The Video Engineer's Guide to Digital Audio

John Watkinson

John Watkinson is an independent consultant and author in the broadcast industry with over 20 years experience in digital and analog technology. With a BSc (Hons) in Electronic Engineering and an MSc in Sound and Vibration, he has held teaching posts at a senior level with the Digital Equipment Corporation, Sony Broadcast and Ampex before forming his own consultancy.

He regularly delivers papers at conferences including AES, NAB, SMPTE, IEE, ITS and IBC and has written a number of definitive books including "The Art of Digital Audio".

An NVISION® Guide

Publishers Forward

Our thanks to John Watkinson for agreeing to write this book for us. John's work has helped video and audio engineers worldwide to better understand the techniques required to make a safe transition to the digital environment. At NVISION, we have endeavored to design and manufacture product that ensures that the inclusion of digital audio within a television facility can be achieved, painlessly and cost effectively.

The most important task in designing a digital audio system within a video facility, is understanding the importance of issues that were irrelevant in the analog domain. It is our belief that this new book will be an invaluable addition to every engineers toolbox and will help us all to avoid the pitfalls that exist on the road to digital.

Nigel Spratling
Vice President - Marketing
NVISION®

ISBN 0-9640361-3-4

CONTENTS

INTRODUCTION

This guide is intended for the television engineer faced with the problem of integrating digital audio systems in a television environment. Digital audio is a complex subject which is often described by specialists in a way which leaves many of us confused. This guide is not like that. It is for normal people who have a job to do and need to make decisions without getting lost in pages of math. So there are no equations here, only clear explanations. And there's a Glossary where all the buzzwords are defined.

Television production equipment is changing over to digital technology at a rate which is accelerating. The advantages of digital apply equally well to audio as they do to video. In fact, retaining an analog audio system can prevent the full freedom of digital video being realized.

There are many quality related reasons why digital audio is advantageous, and it was the search for quality which led to the development of the technology in the first place. However, the main pressure today is economic. Essentially a sensibly planned digital audio installation costs less to run than an analog plant doing the same job. It will have less down time and need less maintenance. Alongside the economic advantage, the excellent sound quality is a useful bonus.

Television sound quality is becoming increasingly important. With the expectations of the viewer set by his Compact Disc system, TV sound has to match up. More and more people are equipping their TV sets with quality speakers, often in surround sound systems. In some countries, television broadcasts are already accompanied by NICAM digital stereo sound. Before long, Digital Video Broadcasting (DVB) will become the preferred way to deliver digital audio and pictures to the home.

Against this background, going digital in some way will be essential in order to remain competitive, the only remaining question being one of degree. Whilst a brand new facility can start out with a totally digital system, in most installations there is an investment in existing equipment and engineers who can handle it. A total changeover to digital makes no sense and one of the jobs of this guide is to illustrate how the replacement of a few key items could offer a more cost effective solution in which digital and analog co-exist for a transition period. In the real world, there is seldom one universal right answer, and where possible we present more than one solution to a given problem.

Chapter 1

What is Digital Audio?

1.1 Although there are other approaches, there is one system, known as Pulse Code Modulation (PCM) which is in virtually universal use for professional audio production. Fig.1.1 shows how PCM works. The time axis is represented in a discrete, or stepwise manner and the waveform is carried by measurement at regular intervals. This process is called sampling and the frequency with which samples are taken is called the sampling rate or sampling frequency Fs. The sampling rate is fixed and the sampling clock is jitter-free so that every sample will be made at an exactly even time step. If there is any subsequent timebase error, the instants at which samples arrive will be changed but the effect can be eliminated by storing the samples temporarily in a memory and reading them out using a stable, locally generated clock. This process is called timebase correction and all properly engineered digital audio systems must use it. Clearly timebase error is not reduced; it is totally eliminated. A further consequence of this approach is that the audio channels of a DVTR don't have any wow or flutter.

The sampling process seems to be taking no notice of what happened between the samples. This is not true because all analog signal sources from microphones, tape decks, pickup cartridges and so on have a frequency response limit, as indeed do our ears. When a signal has finite bandwidth, the rate at which it can change is limited, and the way in which it changes becomes predictable. When a waveform can only change between samples in one way, nothing has been lost by sampling.

In a sense sampling has made audio more like video, which is sampled into discrete frames. One consequence of using digital audio with video is that we now have to deal with two sampled signals. It is clear that for practical purposes some sensible synchronization has to be achieved between the two rates. A whole section of this guide is devoted to synchronization.

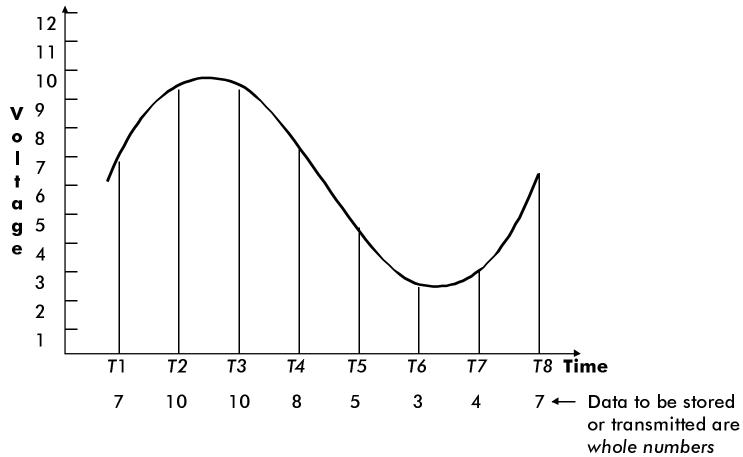


Figure 1.1 In pulse code modulation (PCM) the analog waveform is measured periodically at the sampling rate. The voltage (represented here by the height) of each sample is then described as a whole number. The whole numbers are stored or transmitted rather than the waveform itself.

Fig.1.1 also shows that each sample is also discrete, or represented in a stepwise manner. The length of the sample, which will be proportional to the voltage of the audio waveform, is represented by a whole number. This process is known as quantizing and results in an approximation, but the size of the error can be controlled until it is negligible. The advantage of using whole numbers is that they are not prone to drift. If a whole number can be carried from one place to another without numerical error, it has not changed at all. By describing audio waveforms numerically, the original information has been expressed in a way which is better able to resist unwanted changes in transmission.

Essentially, digital audio carries the original waveform numerically. The number of the sample is an analog of time, and the magnitude of the sample is an analog of the pressure (or the velocity) at the microphone. In simple terms, the audio waveform is conveyed in a digital system as if the voltage had been measured at regular intervals with a digital meter and the readings transmitted. The rate at which the measurements were taken and the accuracy of the meter are the only factors which determine the quality. The quality

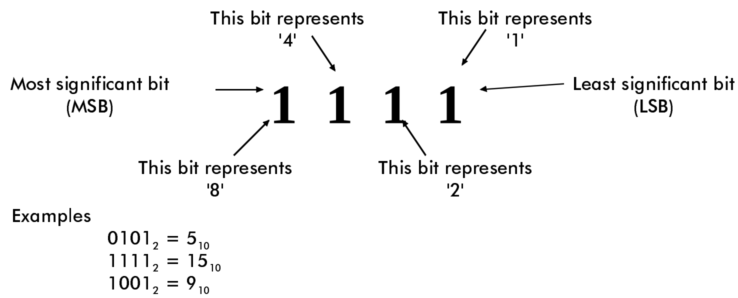


Figure 1.2 In a binary number, the digits represent increasing powers of two from the LSB. Also defined here are MSB and wordlength. When wordlength is 8 bits, the word is a byte. Binary numbers are used as memory

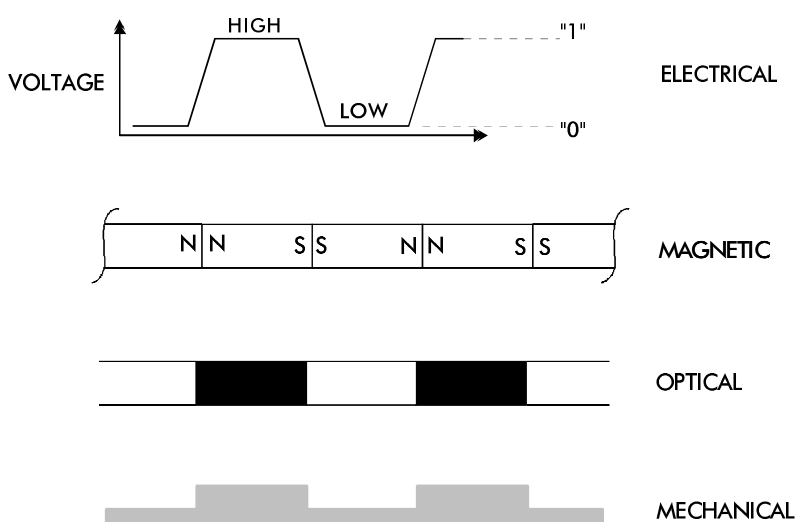


Figure 1.3 Several ways of handling binary data are shown here. In electronic circuits, two different voltages are needed, commonly achieved with a simple on/off switch. In magnetic recording, two flux directions are used. Optical recording may use alternative opaque or dark areas and mechanical storage may use raised or depressed areas on a surface.

is determined only by the accuracy of conversion and is independent of the quality of the signal path.

1.2 Digital transmission has the advantage over analog in that it can reject noise and jitter. If the whole numbers representing the video waveform voltage are converted into binary, using the conversion table shown in Fig.1.2, the binary digits, or bits, have only two states, 1 and 0. These two states can then be represented by any electrical, magnetic, optical or mechanical system which can exist in two states. Fig.1.3 shows binary represented electrically by two

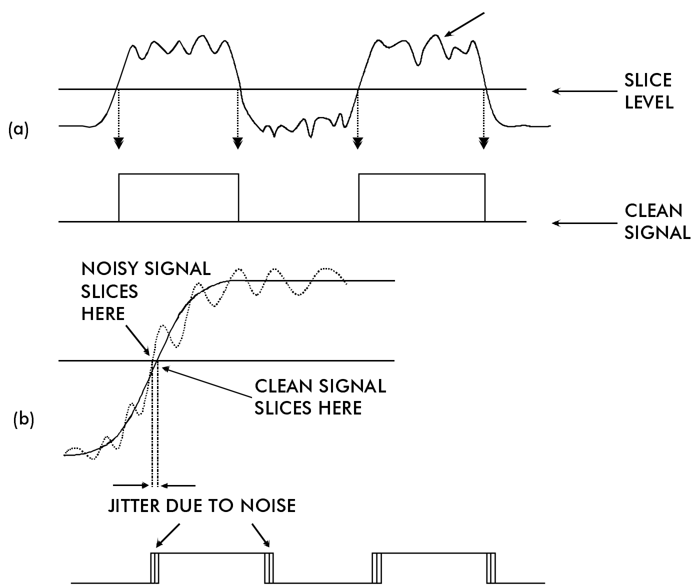


Figure 1.4 A Binary signal has the advantage that a simple slicer shown at (a) can be used to recreate a square signal. The result is that noise only causes jitter as shown in (b). This can be removed with a well designed

different voltages, magnetically by the reversal of the direction of magnetization, optically by alternate opaque and transparent areas of a recording, and mechanically by the presence of pits in the surface of a laser disk.

With only two states, more noise can be rejected than in any other system as Fig.1.4a) illustrates using the example of an electrical interface. Although the signal transmitted is a clean, two-level waveform, by the time it reaches the receiver it will be suffering from noise and jitter. The receiver compares the voltage of the signal with a reference which is mid-way between the transmitted levels in a process called slicing. Any voltage above the slicing level is considered a 1 and any voltage below is considered a 0. This slicing process will reject considerable amounts of noise and restore the signal to clean binary once more.

Slicing resists noise, but it is powerless against jitter which requires a further process. Fig.1.4b) shows that jitter has the effect of shifting voltage changes, or transitions, along the time axis in a random manner. However, the average timing of a large number of transitions is unaffected. A phase-locked loop (PLL) is an electronic circuit which can average the timing of many transitions to recreate a stable clock from a jittery one. It acts like the flywheel on a piston engine which averages the impulses from the pistons to produce smooth rotation.

The combination of a slicer and a phase-locked loop is called a reclocker. The waveform leaving a reclocker is identical to the waveform at the transmitter. Noise and jitter have been rejected so there has been no loss of data due to the transmission. Consequently, reclockers can be cascaded or tandem connected as often as necessary yet the same data will come out of the end.

