

Better Latent Than Never – Some facts and fiction about latency Mike Rivers

To paraphrase author-humorist Mark Twain's comment about the weather, everyone talks about latency, but nobody does anything about it. Since most of today's recording musicians are using some sort of computer-based multitrack DAW, "latency" is the frequent topic of tutorials and on-line discussions. Everyone knows that low latency is good, but what constitutes "low latency," when is it important, and when isn't it?

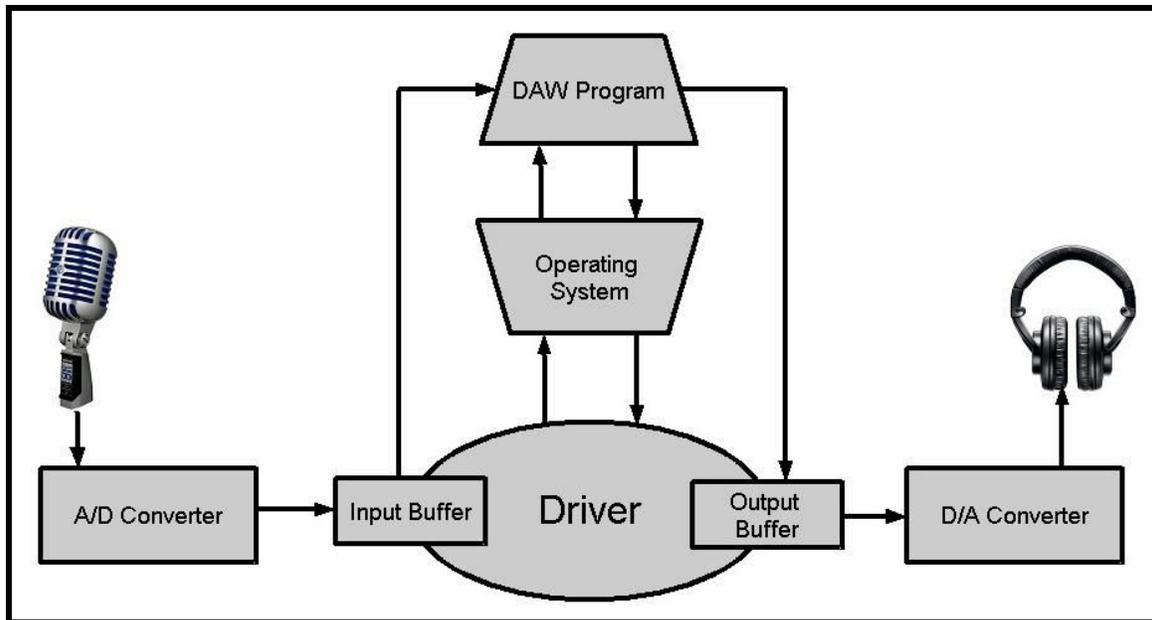
Latency is simply a time delay. Time delay is used to create audio effects such as echo or flanging, or for time-aligning tracks or loudspeakers. When delay becomes a nuisance, however, we call it "latency." In a DAW, latency can involve both audio and MIDI signals, or, when playing a soft synth, a combination of the two. While latency can appear at many places in a signal chain, the most cursed-at in a DAW system is in the monitor path between the microphone and headphones. With a few diversions, that's what we'll address here.

In the early days of computer-based audio production, the computer served primarily as a recorder and editor. Most first-generation users migrated from an analog recording setup which nearly always had a mixer which served as the hub of the system. This was how mics or instruments were routed, both to the computer sound card inputs and to headphones or monitor speakers. An analog signal path has essentially no delay, hence latency only became an issue when overdubbing.

Nearly all of today's computer-based DAWs include a software mixer. By integrating mixing with recording and editing functions, there's less wiring, less setup confusion for a novice, and lower cost. Like its analog hardware counterpart, the DAW's mixer can also provide a headphone feed for tracking or overdubbing. However, getting the signal through the whole system, from mic to monitor, takes some time. The time between when you sing a note and when you hear yourself in the phones is what we call monitoring latency, and this is generally the "latency annoyance" that we notice first when setting up a DAW. So what can we do to reduce DAW monitor latency?

The DAW Signal Path

To best attack latency, it's helpful to understand the various sources of delay along the signal path, determine what you can change, and what you can't. Here's a block diagram of the signal flow from mic to headphones in a basic DAW:



The analog-to-digital and digital-to-analog converters each contribute a fixed amount of delay. This is typically around 1.5 ms at 44.1 kHz sample rate. It's usually shorter at higher sample rates, though with modern DSP-based audio interfaces, this relationship isn't linear – if it's 1.5 ms at 44.1 ms, it will be longer than 0.7 ms at 96 kHz sample rate.

