Serial Digital Audio Routing Switchers?

When updating a facility to digital, one of the first things to consider is replacing the old patch bays with a centrally located routing switcher. There are many reasons for considering the routing switcher first, for example, improved signal integrity, no mechanical contacts to wear out due to constantly plugging and unplugging patch cables, distributed control of the system, automatic audio follow video switching, and the ability to share equipment throughout a facility. This paper deals with the step-child of a typical digital video installation namely, the study of serial digital audio.

Serial digital audio streams come in many forms. First the audio sample rate could be either 8Khz, 32Khz, 44.1Khz, 48Khz, or 96Khz. Then there is the number of bits per audio sample, which are commonly 8, 16, 20 or 24. Do not forget the encoding format. Is the audio stream AES/EBU, SP-DIF, AC-3, or could it be some other standard? And finally, since we are exploring routers, is the facility's requirement for a traditional matrix, a multi-stage matrix, or for a time division multiplexed system?

All of these considerations make choosing a central digital audio router confusing to say the least. In order to make an informed decision, stock must be taken of exactly what equipment exists in the facility currently, while trying to keep in mind what equipment may be added in the future. The first thing to examine is the type of audio data streams that will be passed through the routing switcher.

- **Taking apart the audio stream**

A digital audio stream is a bit-serial communications signal for transmitting audio data through a single transmission line. The stream may be sent down a single piece of shielded cable or differentially using a shielded twisted pair cable.

The AES/EBU standard calls for a differential stream terminated at both ends with 110 ohm (which is the characteristic impedance for typical shielded twisted pair cables). SP-DIF is the consumer standard that uses coax or equivalent single conductor shielded cable terminated with 75 ohms. Although the physical characteristics of the transmission media is different for these two standards, the data in the audio stream is very similar.

- **AES/EBU and SP-DIF streams**

The AES/EBU and SP-DIF streams each provide for two channels of digital audio (each channel can have up to 24 bits per sample) along with synchronization, control, status, and error detection data. The stream is encoded using a bi-phase mark method. Bi-phase encoding is used to minimize the DC content of the transmitted signal and also allows clocking information to be included along with the data, so the receiver may lock to and extract audio data without requiring a clock signal to be sent separately.

In both of these standards, the audio stream is made up of differing groups of data called blocks, frames, and sub-frames. Each block, frame, and sub-frame begins with a unique data pattern called a preamble that is used by the receiver to synchronize the data stream. One sample of a single channel of audio data is encoded into a sub-frame. A frame consists of two sub-frames, one audio sample for the right channel and one for the left channel. A frame, therefore, can be thought of as a single sample of a stereo audio signal. Finally, a block consists of a group of 192 frames.

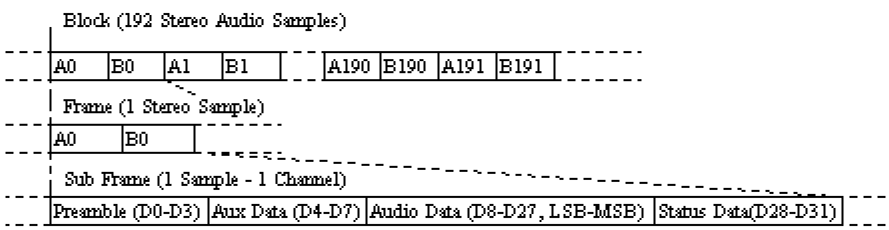


Fig. 1 AES/EBU audio data stream expanded to show its different parts.

Above is a drawing of an AES/EBU audio data stream expanded out to show its different parts.

In addition to the audio data, each sub-frame contains extra data which is accumulated over a block. This accumulation results in 24 bytes of channel status data. This data is used to carry additional information about the audio data. One of the channel status bits in the stream is used to determine whether this is a professional or a consumer audio stream. Typically an AES/EBU stream is used in professional applications while a SP-DIF stream is used for consumer applications. The professional/consumer bit defines the meaning of the rest of the status bytes which allows different information to be included in the status of these two standards. Another bit, which is common to both professional and consumer streams, is an audio valid bit which is set to 0 if the stream is either AES/EBU or SP-DIF, and set to 1 if the stream contains some other type of data such as encoded AC-3.

Including the preamble, auxiliary data, the audio sample and the status bits, each sub-frame contains the equivalent of 32 bits of data. Because of the bi-phase encoding method, the maximum fundamental frequency component can be calculated to be 64 times the sample rate. Thus a digital audio stream sampled at 32Khz creates a 2.048Mhz pattern, while a 48Khz sample rate creates a 3.072Mhz pattern and a 96Khz sample rate creates a stream at 6.144Mhz. As can be seen, audio is not a low frequency signal.