The reclocker is found everywhere in digital systems. Routers use it, Digital VTRs use it in the video and audio signals. Hard disk drives use it, CD players and DAT machines use it. The universality of reclocking is because it eliminates generation loss. A device which contains a reclocker launches a clean signal which is just as robust as the signal launched from the previous device.

1.3 There are two main advantages which follow from the characteristics of PCM audio. Firstly, in the absence of a compression system, the quality of reproduction of a well engineered digital audio system is independent of the transmission or recording medium and depends only on the quality of the conversion processes. Secondly, the freedom to interchange audio data without generation loss between a variety of media allows tremendous opportunities which were denied to analog signals.

Someone who is only interested in sound quality will judge the former the most relevant. If good quality converters can be obtained, all of the shortcomings of analog audio can be eliminated to great advantage. Wow, flutter, particulate noise, print-through, dropouts, modulation noise, HF squashing, azimuth error, interchannel phase errors and time-code breakthrough all disappear from recorders. Routing is performed without crosstalk from other audio signals or from video signals.
Digital circuitry costs less to manufacture. Switching circuitry which handles binary can be integrated more densely than analog circuitry. More functionality can be put in the same chip. Analog circuits are built from a host of different component types which have a variety of shapes and sizes and are costly to assemble and adjust.

Digital circuitry uses standardized component outlines and is easier to assemble on automated equipment. Little if any adjustment is needed and periodic maintenance is virtually eliminated.

Once audio is in the digital domain, it becomes data, and as such is indistinguishable from any other type of data. Systems and techniques developed in other industries for other purposes can be used for audio. Computer equipment is available at low cost because the volume of production is far greater than that of professional audio equipment. Disk drives and memories developed for computers can be put to use in audio products. A word processor adapted to handle audio samples becomes a workstation. There seems to be little point in waiting for a tape to wind when a disk head can access data in milliseconds. The difficulty of locating the edit point and the irrevocable nature of tape-cut editing are hardly worth considering when the edit point can be located by viewing the audio waveform on a screen or by listening at any speed to audio from a memory. The edit can be simulated and trimmed before it is made permanent.

Fig.1.5 shows a simple digital system consisting of an analog to digital converter (ADC), a digital router and a digital to analog converter. The router and the DAC reclock the data so that they are identical to those leaving the ADC. Consequently the presence of the router has no effect whatsoever on the audio quality. This is true provided that there is no data corruption due to noise or jitter which is in excess of what the reclockers can handle. Thus in digital systems there are two important areas where attention must be paid:

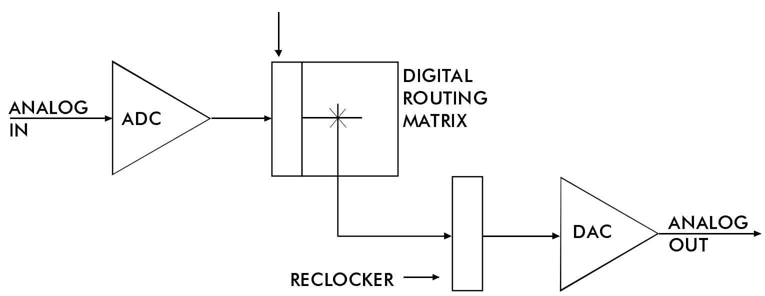


Figure 1.5 When reclocking is used, the routing system is transparent. Quality is determined by the ADC and DAC, therefore well designed products are essential.

1/ ADCs and DACs determine sound quality and the best available should be used. Cheap converters are false economy as they drag down the performance of the rest of the system.

2/ The quality is determined by the converters only if there are no data errors. Therefore a major goal in digital systems is data integrity. Data integrity cannot be measured with traditional analog test gear but has to be built in by good installation practice.

Chapter 2

Conversion

The quality of digital audio is determined by the accuracy of conversion between the analog and digital domains. This section considers the theory and practice of this critical aspect of digital audio.

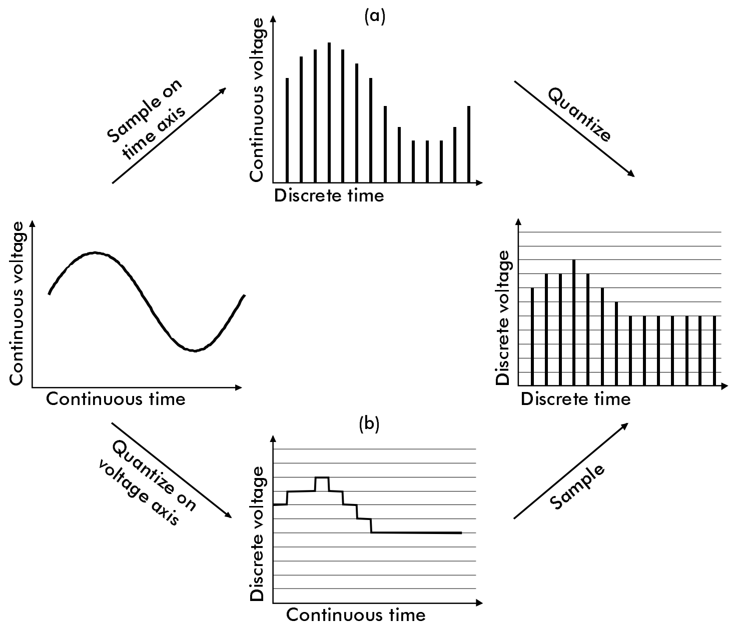


Figure 2.1 Since sampling and quantizing are orthogonal, the order in which they are performed is not important. In (a) sampling is performed first and the samples are quantized. This is common with audio converters. In (b) the analog input is quantized into an asynchronous binary code. Sampling takes place when this code is latched on the sampling clock edges. This approach is universal in video converters.

2.1 The ear can detect tiny amounts of distortion, and will accept an enormous dynamic range. These characteristics require audio signals to have a precision far in excess of that required by video. The only concession is that much less bandwidth is required. The input to an ADC is a continuous-time, continuous-voltage waveform, and this is changed into a discrete-time, discrete-number format by a combination of sampling and quantizing. These two processes are totally independent and can be performed in either order and discussed quite separately in some detail. Fig.2.1a) shows an analog sampler preceding a quantizer, whereas b) shows an asynchronous quantizer preceding a digital sampler. Both approaches will be found in real equipment.

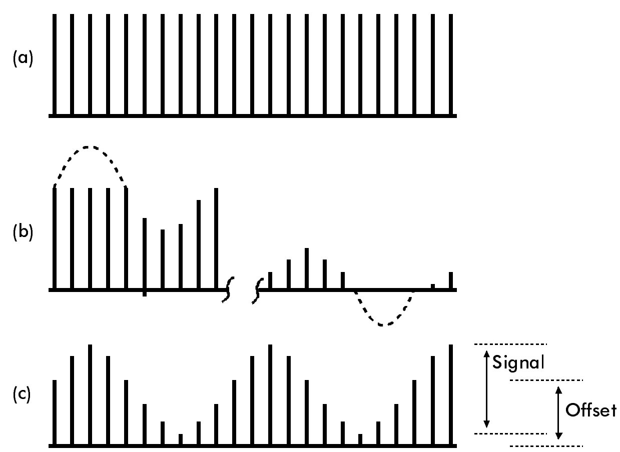


Figure 2.2 The sampling process requires a constant-amplitude pulse train as shown in (a). This is amplitude modulated by the waveform to be sampled. If the input waveform has excessive amplitude or incorrect level, the pulse train clips as shown in (b). For an audio waveform, the greatest signal levelis possible when an offset of half the pulse amplitude is used to

2.2 The sampling process originates with a pulse train which is shown in Fig.2.2 to be of constant amplitude and period. The audio waveform amplitude-modulates the pulse train in much the same way as the carrier is modulated in an AM radio transmitter. In the same way that AM radio produces sidebands or images above and below the carrier, sampling also produces sidebands although the carrier is now a pulse train and has an infinite series of harmonics as shown in Fig.2.3a). The sidebands repeat above and below each harmonic of the sampling rate as shown in b).

The sampled signal can be returned to the continuous-time domain simply by passing it into a low-pass filter known as an anti-image or reconstruction filter.

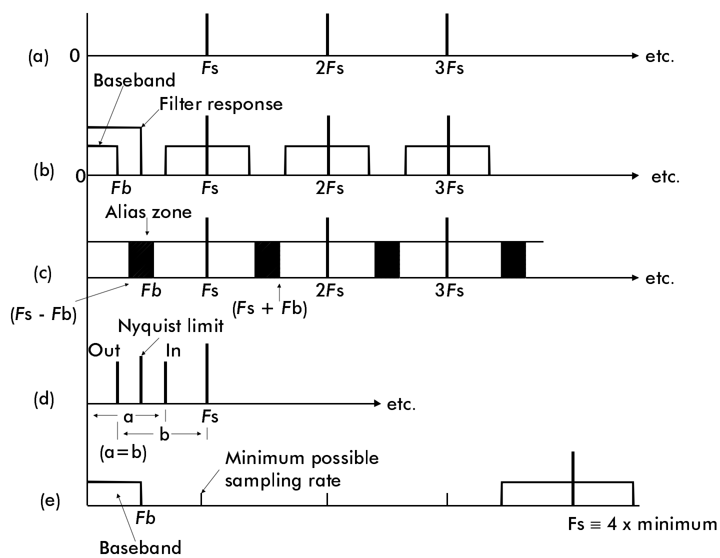


Figure 2.3 (a) Spectrum of sampling pulses. (b) Spectrum of samples. (c) Aliasing due to sideband overlap. (d) Beat-frequency production. (e) 4 x oversampling.

This filter has a frequency response which prevents the images from passing, and only the baseband signal emerges, completely unchanged. It could be argued that the reconstruction filter is unnecessary, since all the images are outside the range of human hearing. However, the slightest non-linearity in subsequent analog stages would result in gross intermodulation distortion and if such a signal reached a TV transmitter the images would take the transmitted spectrum way out of tolerance.

If an input is supplied having an excessive bandwidth for the sampling rate in use, the sidebands will overlap, (Fig.2.3c)) and the result is called aliasing, where certain output frequencies are not the same as their input frequencies but instead become difference frequencies (Fig.2.3d)). It will be seen from Fig.2.3 that aliasing does not occur when the input frequency is equal to or less than half the sampling rate, hence the rule which says that the sampling rate must be at least twice the input bandwidth.

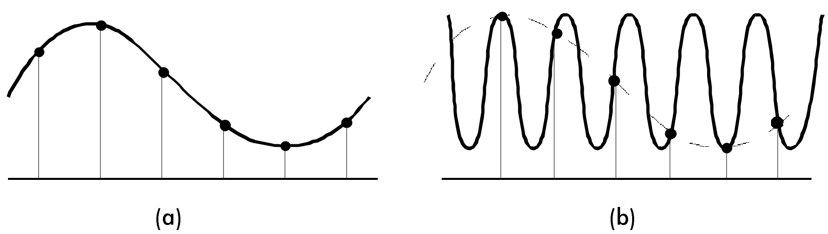


Figure 2.4 In (a) the sampling is adequate to reconstruct the original signal. In (b) the sampling rate is inadequate, and reconstruction produces the wrong waveform (dashed). Aliasing has taken place.

Aliasing can be considered equally well in the time domain. In Fig.2.4a) the sampling rate is obviously adequate to describe the waveform, but at b) it is inadequate and aliasing has occurred. Aliasing sounds quite unnatural and has to be prevented in practical systems. A low-pass, or anti-aliasing filter is provided at the analog input to prevent frequencies of more than half the sampling rate from reaching the sampling stage.

The ideal filter with a vertical "brick-wall" cut-off slope is difficult to implement. As the slope tends to vertical, the delay caused by the filter goes to infinity: great quality but you don’t live to hear it. In practice, a filter with a finite slope has to be accepted. The cut-off slope begins at the edge of the required band, and consequently the sampling rate has to be raised a little to drive aliasing products to an acceptably low level.