A driver is the interface between hardware and software. It passes data between the converters, operating system, and application program. Since the computer has other tasks in addition to handling your audio, input and output data are buffered to assure a steady flow in and out of the system.

The input buffer acts as a “holding tank” which fills up as data comes in from the A/D converter. The buffer communicates with the operating system and releases data when the processor is ready to use it. Similarly, the output buffer receives data from the processor in spurts and sends it out in a steady stream. Number crunching for straightforward audio recording and playback is quite fast, but signal processing plug-ins and virtual instruments require additional processing time, creating the need for larger buffers. Buffers must be large enough so that they never overflow, but the larger the buffer, the greater the delay. A buffer that's too small will result in audio clicks or stuttering. A buffer that's too large will haunt you with audible delays.

Buffer size is usually expressed as a number of samples. The number of samples buffered divided by the sample rate, plus the converter latency closely approximates the total system latency between input and output.

How Much Latency Is Too Much (or too little)?

It depends. When there's no need for input monitoring through the DAW, for example when mixing, or when recording a band playing on stage or live in the studio, latency is usually of no concern. Latency matters when the only way you can hear yourself is to listen to the output of the D/A converter.

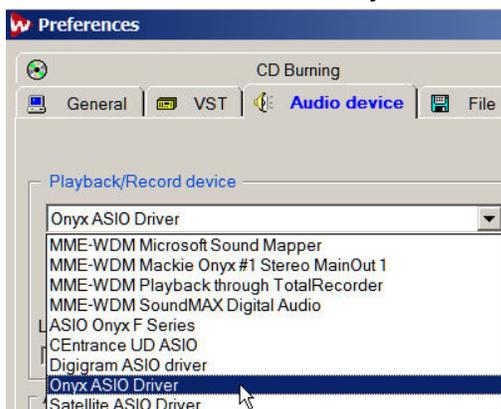
Most people aren't uncomfortable with a mic-to-headphone delay of less than ten milliseconds. Since sound travels through air at about one foot per millisecond, a ten millisecond delay is similar to playing your guitar ten feet away from the amplifier. Anyone can do that, but beyond 30 milliseconds, it starts sounding like an echo. On the other end, a very short delay (1-3 ms) can cause comb filtering, making your own voice sound odd in your headphones even though it sounds fine to the engineer over the control room monitors and when listening to playback. Some singers are really bothered by this, and it's a bad idea to try to adjust EQ on the vocal channel while listening to your own voice on headphones through a short delay.

So What Can You Do?

The simple answer to reducing monitor latency is to make the buffers smaller. It could be as simple as checking a "Use low latency monitoring" check box, though usually it's more involved than that. Since I'm a Windows user, these tips are Windows-centric, but the principles apply to any platform.

An out-of-the-box Windows computer has many things running that are superfluous for audio processing and which can be safely disabled or configured to start only when needed. Reducing the number of services and "startup" programs leaves more time and more memory for audio work. There are a number of web sites (see the Resources side bar) which, while a bit dated, offer good tips for improving audio streaming.

Make sure your software is up to date. Visit the manufacturer's web site and install the latest driver for your audio interface or computer's sound card. Be sure that the ASIO driver is selected in your DAW's audio device preferences setup menu.



All professional audio interfaces offer ASIO (Audio Stream Input Output) drivers with adjustable buffers and faster throughput than standard WDM (Windows Driver Model) or DirectX drivers. Most driver setup programs install WDM and/or DirectX drivers along with ASIO so that your pro

audio hardware can work with Windows multimedia applications which aren't aware of ASIO drivers.

I've never run across a built-in sound card, either one integrated on the motherboard or as an separate card that comes along with the computer, that has an ASIO driver. If you're not yet upgraded to a professional audio interface (and if you're serious about audio recording, you really should), try installing ASIO4ALL. It's an alternate (and free) driver for WDM audio devices that provides many ASIO advantages. It doesn't work with all hardware and software, but it's worth it a try and may allow you to reduce your system latency by a significant amount.

Demon of the Month Club

Computer solutions come and go as quick as a wink. Every few months someone will discover a new source of audio troubles and propose a solution. Once the word gets out, Some users attempting to implement the solution will be able to make an improvement while others won't. While such problems may eventually be resolved (often by attrition), the solutions seem to live on indefinitely. Google™ "VGA kills audio" and check the dates, but if you go on line looking for a solution to clicks and pops, people will still tell you to check "hardware acceleration" even though it no longer applies. I'd be remiss, though if I didn't mention (this was written in 2010) the latest demon to exorcise, the Deferred Procedure Call (DPC).