- **Digital Dolby - AC-3**

One commonly used digital audio standard is called Dolby AC-3. This is a standard used to encode 5.1 channels of audio at a compressed data rate into a single audio stream. It was originally developed for use in films and has subsequently been chosen for use in the transmission of digital TV signals to the home.

For those not familiar with the term 5.1, this refers to several individual audio channels that can be used as left front, right front, center front, right rear, left rear, and base channels. Because the base channel is bandwidth limited to very low frequencies while the other 5 channels are full bandwidth audio channels, the base channel is called .1.

It should be noted that the 5 full bandwidth audio channels do not need to contain related information. Each channel is fully independent of the others. So, for instance, 4 different languages could be sent using a single AC-3 audio stream.

AC-3 uses a very low bit rate compression algorithm. The bit rate was determined by space available on film to place digital audio data in an optical form. As a result the bit rate of an AC-3 stream is 320Khz.

- **So how should these different digital audio streams be routed?**

There are several methods used in today's digital audio routers to get the inputs to the outputs. Each of these methods has its advantages and disadvantages. The requirements of the installed location must be examined and understood in order to make an intelligent decision to determine which type of router is best for the application.

- **Traditional routing matrix**

Every type of routing system produced today has at its roots research done at Bell Labs for telephone switching. However, requirements for telephone routing and audio (or video) routing are different. A traditional routing matrix is the most common form of audio/video router today. This design consists of X inputs and Y outputs with X•Y switches allowing any X to be sent to any Y. The input streams need not be synchronized with each other.

This type of router is great for smaller systems, but as the number of inputs and outputs grow the number of switches required becomes large. For example a 32x32 matrix required 1024 switches, a 64x64 requires 4096, and a 128x128 requires 16384 switches. As the number of switches grows, so does the cost of the router.

In addition, the physical design of the interconnections in a very large router becomes quite a challenge.

- **Time division multiplexing**

This type of switching router doesn't use switches at all. Instead, there is a common global signal highway that carries data continuously. Each and every input signal has its own assigned time to occupy the highway, and, to make a route, a receiver is designed to pick off the correct input signal from the highway. This, of course, means that every input stream is synchronized to a reference frequency that is based on the data rate of the highway.

The problem with this matrix comes when a very large number of inputs is required. The data rate of the highway is equal to the sample rate of the audio times the number of inputs to be put onto the highway. A 64x64 router passing 48Khz audio data requires a 3.072Mhz highway. To make a 1028x1028 router, the highway must run at 49.344Mhz.

A time division multiplexed router is very good for small matrices and where the audio signal is being processed as in a mixing console. But the function of a router is to pass signals unaltered where they can be processed elsewhere, and this is not the most cost efficient solution.

- **Multi-stage routing matrix**

A multi-stage router is comprised of several smaller routers connected in such a way as to minimize the number of crosspoint switches needed. There has been a lot of research done on this type of router by the telephone company. Like a traditional matrix, the input streams need not be synchronized if this is not a requirement for the installation.

The main problem is that in most multi-stage router designs it is possible to find a route that cannot be made. This condition is called a blocked route, and can be economically devastating in a typical digital audio installation.

Previous solutions to the problem of blocking have not sufficiently realized the impact of multi-casting sources vs. the telephone situation where only single-casting of sources is allowed. After several years of engineering development, Sierra Video Systems has devised a unique new design (Patent Pending) which uses a redundant multi-stage design with multi-casting to solve the problem of blocked routes while providing path redundancy at the same time.

- **Data unaware routing**

The simplest form of a digital audio router is one that is not aware of the data being routed through it. This type of router consists of a number of: 1) input buffers each with the correct termination impedance, 2) output drivers also with the correct termination impedance, and 3) crosspoint matrix of some kind in the middle. This type of router is the direct equivalent to a patch panel, but allows connections to be made electronically from control panels that can be in remote locations away from the router itself.

This type of router can be designed to handle a broad range of data rates, so the same router can switch 32Khz, 44.1Khz, or 48Khz AES/EBU data streams as well as AC-3 or other compressed data streams with no hardware changes. The format of the signal, however, must be the same on both the source and the destination of the route. For instance, if an AC-3 stream is routed to a digital beta VTR, the stream will probably not be handled properly at the destination.