Every signal which has been through the digital domain has passed through both an anti-aliasing filter and a reconstruction filter. These filters must be carefully designed in order to prevent audible artifacts, particularly those due to lack of phase linearity as they may be audible. The nature of the filters used has a great bearing on the subjective quality of the system. It is difficult to produce an analog filter with low distortion, but fortunately today the problem is sidestepped using oversampling. The audible superiority and economy of oversampling converters has led them to be almost universal. Strictly speaking, oversampling means no more than using a higher sampling rate than theory demands. In the loose sense an "oversampling converter" generally implies that some combination of high sampling rate and various other techniques has been applied. Oversampling is treated in depth in a later paragraph of this section.

2.2 Sampling theory is only the beginning of the process which must be followed to arrive at a suitable sampling rate. The finite slope of realizable filters will compel designers to raise the sampling rate. For consumer products, the lower the sampling rate the better, since the cost of the medium is directly proportional to the sampling rate: thus sampling rates near to twice 20kHz are found. However, in professional digital audio recorders used in record production, there is a need to operate at variable speed for pitch correction. When the speed of a digital recorder is reduced, the offtape sampling rate falls. With a minimal sampling rate the first image frequency can become low enough to pass the reconstruction filter. If the sampling frequency is raised without changing the response of the filters, the speed can be reduced without this problem. The sampling rate of 48kHz was adopted for professional purposes and has been taken up by DVTR manufacturers to the exclusion of other rates. Thus in the television production environment this is the only rate to consider. Variable sampling rates are seldom used in television as the sampling rate is locked to video timing for practical reasons.

For broadcast purposes a 15kHz audio bandwidth is considered sufficient and a sampling rate of 32kHz is more than adequate. This

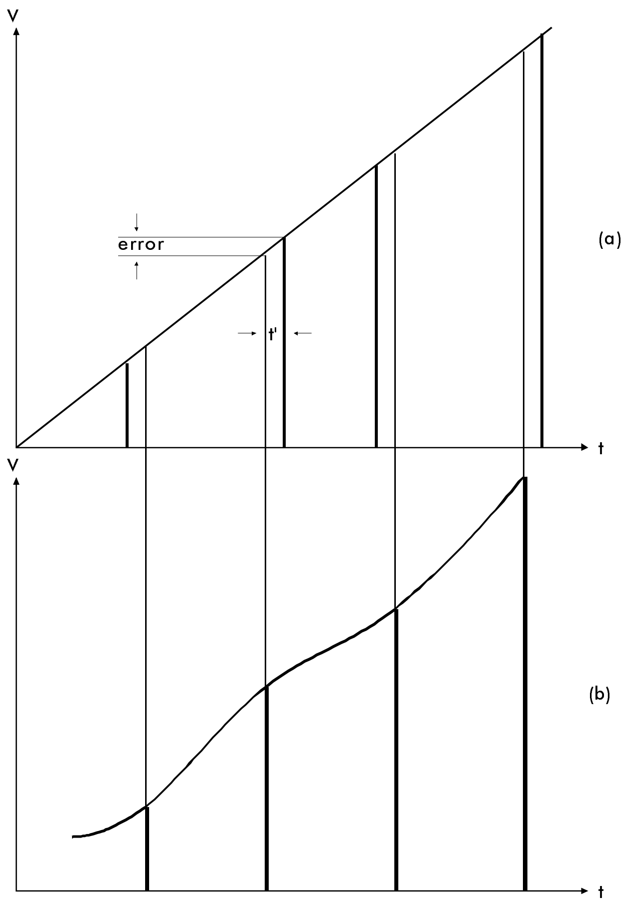


Figure 2.5 The effect of sampling timing jitter on noise, and calculation of the rquired accuracy for a 16 bit system. (a) Ramp sampled with jitter has error proportional to slope. (b) When jitter is removed by later circuits, error appears as noise added to samples. For a 16 bit system there are 2^{16}Q, and the maximum slope at 20KHz will be 20 000π x 2^{16}Q per second. If jitter is to be neglected, the noise must be less than 1/2Q, thus timing accuracy t' multiplied by maximum slope =1/2Q or 20 000π x 2^{16}Qt' = 1/2Q

$$\therefore\ t' = \frac{1}{2 \times 20\,000 \times \pi \times 2^{16}} = 121\ \text{ps}$$

Frequency is used in the NICAM 728 stereo TV sound system and in DAB. The Compact Disc, like most consumer equipment has a sampling rate of 44.1kHz. In the same way that 60Hz U.S. video is incompatible with 50 Hz European video, and needs a standards converter, audio with different sampling rates is incompatible and a sampling rate converter is needed. Section 5 deals with rate conversion.

2.3 The sampling clock has to be generated very precisely. Fig.2.5 shows the effect of ADC sampling clock jitter on a sloping waveform. Samples are taken at the wrong times. When these samples have passed through a system, the timebase correction stage prior to the DAC will remove the jitter, and the result is shown at b). A similar effect will occur if correctly sampled data is converted by a DAC with a jittery clock. The magnitude of the unwanted signal is proportional to the slope of the audio waveform and so the amount of jitter which can be tolerated falls at 6dB per octave. As the resolution of the system is increased by the use of longer sample wordlength, tolerance to jitter is further

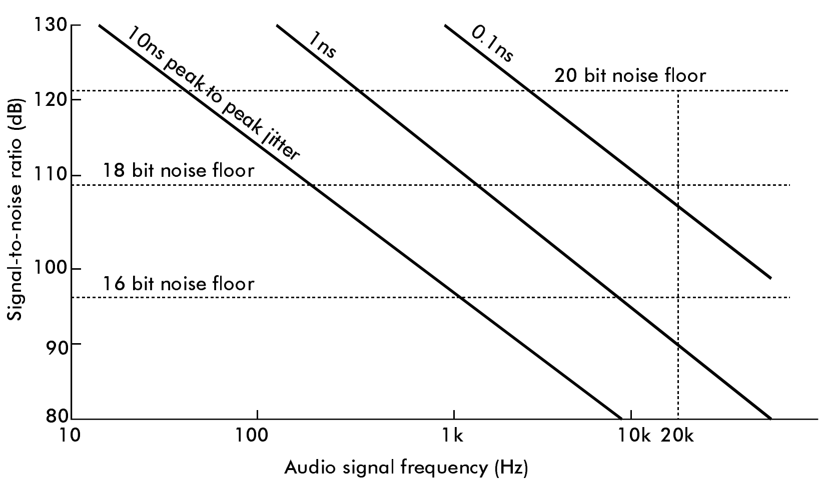


Figure 2.6 Effects of sample clock jitter on signal-to-noise ratio at different frequencies, compared with theoretical noise floors of systems with different resolutions.

reduced. The nature of the unwanted signal depends on the spectrum of the jitter. If the jitter is random, the effect is noise-like and relatively benign unless the amplitude is excessive.

Fig.2.6 shows the effect of differing amounts of random jitter with respect to the noise floor of various wordlengths. Note that even small amounts of jitter can degrade a 20-bit converter to the performance of a good 16-bit unit. There is thus no point in upgrading to higher resolution converters if the clock stability of the system is insufficient to allow their performance to be realized.

Clock jitter is not necessarily random. One source of clock jitter is crosstalk or interference on the clock signal. This is one reason the AES/EBU digital audio interface uses a balanced line transmission.

If an external clock source is used, it cannot be used directly, but must be fed through a well designed, well-damped phase-locked-loop which will filter out the jitter. The phase locked loop must be built to a higher accuracy standard than in most applications. Noise reaching the frequency control element will cause the very jitter the device is meant to eliminate. Some designs use a crystal oscillator whose natural frequency can be shifted slightly by a varicap diode. The high Q of the crystal produces a cleaner clock.

2.4 Quantizing is the process of expressing some infinitely variable quantity by discrete or stepped values. In audio the values to be quantized are infinitely variable voltages from an analog source. Strict quantizing is a process which operates in the voltage domain only; time can be assumed to stand still.

Fig.2.7a) shows that the process of quantizing divides the voltage range up into quantizing intervals Q, also referred to as steps S. In applications such as telephony these may usefully be of differing size, but for digital audio the quantizing intervals are made as identical as possible. When this is done, the binary numbers which result are truly proportional to the original analog voltage, and the digital equivalents of mixing and gain changing can be performed by adding and multiplying sample values. When all quantizing intervals are the same, the term uniform quantizing is used.

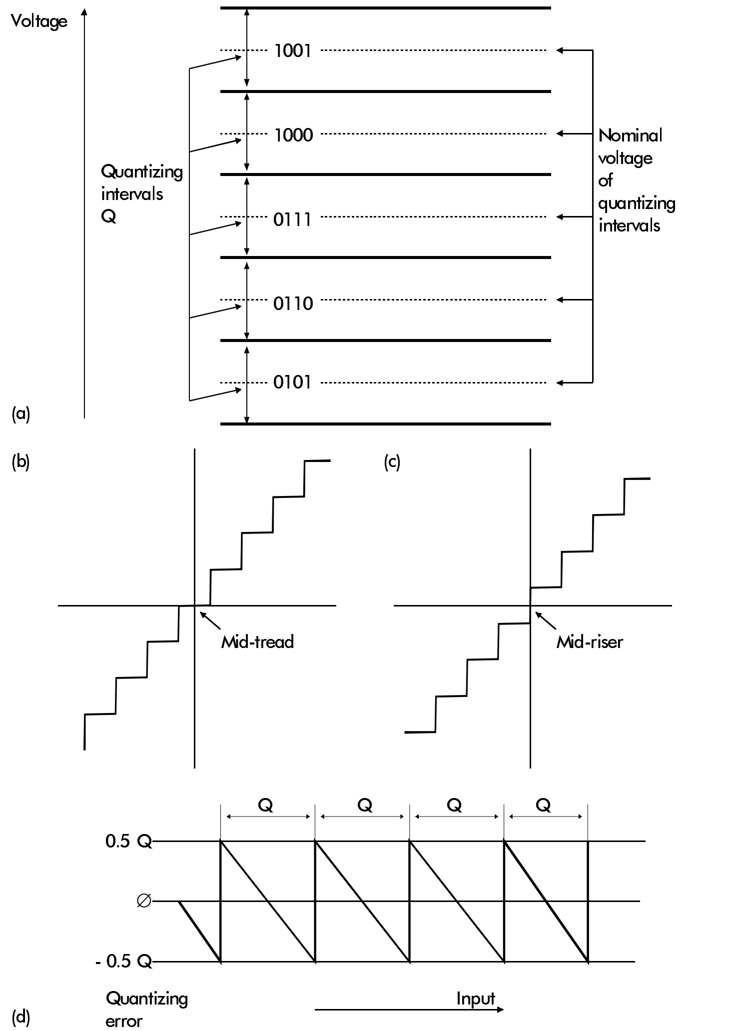


Figure 2.7 Quantizing assigns discrete numbers to variable voltages. All voltages within the same quantizing interval are assigned the same number which causes a DAC to produce the voltage at the centre of the intervals shown by the dashed lines in (a). This is the characteristic of the mid-tread quantizer shown in (b). An alternative is the mid-riser system shown in (c). Here 0 volts analog falls between two codes and there is no code for zero. Such quantizing cannot be used prior to signal processing because the

Whatever the exact voltage of the input signal, the quantizer only locates the quantizing interval in which it lies. The quantizing interval is then allocated a code value which is typically some form of binary number. The information sent is the number of the quantizing interval in which the input voltage lay. The precise location within the interval is not conveyed, putting a limit on the accuracy of the quantizer. When the number of the quantizing interval, i.e. the sample value, is converted back to analog, it will result in a voltage at the center of the quantizing interval as this minimizes the magnitude of the error between input and output. The number range is limited by the wordlength of the binary numbers used. For example, in a sixteen-bit system, 65,536 different quantizing intervals exist, although the ones at the extreme ends of the range have no outer boundary.

2.5 Fig.2.7b) shows that the transfer function of an ideal quantizer is somewhat like a staircase, and zero volts analog, corresponding to all zeros digital or muting, is half way up a quantizing interval, or on the center of a tread. This is the so-called mid-tread quantizer which is universally used in audio. Quantizing causes a voltage error in the audio sample which is given by the difference between the actual staircase transfer function and the ideal straight line. This is shown in Fig.2.7d) to be a sawtooth like function whose amplitude cannot exceed +/- 1/2 Q peak-to-peak unless the input is so large that clipping occurs.