The DPC is a Windows operating system feature that lets high priority tasks such as input handlers to defer lower priority tasks until they've finished what they're doing. The DPC process allows a driver to get top priority for its most critical job of data input or output, and scheduling routine tasks like (in the case of an audio driver) controlling the traffic through the buffer with a lower priority. An impolite driver can grab large amounts of CPU time, and do it repeatedly, necessitating unusually large buffers to preserve smooth data flow.

A cool tool for navigating the busy-driver maze is the DPC Latency Checker (a free download). While it won't point directly to a guilty driver, by watching its display as you disable various devices, you can locate those which are the worst offenders, many of which you can live without, at least while you're working with audio.

Common DPC troublemakers are wired and wireless networking, virus scanners, screen savers, CD auto-play, E-mail clients checking for mail too often, Windows Zero Configuration Wireless, and, in laptops, APIC battery monitoring. The red bars in this screen shot show something getting attention for 4 milliseconds (4000 microseconds) once every 2 seconds. In this case, it was the driver for the wireless network interface, causing a lot of audio stuttering even with very large buffers. Often there will be only one significant troublemaker. Disabling the wireless driver was all that was required to get glitch-free performance with this computer.



Once you've sufficiently reduced the DPC latency (no red bars on the display), fire up the DAW and start reducing the ASIO buffer size. Pile on the tracks and plug-ins. Your goal is to find the smallest buffer that will let your DAW work as hard as it needs to without skipping. It's a cut-and-try process. It's also iterative – if you can't achieve sufficiently low latency with the smallest buffer that works, hunt for more things to disable. A 128 sample buffer is a good goal, but depending on the drivers and hardware, you may not be able to achieve it.

Work Smarter

Adjust the buffer size to suit the task at hand. Use the smallest workable buffers when tracking, use larger buffers for editing and mixing, where latency is irrelevant.

Defer using plug-ins that have substantial latency until mixdown. Use an amplifier or effect pedal instead of an amplifier simulator plug-in while tracking your guitar. For tracks that really need some help (a direct guitar, for example) in order to play against them, apply the amp simulator plug-in, get a working sound, then "freeze" the track with that sound. You may want to do this on a copy of the original track if your DAW program doesn't have an "un-freeze" option. Remove the plug-in, mute the original track in the mix, and monitor the frozen track instead. At mixdown, return to the original track, pile on the effects, and increase the buffer size as needed.

With virtual instruments the DAW *becomes* the instrument, so you have no choice but to monitor the DAW output. Since the input is MIDI, the A/D converter is idle. Your DAW may allow you to disable the the audio input, which may in turn

reduce latency. Once you've recorded a MIDI track, you can "freeze" the virtual instrument, freeing up some processing time.

Doubling the sample rate nearly halves the A/D and D/A conversion time. Although the computer works harder at 2x sample rate, which may require a larger buffer, by doubling the sample rate, you may end up with lower total latency. It's another avenue to explore.

Dedicating a computer to audio work is still valid advice. If you insist on being able to check your e-mail or watch YouTube videos while you're in session, you'll never achieve really low latency.

Alternate Solutions

An analog mixer offers true zero latency monitoring, with the bonus of flexible signal routing and control. Today's inexpensive mixers offer good sound quality as long as you're careful to avoid clipping. You might be short on table space or prefer to spend your money on another microphone or plug-in, but I still heartily recommend a mixer-based system. It resolves many issues beyond monitor latency.

Some audio interfaces include a built-in monitor mixer which is independent of the computer. In stereo interfaces with this feature, the monitor mixer is usually analog, resulting in real zero latency input monitoring. Multi-channel audio interfaces typically use a DSP-based digital mixer. Since the signal goes through the A/D and D/A converters on its way in and out of the DSP chip, converter latency plus the short DSP latency (typically 2-3 ms total) will be present, but it's consistent and requires no tweaking. Why use DSP? In the end, programming and parts cost are cheaper than building a multi-channel analog mixer.

Many mixers are now available with a built-in Firewire or USB interface which combine mic preamps, direct input monitoring, and computer I/O into a single unit. This is a great approach if you need a mixer for your band's PA and like to record your shows, though some of them are lacking in studio niceties.

