

Live Sound International: Networked audio transport

SOURCE: April 2004 issue - POSTED: 05/20/05

Looking at the methods and factors

In the (still) emerging world of networked audio transport, there are two major categories in which a system may fall: a fully standards-based network, or a proprietary network that may or may not use standards-based transport.

Both have their advantages and disadvantages and, of course, are subject in varying degrees to the problems associated with transforming an analog signal into a digital stream of ones and zeros and then back again.

Let's first explore the biggest question that a system designer should ask when someone shows them a new piece of digital gear: "What is the latency as it relates to audio being transported in large networks?"

As you've already learned if you've read the related articles earlier in this issue of Live Sound, latency, or what some manufacturers call "propagation delay," is the amount of time that an audio signal is delayed due to digital processes including analog to digital conversions (A/D), digital to analog conversions (D/A), and digital signal processing (DSP).

For live sound applications, excessive amounts of latency can wreak havoc on the audience as well as the performers by creating listener fatigue and poorly reconstructed audio.

Generally speaking, the extent to which latency will cause a problem is a function of the ratio between the direct sound and the sound that is delayed. In a large-scale sound system, where there is little to no direct sound, the delay does not become a problem until there is a noticeable delay between sight and sound.

However, when the direct sound is within 8 dB or so of the delayed signal, 20 milliseconds (ms) to 30 ms difference will be audible. As for performers on stage, the acceptable time is generally shorter, especially when it comes to their monitors.

Echo from a system - whether it is acoustic or electronic - contributes to performer and audience fatigue. A performer using an in-ear personal monitoring system will be conscious of any latency in the system to as little as 5 ms to 10 ms. Small latencies on the order of 100 microseconds (μ s) can cause phasing problems that can result in high frequency roll-off when they are added back into the mix with an un-delayed source.

A common clock

It should be noted that all digital network components should be synched to a common clock. The best way to achieve this is with the use of a separate word clock. This will allow all devices in the chain to be locked to each other and thus eliminate any phase shifts that may occur between devices from un-synched sampling rates.

With advances in digital technology, latencies for audio gear have dropped dramatically, to the point that some digital mixers are below 3 ms for analog in to analog out with no added time for DSP. The problem lies in cascading multiple digital devices together if not digitally linked. When having to make A/D and D/A conversions for every piece of digital gear in a signal chain, it's easy to see how the latencies from conversion alone can add up to an unacceptable total.

Two key factors in comparing digital audio networks are the sampling rate and the bit resolution that the network supports. The sampling rate is the "rate" at which a digital device "samples" the composite analog sound waveform over time.

The sampling rate is important for the way digital audio can describe the frequencies in a sound. Bit depth, or resolution, describes the potential accuracy of a particular piece of hardware or software that processes audio data. In general, the more bits that are available, the more accurate the resulting output from the data being processed.

For example, audio recorded with a 48 kHz sampling rate and 24 bits of resolution will have 48,000 measurements of which there are 16.7 million different values that each measurement can be per second. Sampling rates and bit resolutions are important to digital audio networks because they need to remain consistent to keep the latency to a minimum. A sampling rate conversion typically takes the same amount of time as an A/D conversion.

The major limitation on the number of channels a network can support is bandwidth, which is the amount of information that can be sent down the chain at one time. The bandwidth required to pass one channel of audio varies with the sampling rate and the bit depth selected. As both increase, so too does the bandwidth demand reducing the maximum number of channels which can be transported over a particular network's architecture.

Well, what if you could have just one A/D conversion at the start of the signal chain and one D/A at the end with all the other gear still in the chain? Now we're talking about the digital audio networks. Again, there are two groups of these digital audio networks, so let's take a closer look at both.

Published rules

A fully standards-based network abides by a published set of rules, which specifies a recommended interface for serial digital transmissions. This includes the description of data format for transport and the method of transport.

The best part of a standards-based audio network is that any manufacturer can implement the format into their products without having to pay the often-expensive licensing fees of a proprietary solution. This allows the use of "manufacturer A's" reverb unit with "manufacturer B's" digital console without the conversion to analog, resulting in a total system latency that is much lower than if a D/A and A/D conversion was necessary.

A look at digital transport for Eric Clapton on tour. Spectrum Sound implemented Soundweb DSP and networking at FOH, with audio routed via CAT5 to QSC RAVE A/D converters with CobraNet technology. Then it's a quick flight for the signal on fiber optic to RAVE D/A converters at the amplifier racks, with another short, CAT5 hop to the system's four-channel CyberLogic amps.

The most common standard is AES3, more often called the AES/EBU (Audio Engineering Society/European Broadcasting Union) standard. AES3 uses a 110-ohm shielded twisted pair cable to send two channels up to 300 feet. The standard allows up to a 24-bit resolution with no maximum sample rate.