The non-linearity of the transfer function results in distortion, which produces harmonics. Unfortunately these harmonics are generated after the anti-aliasing filter, and so any which exceed half the sampling rate will alias. This results in anharmonic distortion within the audio band. These anharmonics result in spurious tones known as bird-singing. When the sampling rate is a multiple of the input frequency the result is harmonic distortion. Where more than one frequency is present in the input, intermodulation distortion occurs, which is known as granulation. As a result practical digital audio devices deliberately use non-ideal quantifiers to achieve linearity.

2.6 At high signal levels, quantizing error on program material is effectively noise. As the audio level falls, the quantizing error of an ideal quantizer becomes more strongly correlated with the signal and the result is distortion. If the quantizing error can be decorrelated from the input in some way, the system can remain linear but noisy. Dither performs the job of decorrelation by making the action of the quantizer unpredictable and gives the system a noise floor like an analog system. The introduction of dither prior to a conventional quantizer inevitably causes a slight reduction in the signal to noise ratio attainable, but this reduction is a small price to pay for the elimination of non-linearity's. The addition of dither means that successive samples effectively find the quantizing intervals in different places on the voltage scale. The quantizing error becomes a function of the dither, rather than a

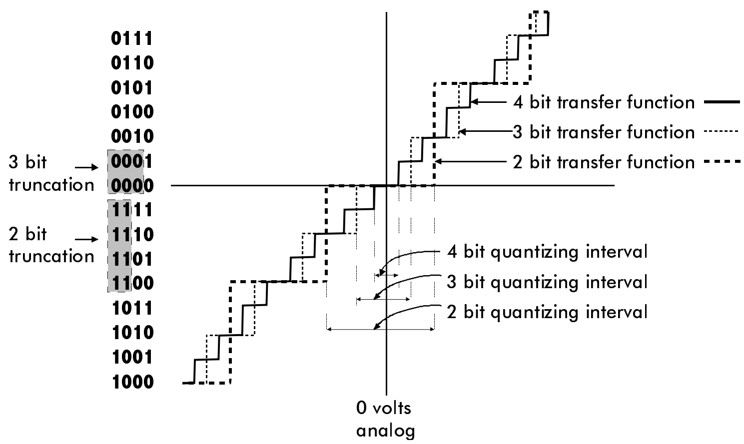


Figure 2.8 Shortening the wordlength of a sample reduces the number of codes which can describe the voltage of the waveform. This makes the quantizing steps bigger, hence the term requantizing. It can be seen that simple truncation or ommission of the bits does not give analogous behavior. Rounding is necessary to give the same result as if the larger steps had been used in the original conversion.

predictable function of the input signal. The quantizing error is not eliminated, but the subjectively unacceptable distortion is converted into a broadband noise which is more benign to the ear.

2.7 Today's ADC technology allows 18- and 20-bit resolution to be obtained. In order to input samples of such wordlength to systems such as NICAM or DVB the words need to be shortened in some way. If the word is simply truncated by discarding the unwanted low-order bits or rounded to the nearest integer the linearising effect of the original dither will be lost. Shortening the wordlength of a sample reduces the number of quantizing intervals available without changing the signal amplitude. As Fig.2.8 shows, the quantizing intervals become larger and the original signal is requantized with the new interval structure. This will introduce requantizing distortion having the same characteristics as quantizing distortion in an ADC.

In practice, the wordlength of samples must be shortened using digital dithering prior to rounding. This is directly equivalent to the analog dithering in an ADC. Digital dither is a pseudo-random sequence of numbers.

2.8 Oversampling is an important topic in digital audio. It implies using a sampling rate which is greater (generally substantially greater) than that required by theory. Fig.2.9 shows the main advantages of oversampling. At a) it will be seen that the use of a sampling rate considerably above the conventional rate allows the anti-aliasing and reconstruction filters to be realized with a much more gentle cut-off slope. There is then less likelihood of phase linearity and ripple problems in the audio passband.

Fig.2.9b) shows that information in an analog signal is two dimensional and can be depicted as an area which is the product of bandwidth and the linearly expressed signal-to-noise ratio. The figure also shows that the same amount of information can be conveyed down a channel with a SNR of half as much (6 dB less) if the bandwidth used is doubled, with 12 dB less SNR if bandwidth is quadrupled, and so on, provided that the modulation scheme used is perfect.

The information in an analog signal can be conveyed using some analog modulation scheme in any combination of bandwidth and SNR which yields the appropriate channel capacity. Raising the sampling rate potentially allows the wordlength of each sample to be reduced without information loss.

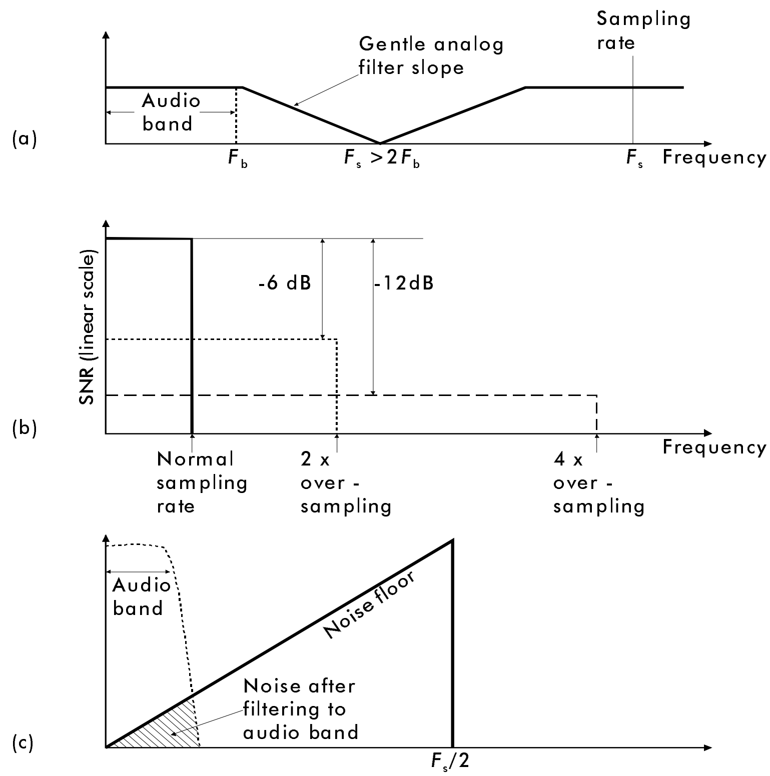


Figure 2.9 Oversampling has a number of advantages. In (a) it allows the slope of analog filters to be relaxed. In (b) it allows the resolution of converters to be extended. In (c) a noise shaped converter allows a

Oversampling enables a converter element of shorter wordlength to be used, suggesting a flash converter as used in digital video. The flash converter is capable of working at very high frequency and so large oversampling factors are easily realized. Certain types of ADC architecture offer a noise floor which rises with frequency. These are called noise-shaped converters. Such a converter can be highly oversampled and the output can be fed to a digital low-pass filter which has the same frequency response as an analog anti-aliasing filter used for conventional sampling. The result is a disproportionate reduction in noise because the majority of the noise is outside the audio band. A high resolution converter can be obtained using this technology without requiring unattainable component tolerances.

Oversampling is confined to converter technology where it gives specific advantages in implementation. The transmission system will employ standard PCM, where the sampling rate is a little more than twice the audio bandwidth. Fig.2.10 shows a digital transmission system using oversampling converters. The ADC runs at n times the Nyquist rate, but once in the digital domain the rate needs to be reduced in a type of digital filter called a decimator. The output of this is conventional Nyquist rate PCM which is transmitted. On reception the sampling rate is raised once more in a further type of digital filter called an interpolator. The system now has the best of both worlds: using oversampling in the converters overcomes the shortcomings of analog anti-aliasing and reconstruction filters and the wordlength of the converter elements is reduced making them easier to construct; the transmission is made with Nyquist rate PCM which minimizes the bandwidth required.

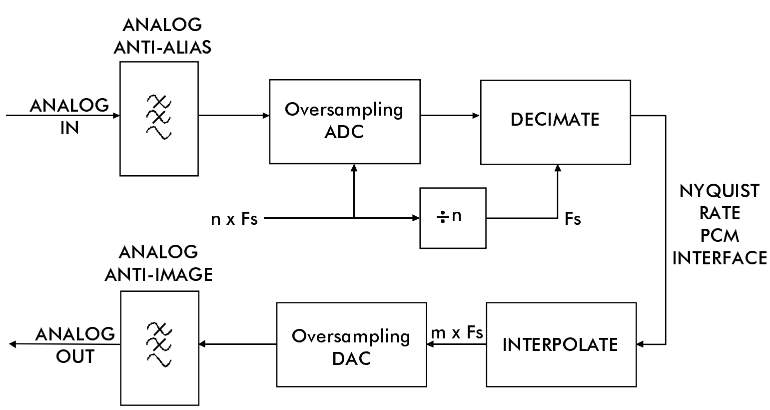


Figure 2.10 Oversampling converters can be used for ADC or DAC applications, but digital filters are used to allow Nyquist rate interfacing.

Chapter 3

Levels and Metering

The subject of level in hybrid digital/analog systems is not straightforward owing to the lack of standardization. In this section we shed light on the principles involved and give practical advice.

3.1 In the analog domain, the 600 Ohm balanced line has been widely used for audio. The use of balancing transformers has been largely supplanted by balanced transformerless circuitry. Fig.3.1a) shows a typical analog interconnect. The source amplifier device has a negligible output impedance which is raised to 600 Ohms by the incorporation of resistors. The load amplifier has a high input impedance which is reduced by the parallel connection of a resistor. Effectively the system acts as a potential divider wasting half of the drive voltage so that the drive amplifier has to produce +10dBu in order to obtain +4dBu on the line. Clearly if an attempt is made to drive two 600 Ohm loads, the line will be shunted by only 300 Ohms and a further 4dB drop will be suffered.

The 600 Ohm source impedance is a hangover from the days when audio was transmitted over long land lines, between towns, for example. In this case, the length of the lines approaches the electrical wavelength of the audio signal and proper source and load termination is needed to prevent reflections. This is why video signals need proper termination because the frequencies are so much higher and the cable lengths easily exceed the signal wavelength. However, in a typical television studio, the distances involved are simply not great enough for this to apply to analog audio. With the exception of long land lines, 600 Ohm source impedance is inappropriate as it has too many drawbacks.

In modern analog installations it is recommended that all source devices should have an output impedance of only 65 - 75 Ohms which is a better match to the cable actually used. Oscillation due to capacitive loads is prevented, and accidental short circuits cause no damage. Receivers should have an input impedance of a few tens of Kilohms. It will be seen from Fig.3.1b) that the line level is virtually independent of the number of loads connected and the source does not have to produce an additional 6dB in level. In the absence of a characteristic impedance, the concept of delivering one milliwatt to get 0dBm is meaningless and all levels are measured in dBu where 0dBu is 0.775 V rms irrespective of impedance.

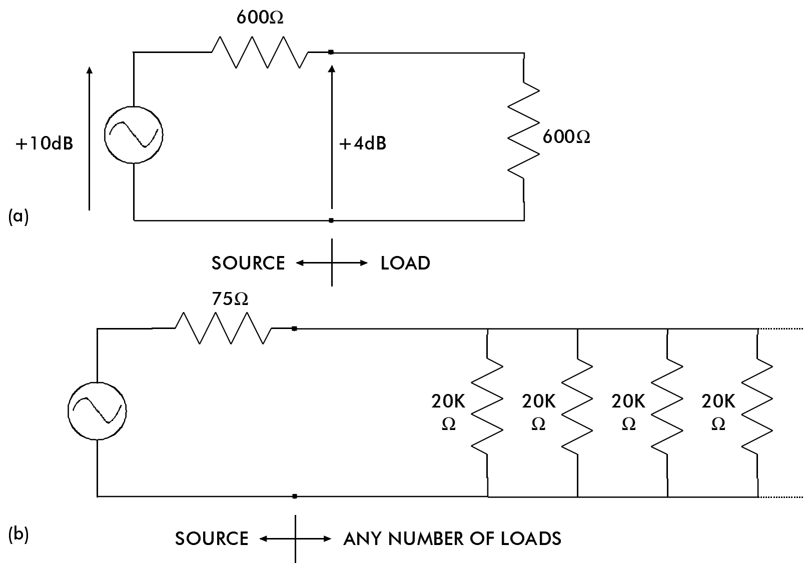


Figure 3.1 (a) Matched impedances waste power because 6dB is lost in the resulting potential divider. (b) Low output impedance and low input impedance allows multiple loads.