When recording vocals, try the time honored technique of removing one earphone to hear your own voice acoustically. Many people sing better this way even when there's no latency problem. Remember to mute the output of the vocal track in the DAW monitor mix so you won't hear your vocal "doubled." When recording a virtual instrument track, listen to a built-in keyboard patch through a small powered speaker or use a small guitar amplifier rather than the virtual instrument you'll be using in the final mix.

Late Enough?

It's difficult to quit fooling with a computer. There are always improvements to be made. Hopefully some of these suggestions will help you to get more comfortable with monitoring while tracking. A comfortable musician is a better musician.

Resources:

Tips for optimizing a Windows PC:

Digital Producer Magazine:

http://www.digitalproducer.com/2002/11_nov/features/11_04/optimize_pcaudio.htm

Sound on Sound Magazine:

<http://www.soundonsound.com/sos/Mar02/articles/pcmusician0302.asp>

ASIO4ALL ASIO Driver for WDM sound cards (free download):

<http://www.asio4all.com/>

DPC Latency Checker (free download)

http://www.thesycon.de/deu/latency_check.shtml

How to disable Auto-Run in Windows XP (Microsoft Knowledge Base)

<http://support.microsoft.com/kb/953252>

An excellent article on using a mixer with a DAW by Ethan Winer:

<http://www.ethanwiner.com/mixer2daw.html>

What Latency Compensation Does (and Doesn't)

Today, "Latency compensation" is fairly common in DAW programs, sometimes so common that it's not even mentioned as a feature since (when it's working right) it makes the program work just as we expect it to.

In the early days of multitrack DAW recording, the main latency problem was that overdubbed parts always seemed to be out of sync with the original tracks. Since the previously recorded tracks you hear when overdubbing are delayed by the output latency, even though your timing is right on based on what you're hearing, the actual recording is a little late. The fix was to manually slide the newly recorded track to align it with the other tracks. Eventually DAW programmers figured out how to determine the amount of output latency and used this information to automatically adjust the time position of the new recording. Smart ones also determine a plug-in's delay and include that in the time adjustment. With software that offers latency compensation, if your tracks are out of sync, you need more practice.

The 2 ms Latency Conspiracy

There's a potential problem with latency in the range of 1-3 milliseconds. When listening to your own voice on headphones, there are two paths from your vocal cords to your eardrum – the natural path through your head and another path through the phones.

You're probably aware of the golden rule of mic placement: Avoid picking up a sound source with two mics at near-equal distances. When those mics are mixed, cancellation occurs at frequencies where the time difference between the mics results in a 180 degree phase difference between the two mic signals. This cancellation is called comb filtering because the frequency response of the system looks (sort of) like the teeth of a comb, with many sharp dips. The same thing happens when your natural voice and the latency-delayed headphone sound combine at your eardrum.

Your voice will sound odd to you because it's being comb-filtered, however, you'll sound normal when listening to the playback, and anyone else listening to the same headphone mix or the control room speakers won't notice anything unusual. This is because you're the only one hearing your voice through two paths. When your singer thinks his voice sounds odd in the phones even when you're using your most expensive tube mics, he just might not be nuts.

Some people are really sensitive about this, some don't notice, and some run the headphone volume so loud that the acoustic sound is swamped out and the comb "teeth" become negligible. Sensitive divas nearly always give a better performance when they're not worried about how their voice sounds, and a good performance trumps technology any time.

The monitoring latency through a typical built-in DSP mixer is in the range where the comb filter notches fall in places in the frequency spectrum that give a voice a lot of its character, around 430, 860, 1290, 1720 Hz and so on.

Measuring Your Latency

Try this simple test setup if you want to learn how much latency you actually have.

Connect the output of Channel 1 of your sound card or interface to the input of Channel 2. The input can be from a microphone (clap your hands or hit a drum), a metronome, or a drum machine. You want a sound that has a sharp attack. The headphones are optional if you want to be sure that things are working.

Open up a DAW program and set it up to record on two tracks, or a single stereo track. Turn Input Monitoring ON. If

you're recording two mono tracks, pan Channel 1 full left and Channel 2 full right. Start up your drum machine or make some noise and adjust levels so that you can see an input on both tracks.

If your DAW has automatic latency compensation, turn it off. You may need to find the latency compensation setting in a menu and manually set it to zero.

Record a couple of seconds. Stop, and zoom in very close. You'll see that Channel 2 is recorded a bit later than Channel 1

The selected portion of this display shows the time offset between the negative-going leading edge of this metronome click. The arrow in the lower right corner is pointing to the selection length in samples. The buffer setting is 64 samples, but the actual latency is 105 samples or 2.5 ms (44.1 kHz sample rate).