The disadvantage of this type of router is that there is no synchronization between the input streams, so the switching between any sources is done at any time during the data streams. The biggest advantages of a pure data router are the low cost and simplicity of the router and the fact that any audio stream (AES/EBU, AC-3, telephone,...) can be simultaneously passed transparently through the router requiring no hardware or software tricks.

- **Clickless/popless switching (word boundary switching)**

The current buzz word in digital audio routing is clickless/popless switching. This term implies that the router will switch between any two digital audio sources without causing a click or a pop to be heard when monitoring the output through a DAC and a loud speaker. Clickless/popless switching is really a marketing breakthrough which has been misused by several router manufacturers.

In order to implement clickless/popless switching, the router is required to synchronize and decode the audio streams checking the value of each and every sample. When a switch is desired, the router must wait until the first audio signal goes through 0 volts, then substitute a 0 volt pattern until the signal that is being switched to goes through 0 volts in the same direction, and ***then*** switch to the new audio stream. This means that each channel of the stereo sample will actually be switched at a different time unless their signals cross 0 at exactly the same time!

In reality, this is **not** how a clickless/popless router is implemented. Instead, the router uses a synchronizing clock to time align all of the audio streams coming into the router and does the switch between two sources on a frame or block boundary. This results in a continuous stream of data that a downstream decoder will remain locked to during the switch from one source to another.

To examine this, let's look at the analog audio waveforms that are equivalent to the digital audio streams being routed. The first example is word boundary switching (Fig 2), while the second example shows true clickless/popless switching (Fig 3).

Each of these examples assumes that the audio data streams have already been synchronized; the top left waveform (Fig 2A and 3A) represents one input to the router, while the bottom left waveform (Fig 2B and 3B) represents another router input. The waveform on the right (Fig 2C and 3C) represents the output of the router after a switch has been made. The time of each sample is represented by vertical dashed lines though all of the waveforms.

It can be seen that switching on a digital word boundary alone can create a large transient which is then heard as a pop or a click when reproduced through a loud speaker. It can also be seen that true clickless/popless switching can be done, but more than just the digital word boundary must be taken into account.

Fig. 2 Word boundary digital audio switching

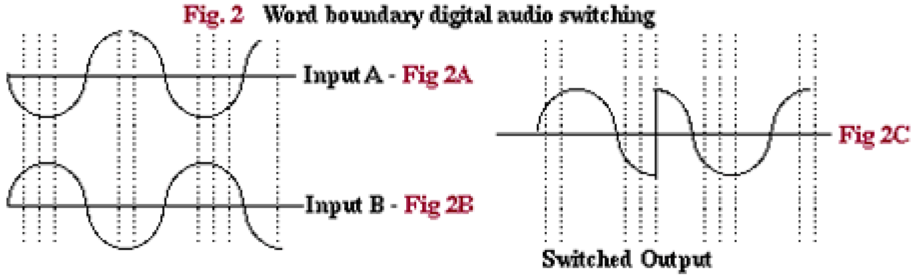
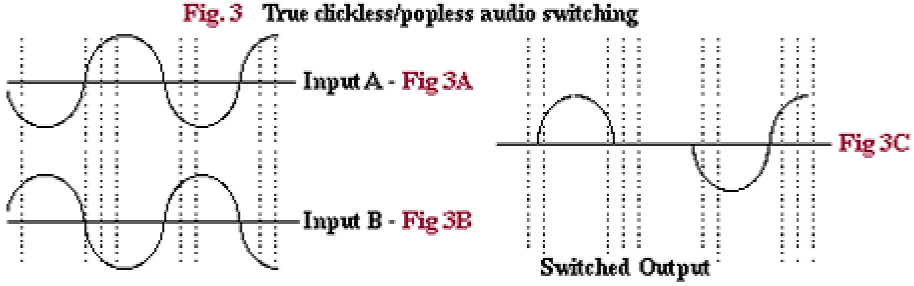