3.2 In the digital domain the range of signal levels has two limits. At the bottom of the range is the noise floor due to the ADC. At the top of the range is a hard clip where the positive and negative ends of the number scale are reached. The dynamic range achieved depends on the wordlength and quality of the ADC, but a good rule of thumb is that each bit in the word gives nearly 6dB of dynamic range. Thus a 16-bit system might achieve 93dB; a 20 bit system 115dB. The largest signal which will fit into a digital numbering system without clipping is defined as 0dBFs (dBs Full

Scale) or FSD (Full Scale Digital). Clearly on such a scale all practical levels will be negative.

It is a characteristic of digital systems that over the whole dynamic range the signal will be completely linear as Fig.3.2a) shows. Accordingly the best transmission through a digital system will be one where the signal amplitude is such that it just fails to clip.

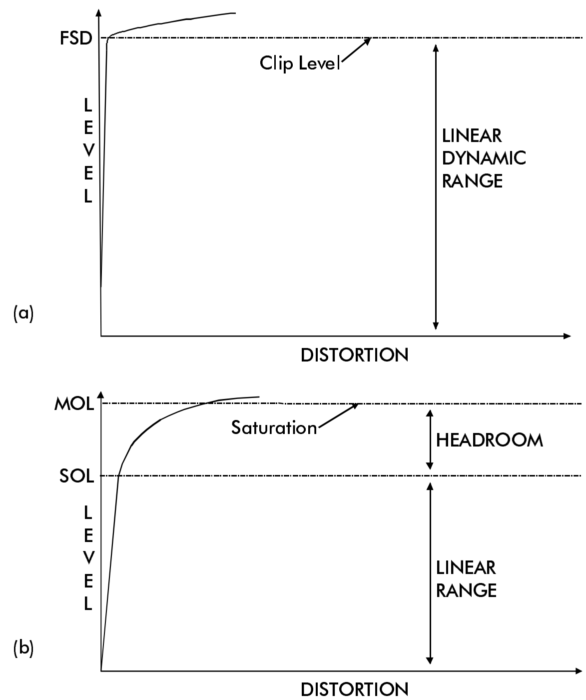


Figure 3.2 The differing level characteristics of analog and digital.

Analog systems are quite different. At the bottom of the dynamic range is the noise floor, but at the top of the range there is no sudden onset of distortion. Instead the signal quality gradually departs from ideal as Fig.3.2b) shows. The best signal quality will be reached at a level where the noise is far enough down, but the onset of distortion is avoided. This level is chosen as a Standard Operating Level (SOL), and the levels above it are known as the headroom. In analog systems, SOL is usually +4dBm. Transients can be recorded in the headroom, as transient distortion is harder for the ear to detect.

3.3 An all-digital system neither has nor needs headroom, but in the real world it will be necessary to interface analog signals having headroom to the digital domain. This requires the digital system to have artificial headroom. This is obtained by adjusting the sensitivity of the ADC such that FSD is not reached at +4dBm but at some higher level. Naturally if the digital channel is to have a 0dB insertion loss the all DACs must have a corresponding gain.

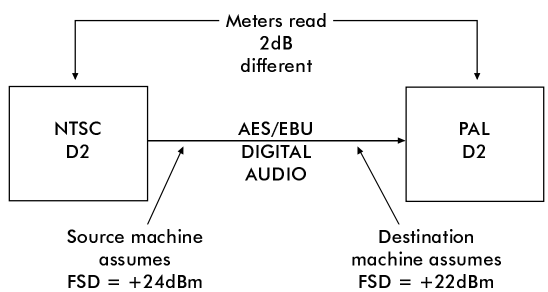


Figure 3.3 A transparent digital transfer results in different meter readings because the two mchines assume different analog levels for FSD.

	Analog	Digital for Audio	Digital for Video	Digital for EBU Video
FSD Nominal Headroom	No Spec 0.0 Vu No Spec	+24 dBu +4 dBu 20dB	+24 or +28 +4 or +8 20dB	dBu +22 dBu dBu +4 dBu 18 dB

Table 3.1 Common operating levels for audio production.

Unfortunately there is no international standard for the amount of digital headroom to leave as Table 3.1 shows. Because of this, it is possible for apparent level changes to occur when digital transfers are attempted. This is difficult to grasp because a digital transfer simply transports a series of numbers from one place to another; a mechanism which cannot change the waveform in any way.

Fig.3.3 shows how it can happen. An NTSC D-2 DVTR having FSD set at 20dB above +4dBm makes a transparent digital audio transfer to a PAL D-2 DVTR which has FSD set at 18dB above +4dBm. Although the numbers transferred are not changed, the level meters on the source machine read 2dB higher than on the destination machine. The 2dB gain error is due to the different scaling of the two machines. Fortunately the digital transfer means that no clipping can occur and the copy will be as good as the original even if it is at the "wrong" level.

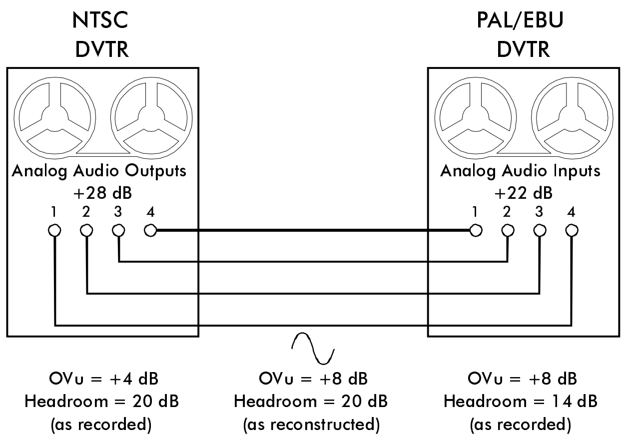


Figure 3.4 FSD level and unexpected gain.

Fig.3.4 shows an analog transfer between two machines with different headroom. The source machine has 24dB of headroom above +4dBm and the destination machine has 20dB of headroom above +4dBm. Meters on the destination machine will read 4dB higher, but of more concern is that signals in the top 4dB of the source machine will clip in the destination machine. If the source DAC is not adjustable, a 4dB pad will need to be inserted between the machines to equalize the levels and to prevent premature clipping.

3.4 In order to avoid level hassle, the following tips will prove invaluable:

1. It is vital to decide upon a single FSD level for the plant and stick to it. NVision ADCs and DACs make this easier because they can easily be set to clip at 20, 24 or 28 dBm. If all ADCs, and DACs in the system are set to the same FSD, level change cannot occur unless program material is coming from elsewhere. If you regularly take or supply significant numbers of digital tapes to another installation, consider using the same FSD level as they do.

2. If possible, avoid transfers through the analog domain between digital audio devices. Using digital transfer avoids generation loss and level discrepancies are eliminated provided all devices have the same FSD level. If different FSD devices cannot be avoided, for example in standards conversion work, a digital level change will be required. This is easily done using a simple mixer such as the NV1055.

3. If analog transfers cannot be avoided, FSD levels of ADCs and DACs must be set correctly. Be prepared to insert pads to obtain a fixed level drop. If level has to be increased, some equipment may require physical modification such as resistor changing in order to get levels to match. Remember if you have a 600 Ohm output impedance device, you can "win" 6dB of gain by feeding it to a Hi-Z input.

Chapter 4

Transmission

The AES/EBU digital audio interface dates from 1985, but was revised in 1992. It was designed to ensure interconnection of professional digital audio equipment irrespective of origin; a goal which has been met.

4.1 As the bit rate of digital audio is fairly low by electronic standards, serial transmission is relatively easy and has the advantage that multi-core cables are not needed. For many purposes existing twisted screened analog cable will pass AES/EBU digital audio, but for new installations cable designed for digital audio should be used. Cable having an impedance of 110 Ohms +/- 10 percent and a capacitance of around 12 - 13 pF per foot will work well.

Transmitting audio data serially is not as simple as connecting the serial output of a shift register to the cable. In digital audio, a common sample value is all zeros, as this corresponds to silence or muting. If a shift register is loaded with all zeros and shifted out serially, the output stays at a constant low level. At the receiver there is nothing to indicate how many zeros were present, or even what the bit rate was. Clearly serialized raw data cannot be transmitted directly, it has to be modulated in to a waveform which contains an embedded clock irrespective of the values of the bits in the samples. At the receiver a circuit called a data separator can lock to the embedded clock and use it to separate strings of identical bits.

4.2 The process of modulating serial data to make it self-clocking is called channel coding. Channel coding also shapes the spectrum of the serialized waveform to make it easier to equalize against cable losses. The AES/EBU interface uses the FM channel code which is DC-free, strongly self-clocking and capable of working with a changing sampling rate. In FM there is always a transition at the bit-cell boundary which acts as a clock. For a data one, there is an additional transition at the bit-cell center. Fig.4.1a) shows that for encoding purposes each data bit can be represented by two channel bits. A channel bit determines whether or not the voltage on the cable changes. For a data zero, the channel bits will be 10, and for a data one they will be 11. Since the first channel bit is always one, each data bit begins with a clock edge. Since there can be two transitions for each data bit, peak-to-peak jitter of half a data bit can be rejected. The lowest frequency in FM is due to a stream of zeros and is equal to half the bit rate. The highest frequency is due to a stream of ones, and is equal to the bit rate. Thus the fundamental frequencies of FM are within a band of one octave. Effective equalization is generally possible over such a band. FM is not polarity conscious and can be inverted without changing the data. As a result an inadvertent exchange of Pins 2 and 3 which would cause an inversion in analog doesn't have the slightest effect on a digital signal.

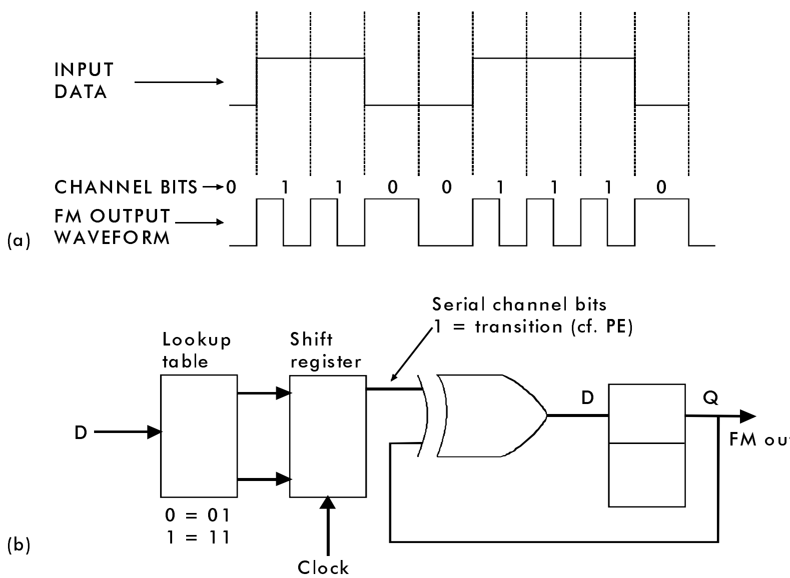


Figure 4.1 In FM coding, each data bit is converted into two channel bits (a), a channel bit 1 causes a transition in the transmitted signal. The FM coder is shown in (b).

Fig.4.1b) shows how an FM coder works. Data words are loaded into the input shift register which is clocked at the data bit rate. This will be 64 times the sampling rate in use. Each data bit is converted to two channel bits in the code book or look up table. These channel bits are loaded into the output register. The output register is clocked twice as fast as the input register because there are twice as many channel bits as data bits. Ones in the serial channel bit output represent transitions whereas zeros represent no change. The channel bits are fed to the waveform generator which is a one bit delay, clocked at the channel bit rate, and an exclusive or gate. This changes state when a channel bit one is input. The result is a coded FM waveform where there is always a transition at the beginning of the data bit period, and a second optional transition whose presence indicates a one.

