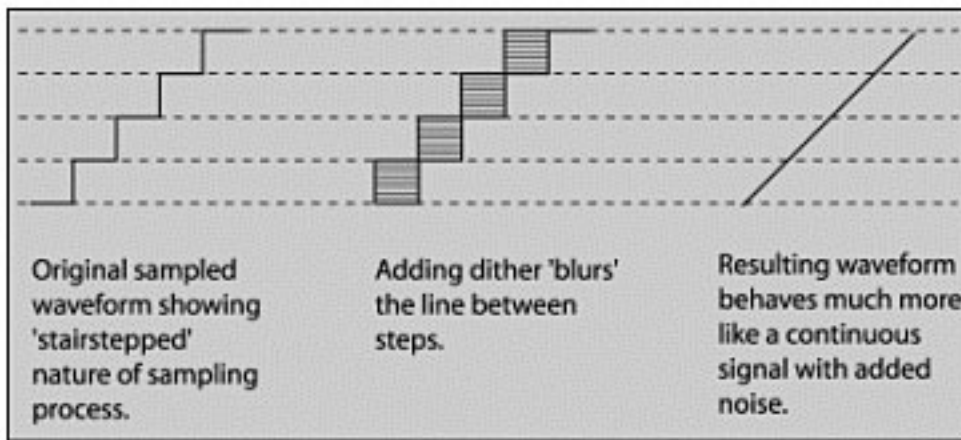
The result of this is distortion in the remaining signal or "quantization error". After the truncation and at higher signal levels, this "quantization error" resembles white noise. As the signal level drops, the noise becomes more correlated (related to the signal) and results in distortion being produced. In the case of a 24-bit system, this distortion is down around the 24th bit — or nearly 144 dB down from full-scale.

By adding dither, the distortion is de-correlated from the signal, which means it is eliminated at the cost of having a slightly increased noise floor. The noise tends to create significantly less havoc with the audio than the unbridled distortion.

One nice side note about summing multiple stages of non-correlated dither is that while two coherent channels of audio will sum to produce a 6 dB level increase, two channels of non-coherent noise will produce only 3 dB more noise.

Dither comes in many flavors and each requires differing amounts of processor overhead. Many of the more exotic dither algorithms produce "noise-shaped" dither. Noise shaping is the process of applying an EQ curve to the raw white dither noise to move some of the energy to a frequency range where our ears are less sensitive. In this way, the audible effects of the dither can be reduced while the distortion-reducing benefit stays exactly the same.

The picture below shows the dither process in terms of what happens to the audio waveform. The first image shows the original sampled signal. Adding dither to this signal is exactly like moving the threshold for each step around or making a "moving target" of each quantization level. If you use the correct level of dither, the quantization distortion is completely eliminated. The second image shows the "moving target" nature of the quantization levels after dither is applied.



The signal that results is exactly as though there were no quantization in the first place, albeit with some added noise (which is not shown.) The bottom line of dithering is that it tends to make the digital medium much more like analog, complete with a very low-level, analog-like noise floor. The proper use of dither entirely eliminates the notion of having quantized "steps." This is why it is difficult to relate bits to resolution with a well-implemented dithered system. It's really better to talk about noise floor and dynamic range.

The End of the Matter?

So, is that it? Not quite. Next time, we'll talk more about recording levels, plug-ins, and the path between the ADC output and the mixer input. We'll also discuss the various mixer plug-ins. I hope we've shown in this article that it is quite possible to mix entirely inside Pro Tools without an external mixer and maintain the highest possible audio quality. Careful listening tests have shown that the differences between mixing internally and externally are inaudible to everyone we've yet encountered.

There can be benefits to using an external console, including being able to use the console's EQ and dynamics as well as taking advantage of any euphonic artifacts produced along the way. For example, older Neve 80 series consoles with 200 series single-ended amplifiers have particularly nice-sounding distortion characteristics when the levels are pushed up.

I am not attempting to talk you out of using a console for these qualities. I am instead showing that there is no data, or information, lost while mixing completely in Pro Tools, that no signal is lost when you pull a channel fader down below unity, and that there is no difference between setting all of the channel faders to -30 with the master at unity and setting the master to -30 with the channel faders at unity. As long as you're not clipping the output and are sending the highest possible output signal before clipping, then all is well.

Note 1: *It is very difficult to set up an accurate mixing listening test; however, it's worthwhile and I highly recommend it. A future white paper will deal with the proper care and feeding of a mixing listening test. Many times, consoles that use VCA-based automation can be unreliable to use as a reference. The VCAs that control the channel's signal level tend to drift over time and we've found that many engineers are capable of rapidly spotting one mix from another when there are simple level differences on the order of 0.2 dB on individual instruments. It's easy to get thrown off during a listening test, only to discover that a level difference has crept in. Once the console is re-calibrated, the differences again go away. Frequently checking the signal levels during a listening test is one way to ensure accurate results. Also, in our past listening tests, we have avoided using any EQ or dynamics from the external console. We've used the console purely as a mixer with level controls, pan knobs, and a mix bus.*

Please forward any future column ideas to: tech_talk@digidesign.com.

The Pro Tools 48-bit mix bus yields close to 288 dB of available dynamic range.

BIT
47
46
45
44
43
42
41
40
39
38
37
36
35
34
33
32
31
30
29
28
27
26
25
24
23
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0

The top eight bits are preserved for headroom. 128 full-level, coherent audio channels would sum to bit 46. The last bit is available because there is an additional 6 dB of gain available from each channel fader. These eight bits provide close to 48 dB of available headroom above '0' before the mix bus clips.

One 24-bit channel of audio, with its fader set to '0', uses bits 16-39. When the master fader (if present) is set to '0', these bits also represent the main output, before any mix bus plug-in processing or dither is added. These 24 bits represent close to 144 dB of available dynamic range. A second full-scale channel with coherent audio (say, a pair of sine waves with identical frequency and phase) would sum to bit 40. In order to avoid clipping the mix bus, the user would either need to pull the channel faders down 6 dB or they would need to create a master fader, or use an Aux track as a sub-group, with the output fader set to -6 dB. In any of these three examples, the full precision of each channel is maintained and no "bits are lost".

The lower 16-bits of the mix bus are used to preserve audio data when the channel faders are pulled down below '0' or unity gain. These 16 bits allow a channel's fader to be pulled down to -96 dB and still contribute a full 24-bits of precision audio to the mix bus.