Another standard developed by the AES is AES10, more commonly called MADI (Multichannel Audio Digital Interface). MADI offers 64 channels at a 48 kHz sample rate and 32 channels at a 96 kHz sample rate with a resolution of up to 24 bits per channel. Transmission over a single 75-ohm coaxial cable has a limitation of 150 feet, but the use of fiber-optic cables can extend the length to two miles.

Other multichannel standards include ADAT Optical (ADI), Sony/Phillips Digital Interface (S/PDIF), and Tascam Digital Interface (TDIF).

The only latency added into standards-based digital audio distribution is the act of making the A/D and D/A conversions and any subsequent DSP that occurs. This is unlike some proprietary solutions that also require additional time to transcode and transport the data.

Viable & affordable

The two most common proprietary systems, which use a standards-based transport, are CobraNet from Peak Audio, and EtherSound from Digigram. These systems transport audio data using standard IEEE 802.3 Ethernet protocols. Ethernet is used because it's relatively inexpensive, reliable, and the technology has been, and will continue to be, upgraded by the computer industry.

Just a few years ago, 10baseT was the best thing going. Now, gigabit (100baseT) Ethernet has become viable and affordable. In order to integrate either of these technologies, a manufacturer must license CobraNet or EtherSound technologies from the developer in order to use them in their product(s).

CobraNet is capable of up to 64 bi-directional channels at a 48 kHz sample rate with a resolution of up to 20 bits per channel over a single 100Mbit link. This could mean that all you have to pull out to the front of house is a single CAT5 cable.

However, CobraNet comes with a healthy dose of latency, 5.33 ms just to transport the signal from one CobraNet device to another. The reason for this delay is due to the 256 samples required to buffer the audio data into Ethernet packets. (The latency due to A/D and D/A conversions must still be calculated into the equation to get to total delay.)

CobraNet is still in its growing stages, and is beginning to be implemented by more manufacturers into different product types including amplifiers, speakers and DSP boxes. However, currently no mixing consoles, digital or otherwise, offer integrated CobraNet. Therefore, a run out to the front of house mix position is going to require two networks, adding more than 10 ms of latency before even thinking about any other digital processing.

This is one example of why cascading digital equipment can lead to unacceptable latency totals. However, note that consoles with integrated CobraNet may be in the not-too-distant future with both Midas' parent company Telex, as well as Mackie, becoming recent licensees of the technology.

Off the shelf

EtherSound is different from CobraNet in that audio only flows in one direction. It offers 64 channels of 48 kHz sample rate at 24-bit resolution. The same off-the-shelf Ethernet switches and cabling can be used for distribution, but any audio that is added into the signal chain is only available to be peeled back off down stream of where it was added.

At the 2003 New York AES convention, Digigram announced that it will be offering a bi-directional EtherSound architecture. Nevertheless, there is still a trade-off. The system can only operate using the daisy-chained "ring" network topology or in direct transmitter to receiver applications. The latency in an EtherSound network is six samples, with .00122 ms added for each EtherSound device in the chain. EtherSound's licensees include InnovaSON, Fostex, and Nexo.

Another lesser-known network is the Media-accelerated Global Information Carrier (MaGIC) developed by Gibson Labs, the technology division of Gibson Guitar Corp. MaGIC grew from the development of the digital electric guitar. Like CobraNet and EtherSound, MaGIC conforms to IEEE 802.3 physical layer and offers up to 32 bi-directional 32-bit channels with up to a 192 kHz sample rate.

The problem with Ethernet is that it is a non-deterministic system, meaning that the data will arrive when it feels like it. CobraNet and EtherSound have developed systems that make the arrival times very predictable and allow for the network to be synchronized with only a small margin of error.

For instance, CobraNet's master unit regularly broadcasts beat packets onto the network either from its internal clock or an external master clock. Other devices on the network lock onto the arrival time of this packet and regenerate the clock locally. The error in clock delivery is $\pm 1/4$ sample period; this translates to about .005 ms at a 48 kHz sampling rate.

While digital seems to be the future of audio networks, it still requires attention to detail and proper setup. Monitoring overall system latency and keeping a consistent sample rate and bit resolution are both new requirements for the digital age. The answer to the question of what system is best for your situation is still the one that best meets your needs, which might be good old-fashioned analog.

Digital is just another hammer in the sound designer's tool belt which can either drive the nail home if used correctly or smash a thumb if its requirements are ignored.

David A. McNell, CTS

David McNell is an AV engineer working in the Special Technologies Group of Newcomb & Boyd, Atlanta. He is a Certified Technology Specialist (CTS) and an EIT