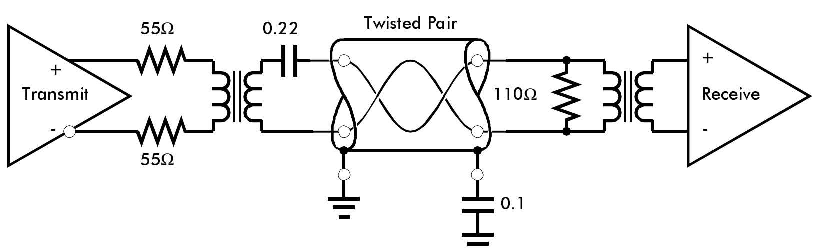


Figure 4.2 Recommended AES3 interconnect circuitry

4.3 The standard driver and receiver chips for RS-422A data communication are employed in AES/EBU, but using transformer coupling for better common mode rejection and avoidance of ground loops. Equalization may be employed on longer cable runs. Fig.4.2 shows the standard configuration. The AES spec does not specify how to handle the shield (Pin 1) at the receiver. Good practice requires Pin 1 to be connected to electrical chassis, not signal ground, so that shield currents cannot produce interference by flowing in common impedance's. Not all equipment does this, and the practice of leaving the shield floating at the receiver was adopted as a palliative. Unfortunately this results in an open house for RF with the potential for interference. One solution is to fit a screen decoupling (or shield bypass) capacitor between shield and chassis which shunts RF energy to ground.

The output impedance of the drivers is 110 ohms, and the impedance of the cable used should be similar at the frequencies of interest. The driver was specified in AES-3-1985 to produce between 3 and 10V pp into 110 Ohms but this was changed to between 2 and 7 volts in AES-3-1992 to better reflect the characteristics of actual RS-422 driver chips.

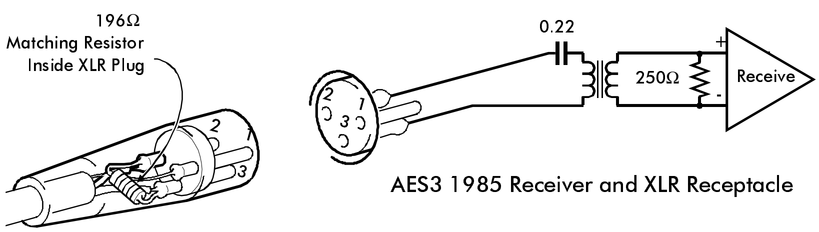


Figure 4.3 Impedance correction of AES3 1985 receivers

The receiver impedance according to AES-3-1985 was set at a high 250 ohms, with the intention that up to four receivers could be driven from one source. This has been found to be inadvisable because of reflections caused by impedance mismatches and AES-3-1992 specifies a point-to-point interface with source, cable and load impedance all set at 110 Ohms. It is important for highest data integrity to convert all old equipment to have the lower input impedance. It is impossible to measure input impedance with an ohm meter because of the transformer, and so it will be necessary to refer to the equipment spec or the schematics to establish the impedance. Fig.4.3 shows how 250 Ohm equipment can be converted without dismantling. A 196 Ohm 1 percent resistor is fitted inside the XLR connector between pins 2 and 3.

This technique cannot be used if the system has the previously legal configuration of several loads connected to one source. In this case a digital audio distribution amplifier such as the NV1021 must be used. However, if the DA is mounted close to the original loads, they need not be modified.

Although not in the AES standard, loop-through can be used in certain circumstances to reduce the number of DAs required. If a large number of devices in the same rack require to be fed with the same reference, they can be looped as shown in Fig.4.4 provided that distances are short and that all devices except the last have a high input impedance rather than the usual 110 Ohms. NVision equipment has a Hi-Z setting to allow this to be done. Looping should not be attempted with equipment which does not offer this facility.

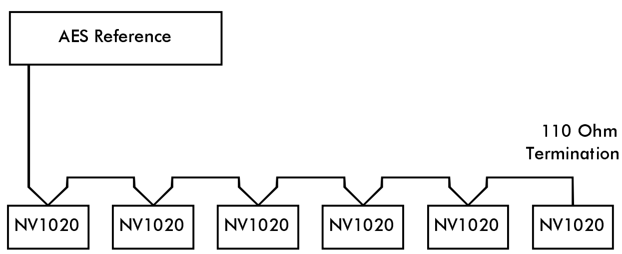


Figure 4.4 Accurately synchronized CODEC's with looping

In some cases it may be uneconomic to pull new cable runs to replace an existing analog run. At the frequencies of digital audio, conventional analog cable has an impedance of around 45 Ohms. Best results will be obtained if the receiver is matched to this value. The "resistor in plug" trick can be used to lower the input impedance. 55 Ohms across a 250 Ohm input matches 45 Ohms, as does 77 Ohms across a 110 Ohm input. The equalizing facilities of the NV1021 will allow reception after considerable signal rounding.

4.4 In Fig.4.5, the specification of the receiver is shown in terms of the minimum eye pattern which can be detected without error. It will be noted that the voltage of 200 mV specifies the height of the eye opening at a width of half a channel bit period. The actual signal amplitude will need to be larger than this, and even larger if the signal contains noise. Fig.4.6 shows the recommended

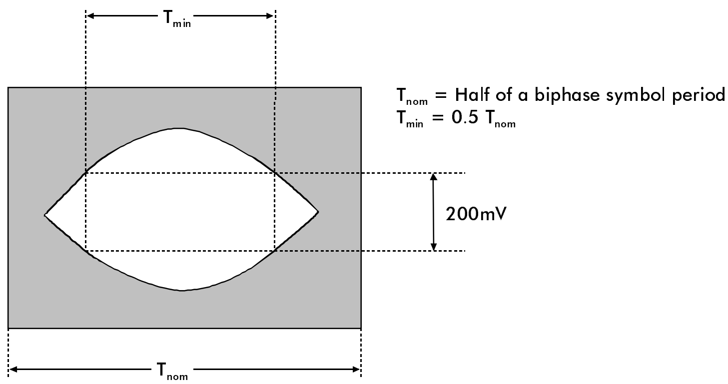


Figure 4.5 The minimum eye pattern acceptable for correct decoding of standard two-channel data.

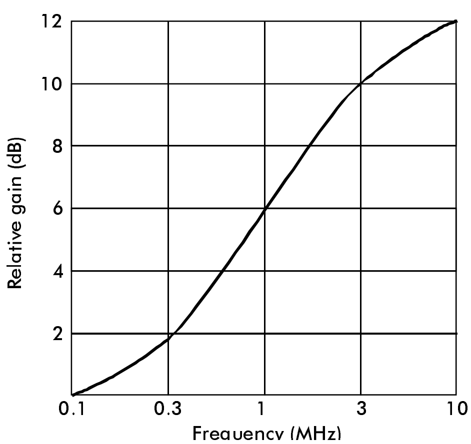


Figure 4.6 EQ characteristic recommended by the AES to improve reception in the case of long lines

equalization characteristic which can be applied to signals received over long lines. The NV1021 contains such an equalizer which will maximize data integrity on long runs.

4.5 There is a further standard, known as AES-ID, using coaxial cable to achieve distances of around 1000m. This is simply the usual AES/EBU protocol but with a 75 Ohm coaxial cable carrying a one volt peak to peak signal (+/- 20 percent); the same level as is used in analog video systems. Fig.4.7 shows what the electrical interface looks like. It is not compulsory to use the transformers, but as most implementations are adaptations of existing circuits the transformers generally remain. Standard video BNC connectors and analog video grade cable are used. Note the shield bypass capacitor at the receiver. This is not in the standard, but allows increased rejection of high frequency interference.

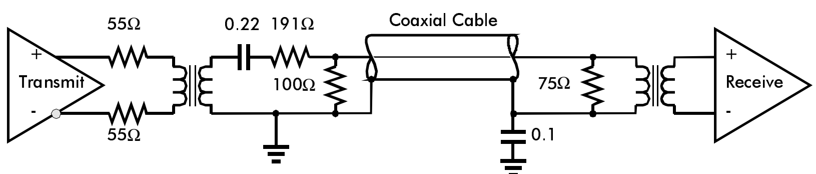


Figure 4.7 Recommended AES3 - ID interconnectivity circuitry.

Many television studios which have gone over to digital video are left with surplus analog video DAs, routers and cabling. It might be thought that this equipment would be perfect for digital audio distribution, but this is not necessarily the case. Such installations should be treated with caution. Whereas a proper digital audio DA or router contains reclocking and slicing which launches a clean signal at the output, video DAs and routers are purely analog devices which neither slice nor reclock. They will introduce generation loss which reduces data integrity. It is possible for two analog video devices to work separately with digital audio, yet when put in series the combination does not work. In such cases it is hard to identify the fault.

4.6 In Fig.4.8 the basic structure of the AES/EBU transmission can be seen. One subframe consists of 32 bit-cells, of which four will be used by a synchronizing pattern. Subframes from the two audio channels, A and B, alternate on a time division basis. Up to 24-bit sample wordlength can be used, which should cater for all conceivable future developments, but normally 20-bit maximum length samples will be available with four auxiliary data bits, which can be used for a voice-grade channel in a professional application.

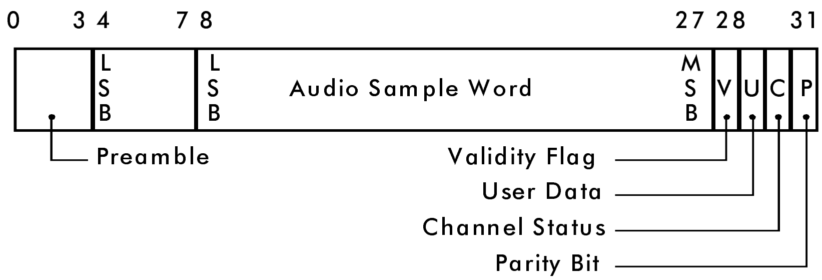


Figure 4.8 AES3 subframe format.

The least significant bit is sent first. One advantage of this approach is that simple serial arithmetic is then possible on the samples because the carries produced by the operation on a given bit can be delayed by one bit period and then included in the operation on the next higher-order bit.

The format specifies that audio data must be in twos complement coding. If different wordlengths are used, the MSBs must always be in bit 27 irrespective of wordlength. Shorter words are leading zero filled up to the 20-bit capacity. The channel status data included from AES-3-1992 signaling of the actual audio wordlength used so that receiving devices could adjust the digital dithering level needed to shorten a received word.

Four status bits accompany each subframe. The validity flag will be reset if the associated sample is reliable. AES-3-1992 described the V-bit as indicating that the information in the associated subframe is "suitable for conversion to an analog signal". Thus it might be reset if the interface was being used for non-audio data. The parity bit produces even parity over the subframe, such that the total number of ones in the subframe is even. This allows for simple detection of an odd number of bits in error, but it also makes successive sync patterns have the same polarity, which can be used to improve the probability of detection of sync. The user and channel-status bits are discussed later.

Two subframes make one frame and frames repeat at the sampling rate in use. The first subframe will contain the sample from channel A, or from the left channel in stereo working. The second subframe will contain the sample from channel B, or the right channel in stereo. At 48 kHz, the bit rate will be 3.072 MHz. In order to separate the audio channels on receipt the synchronizing patterns for the two subframes are different as Fig.4.9 shows. These sync patterns begin with a pulse length of 1.5 bits which violates the FM channel coding rules and so cannot occur due to any data combination. The type of sync pattern is denoted by the position of the second pulse which can be 0.5, 1.0 or 1.5 bits away from the first. The third transition is positioned to make the sync patterns DC free.

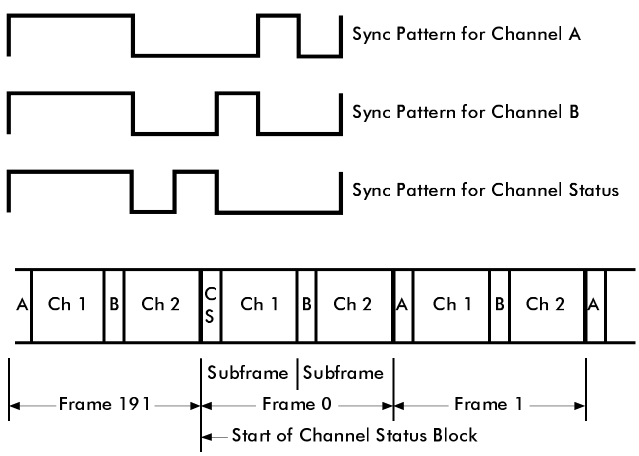


Figure 4.9 AES3 frame format.