Fig. 3 True clickless/popless audio switching



- **Data rate conversion**

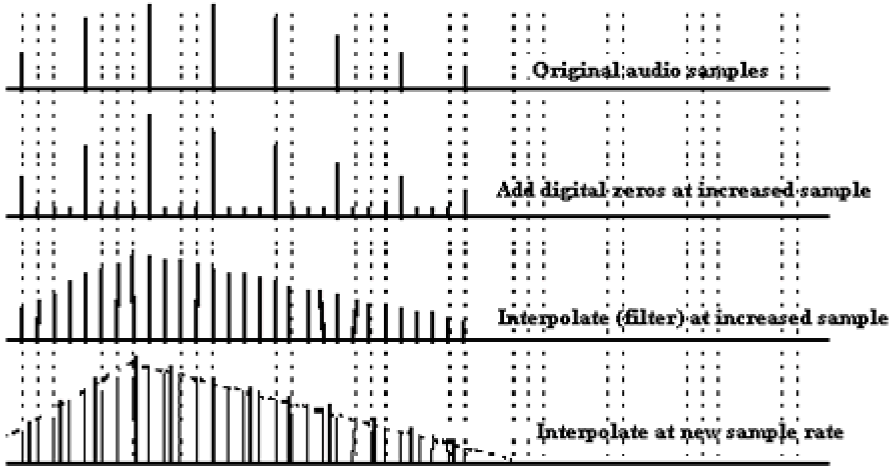
Word boundary switching can only occur if the sources being switched have the same data rates. This is fine if the entire facility uses the same type of equipment and all is referenced to a house sync just like a video signal. But what happens if you try to switch between a professional digital audio tape recording using a 48Khz sampling rate, and a CD player which uses a sampling rate of 44.1Khz? You could choose a data router as described above which doesn't care what the streams consists of and allow the equipment downstream of the router deal with the change in sample rates. Another approach is to do a sample rate conversion to the incoming data stream and synchronize the audio data, so switching can be done on a frame boundary.

A data rate converter is basically a DSP filtering process. In one method of sample rate conversion (the interpolation/decimation method), the data is converted to a series of samples at an intermediate data rate which is much higher than either the original sample or the final rates and then down-converting from the newly created data stream.

For instance, using an input sample rate of 48Khz the filter uses an up-sample rate of 4 times the original or 192Khz. In this example (shown below) every 4th sample (at the higher rate) is first filled in with one of the original data values with the other 3 samples being filled with the value 0. This new data stream is fed to a FIR type digital filter running at the intermediate sample rate which calculates yet another new data stream with the interpolated in-between values in place of the original 0s. This new data stream is sample-rate reduced to the final sample rate by using interpolated values from the surrounding samples and creating the output data stream.

In practice, the interpolation/decimation method of sample rate conversion uses a much higher intermediate sample rate than shown in this example and a very large FIR filter (thousands of taps) to reduce the errors in the calculated in-between values. While this may seem like a brute force approach, digital audio rate converter IC's with outstanding performance (distortion and noise products below -94dB) have recently become available, making this solution possible and affordable.

Fig. 4 Interpolation/Decimation method of sample rate conversion



- **The Sierra Video System solution - you get to pick**

The SVS families of digital audio routers have been designed to be as flexible as possible. Based on the customer's needs, a simple (and inexpensive) data router may be all that is required, or a full blown word boundary switching router with data rate conversion may be needed. The Sierra Video Systems design is based on different input modules that are either synchronizing inputs or not (based on the requirements of the system), smart crosspoint circuitry, and high capacity output drivers. The input modules may be mixed within a router, allowing word boundary switching for some of the inputs and not for others.

All of Sierra Video System's synchronizing input modules have a built in sample rate converter on each input. This allows synchronization on all of the inputs (as required for word boundary switching) in addition to sample rate conversion when required. These modules do not care what the data rate is of the input audio stream. The output audio data will *always* be synchronized to the house reference data rate. Additionally the input data rate converters may be switched to be bypassed (either automatically or through user setup) to allow AC-3 and other forms of non AES/EBU data streams to be passed through the router unaltered.

The type of routing matrix in a Sierra Video digital audio router will either be a traditional XY matrix or a full orthogonal multi-stage based on the number of inputs and outputs required. Both of these matrices can be switched synchronously or asynchronously. These matrices allow for the best price possible for a given system's requirements. See the Yosemite and Shasta families.

- **Let's not forget the rest of the system!**

In addition, SVS digital audio routers are designed to be part of much larger systems which may contain digital video, analog video, analog audio, and RS-422 routing. The entire facility can have a Sierra Video System router containing any or all of these layers switched through a network of control panels that have a common design. Operators only need to learn how to drive one type of router control system for a large (or small) mixed signal facility, and the router can be tailored to the exact requirements of the installation.

Jay Baker

About the Author: Jay S. Baker brings to Sierra Video systems twenty-four years of experience designing professional broadcast equipment. Prior to joining SVS, he held key positions at Ampex, Dolby Labs, Sony, Grass Valley Group, and Abekas. jbaker@sierravideo.com.