The channel status and user bits in each subframe form serial data streams with one bit of each per audio channel per frame. The channel status bits are given a block structure and synchronized every 192 frames, which at 48 kHz corresponds to a period of four milliseconds. The channel A sync pattern is replaced for one frame only by a third sync pattern denoting the start of the channel status block.

4.7 When 24-bit resolution is not required, which is most of the time, the four auxiliary bits can be used to provide voice co-ordination between studios as well as program exchange on the same cables. Twelve bit samples of the talkback signal are taken at one third the main sampling rate. Each twelve bit sample is then split into three four-bit words which can be sent in the auxiliary data slot of three successive samples in the same audio channel. As there are 192 slots per channel status block period, there will be exactly 64 talkback samples in that period. The reassembly of the four-bit words can be synchronized by the channel status sync pattern. Channel status byte 2 has a bit to indicate the use of auxiliary data in this way.

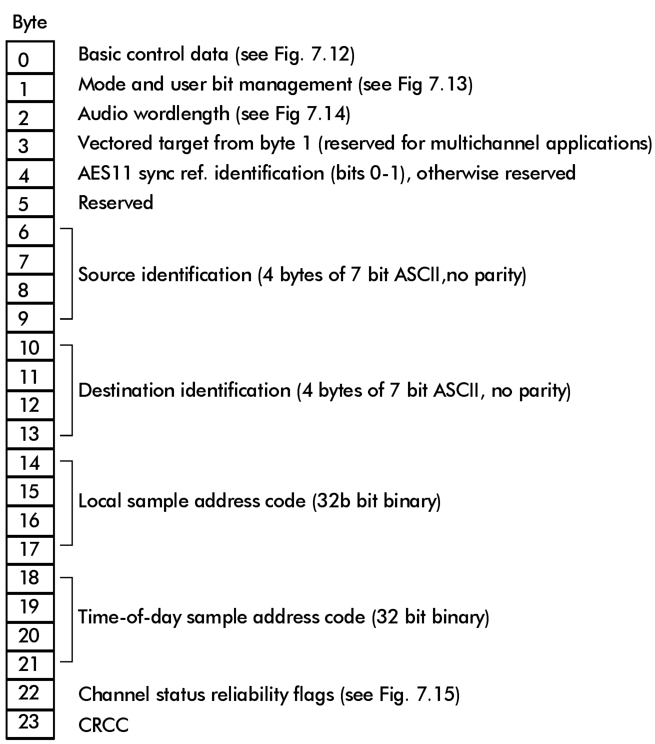


Figure 4.10

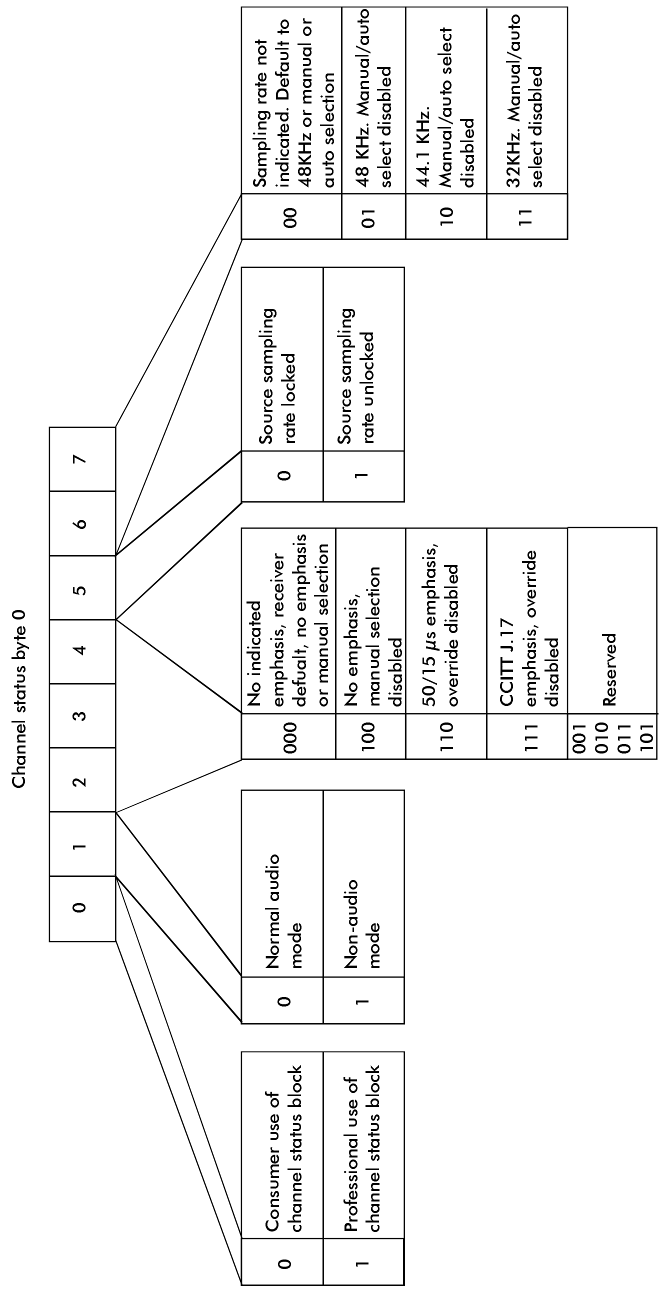


Figure 4.11 The first byte of the channel-status information in the AES/EBU standard deals primarily with emphasis and sampling-rate control.

4.8 The professional channel status structure is shown in Fig.4.10. Byte 0 determines the use of emphasis and the sampling rate, with details in Fig.4.11. Byte 1 determines the channel usage mode, i.e. whether the data transmitted are a stereo pair, two unrelated mono signals or a single mono signal, and details the user bit handling. Fig.4.12 gives details. Byte 2 determines wordlength as in Fig.4.13. This was made more comprehensive in AES-3-1992. Byte 3 is applicable only to multichannel applications. Byte 4 indicates the suitability of the signal as a sampling rate reference.

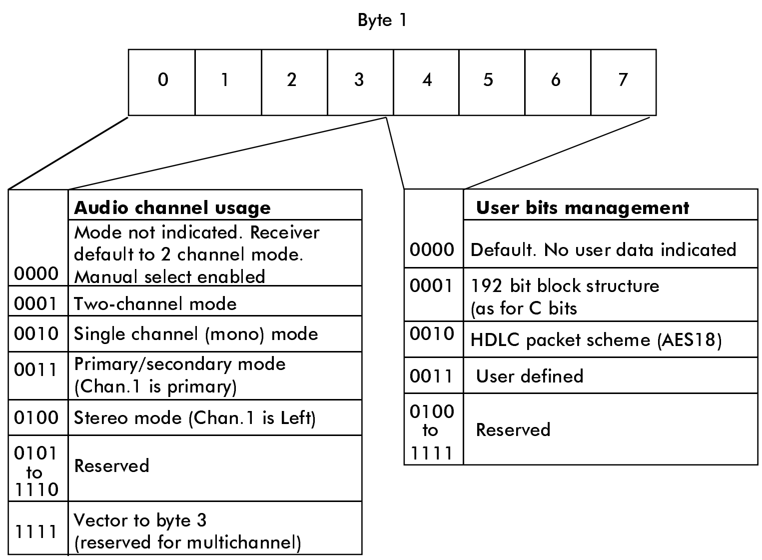
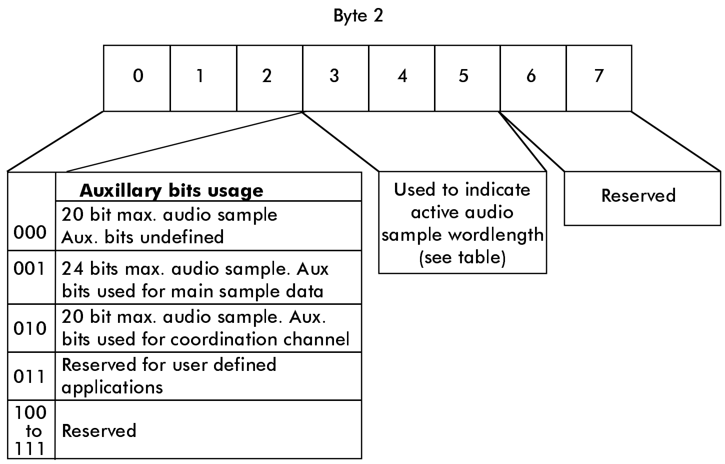


Figure 4.12 Format of byte 1 of professional channel status.

There are two slots of four bytes each which are used for alphanumeric source and destination codes. These can be used for routing. The bytes contain seven bit ASCII characters (printable characters only) sent LSB first with the eighth bit set to zero according to AES-3-1992. The destination code can be used to operate an automatic router, and the source code will allow the origin of the audio and other remarks to be displayed at the destination.



Bits states 345	*Audio wordlength (24 bit mode)*	*Audio wordlength (20 bit mode)*
000	Not indicated	Not indicated
001	23 bits	19 bits
010	22 bits	18 bits
011	21 bits	17 bits
100	20 bits	16 bits
101	24 bits	20 bits

Figure 4.13 Format of byte 2 of professional channel status.

Bytes 14-17 convey a 32-bit sample address which increments every channel status frame. It effectively numbers the samples in a relative manner from an arbitrary starting point. Bytes 18-21 convey a similar number, but this is a time-of-day count, which starts from zero at midnight. As many digital audio devices do not have real time clocks built in, this cannot be relied upon. AES-3-92 specified that the time-of-day bytes should convey the real time at which a recording was made, making it rather like timecode. There are enough combinations in 32 bits to allow a sample count over 24 hours at 48 kHz. The sample count has the advantage that it is universal and independent of local supply frequency. In theory if the sampling rate is known, conventional hours, minutes, seconds, frames timecode can be calculated from the sample count, but in practice it is a lengthy computation and users have proposed alternative formats in which the data from EBU or SMPTE timecode

The next to last byte contains four flags which indicate that certain sections of the channel-status information are unreliable (see Fig.4.14). This allows the transmission of an incomplete channel-status block where the entire structure is not needed or where the information is not available. For example, setting bit 5 to a logical one would mean that no origin or destination data would be interpreted by the receiver, and so it need not be sent.

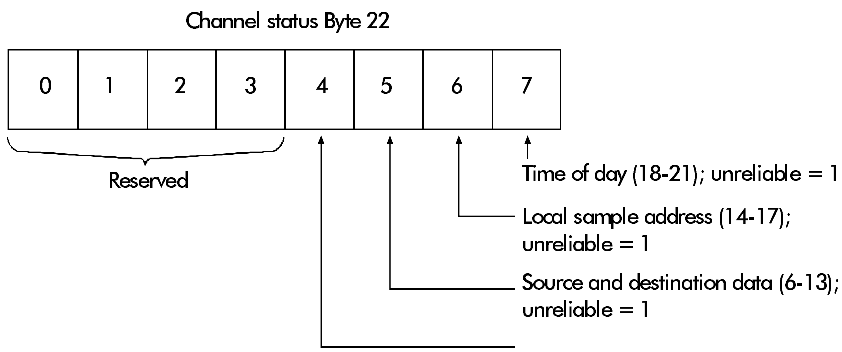


Figure 4.14 Byte 22 of channel status indicates if some of the information in the block is unreliable.

The final byte in the message is a CRCC which converts the entire channel-status block into a codeword. The channel status message takes 4 mSec at 48 kHz and in this time a router could have switched to another signal source. This would damage the transmission, but will also result in a CRCC failure so the corrupt block is not used. Error correction is not necessary, as the channel status data are either stationary, i.e. they stay the same, or change at a predictable rate, e.g. timecode. Stationary data will only change at the receiver if a good CRCC is obtained.

4.9 It is possible to include digital audio in the serial digital video interface signal (SDI) using the auxiliary data capacity which exists during vertical and horizontal sync pulses. This technique is known as embedded audio. The audio data are identified by a special code which video receivers will ignore.

The higher clock rate of component SDI means that there is capacity for up to sixteen audio channels sent in four Groups. Composite SDI has to convey the digitized analog sync edges and bursts and only sync tip is available for ancillary data. As a result of this and the lower clock rate composite SDI only has capacity for four audio channels in one Group. SDI has various levels of support for the wide range of sampling rates and wordlengths encountered in audio. A default level of the standard has been designed specifically for the television production environment and it is unlikely that any more complex system will be required in practice. This is known as Level A which operates only with a video-synchronous 48 kHz sampling rate and transmits V,U,C and the main 20-bit sample only. As Level A transmission is a default receivers will assume it unless told otherwise by signaling and so only Audio Data Packets need be sent.

Fig.4.15 shows the structure of the audio data packing. In order to prevent accidental generation of reserved synchronizing patterns, bit 9 is the inverse of bit 8 so the effective system wordlength is 9 bits.

Address / Bit	x3	x3 + 1	x3 + 2
B9	$\overline{B8}$	$\overline{B8}$	$\overline{B8}$
B8	A (2^5)	A (2^{14})	P
B7	A (2^4)	A (2^{13})	C
B6	A (2^3)	A (2^{12})	U
B5	A (2^2)	A (2^{11})	V
B4	A (2^1)	A (2^{10})	A MSB (2^{19})
B3	A LSB (2^0)	A (2^9)	A (2^{18})
B2	CH (MSB)	A (2^8)	A (2^{17})
B1	CH (LSB)	A (2^7)	A (2^{16})
B0	Z	A (2^6)	A (2^{15})

Figure 4.15 AES/EBU data for one audio sample is sent as three 9 bit symbols. A = audio sample. Bit Z = AES/EBU channel-status block start bit.

Three nine-bit symbols are used to convey all of the AES/EBU subframe data except for the four auxiliary bits. Since four audio channels can be conveyed, there are two "Ch" or channel number bits which specify the audio channel number to which the subframe belongs. A further bit, Z, specifies the beginning of the 192 sample channel status message. V, U and C have the same significance as in the normal AES/EBU standard, but the P bit reflects parity on the three nine bit symbols rather than the AES/EBU definition. The three-word sets representing an audio sample will then be repeated for the remaining three channels in the packet but with different combinations of the CH bits.

One audio sample in each of the four channels of a Group requires twelve video sample periods and so packets will contain multiples of twelve samples. At the end of the packet a checksum is calculated on the entire packet contents.

As the ancillary data transfer is in bursts, it is necessary to provide a little RAM buffering at both ends of the link to allow real time audio samples to be time compressed up to the video bit rate at the input and expanded back again at the receiver. In such a system all that matters is that the average audio data rate is correct. Instantaneously there can be timing errors within the range of the buffers. Audio data cannot be embedded at the video switch point or in the areas reserved for EDH packets, but provided that data are evenly spread throughout the frame, 20-bit audio can be embedded and retrieved with about 48 audio samples of buffering. The buffering stages cause the audio to be delayed with respect to the video by a few milliseconds at each insertion. Whilst this is not serious, Level I allows a delay tracking mode which allows the embedding logic to transmit the encoding delay so a subsequent receiver can compute the overall delay. If the range of the buffering is exceeded for any reason, such as a non-synchronous audio sampling rate fed to a Level A encoder, audio samples are periodically skipped or repeated in order to bring the delay under control.

The extractor recognizes the ancillary data TRS or flag and then decodes the ID to determine the content of the packet. The Group and channel addresses are then used to direct extracted symbols to the appropriate audio channel. A FIFO memory is used to timebase expand the symbols to the correct audio sampling rate. Care is needed with embedded audio when it passes through SDI routers. Special techniques are necessary to prevent clicks on hot cuts. Section 6 deals with this issue.

Chapter 5

Synchronizing

5.1 When multiple digital audio signals are to be assembled from a variety of sources, either for mixing down or for transmission through a TDM (Time Division Multiplexing) system, the audio samples from each source must be synchronized to one another. Such a technique has been used since the earliest days of television in order to allow vision mixing, but now that audio is conveyed in discrete samples, these too must be genlocked or backtimed to a reference for most production purposes. Each source of digital audio must be supplied with a reference sampling rate signal from some central generator, and must return samples at that rate. Failure to properly genlock audio sources results in intermittent pops and crackles as sample values are corrupted. Simply having the same nominal sampling rate is not enough. If digital audio is being used in the television environment further synchronizing will be required. In VTRs the digital audio and video are recorded on the same tape tracks. The scanner or drum speed is locked to video. It follows that the audio block rate and consequently the audio sampling rate must be locked to video. Fig.5.1 shows how to lock external converters to a DVTR. A spare digital audio output is used as a reference.

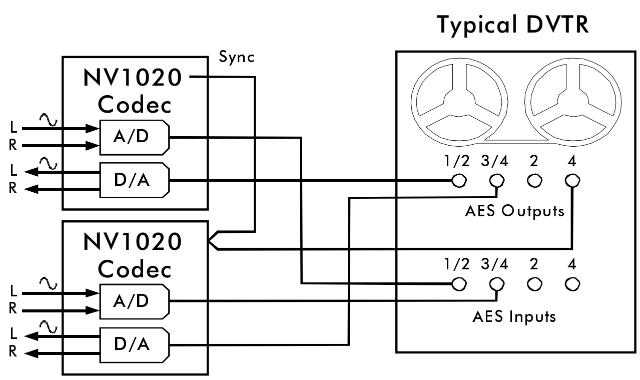


Figure 5.1 Small island synchronization

Whilst all professional digital audio devices have an AES/EBU or SDIF-2 reference input, some recent devices may also have a video reference input which multiplies the video frequency by an appropriate factor to produce a synchronous audio sampling clock. It is, however, unknown for video devices to genlock to audio.

There are two basic ways of locking digital audio in a television environment. In the first, a master generator such as the NV5500 produces locked video and audio references, and these are distributed separately. In the second, only video syncs are distributed, and all audio devices are fitted with video-locking sync generators such as the NV1080. In both cases the NV1022 DA can be used to obtain the required fan-out. In cases where older audio equipment requires SDIF-2 wordclock, the NV1082 can be used as shown in Fig.5.2.

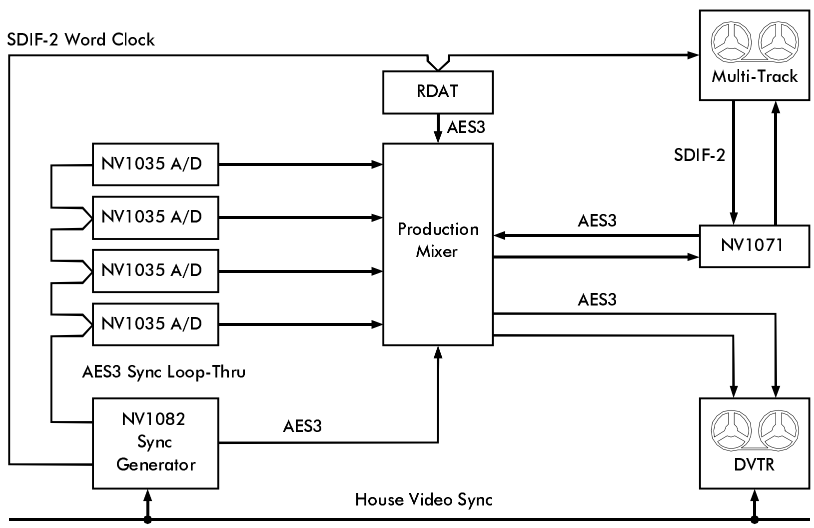


Figure 5.2 The NV1802 reference generator and local island

At first sight it would appear that the two approaches achieve identical results, but this is not quite the case. Fig.5.3 shows how a video-locked audio sampling rate reference is obtained. First, video H-rate is multiplied in a phase locked loop to a frequency of 13.5 MHz which is a common multiple of PAL/NTSC H-rate.

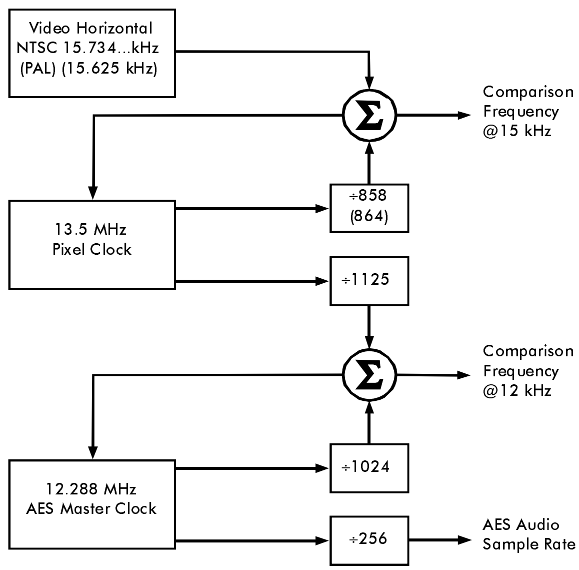


Figure 5.3 Locking digital audio to video horizontal frequency.

The highest common subdivision frequency between 13.5 MHz and 48 kHz is 12 kHz. 13.5 MHz is divided down to 12 kHz and 48 kHz is obtained by multiplication in a further phase locked loop. Whilst the frequencies are in exact ratios, the starting conditions of the four dividers are unknown and so the relative phase of the two signals is arbitrary. This is not a problem in the digital domain, but if such a clock is used as a converter reference, analog phase errors can occur. If two ADCs are driven by two different video-locked references, the relative phase of the two analog signals is lost and if these signals are a stereo pair the imaging will be impaired. The simple solution is that if video locking is used, all converters which are expected to operate with phase coherence should run from the same 48 kHz reference.

5.2 Fig.5.4 shows that the standard 48 kHz sampling rate has fixed relationships to all standard video and film frame rates and to other audio sampling rates. This makes it possible to synchronize digital audio transfers between different picture formats. In most dual-format installations, PAL and NTSC equipment are not locked to

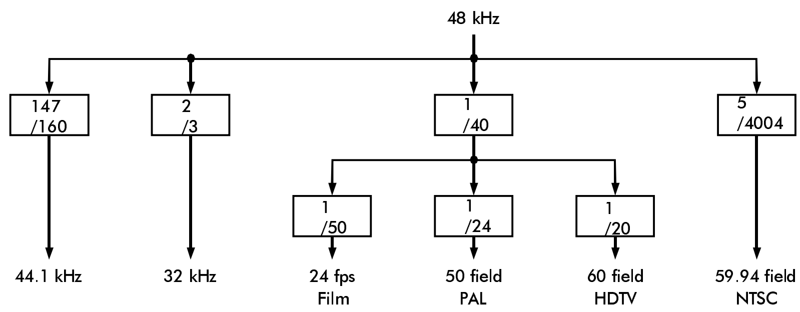


Figure 5.4 Multi-format synchronization relationships.

each other because video standards converters are designed to run with asynchronous signals. With analog audio transfers there is no problem. With digital audio, PAL and NTSC equipment must be locked together. This is easily done as Fig.5.5 shows. The NV5500 Universal Sync generator provides simultaneous PAL and NTSC sync outputs and a synchronous 48 kHz audio output. As NTSC and PAL have no meaningful phase relationship, the NV5500 does not attempt to provide one. Existing video sync generators are genlocked to the NV5500 outputs. In this way all video references within the plant are locked, and any device which multiplies video syncs to produce an audio sampling rate will be automatically genlocked to any other device doing the same thing even if the two devices are on different video standards.

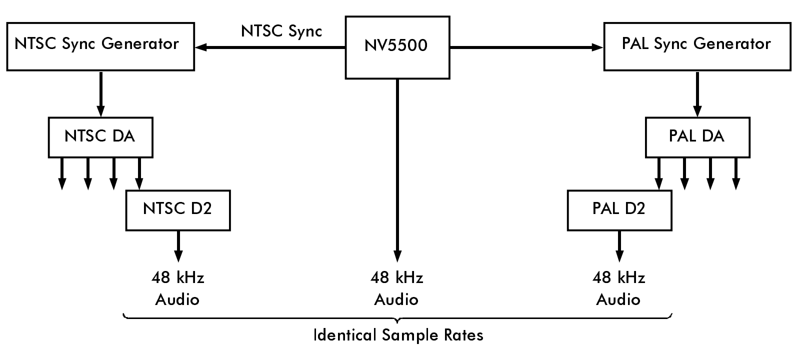


Figure 5.5 NV5500 master timing for system synchronization.

With locked video generators, the digital audio soundtracks can be easily dubbed during a standards conversion operation. Note, however that video standards converters necessarily cause a significant delay in the video path. An NV 1060 digital audio delay unit shown in Fig.5.6 allows the standard converter delay to be matched in the audio, maintaining precise lip-sync.

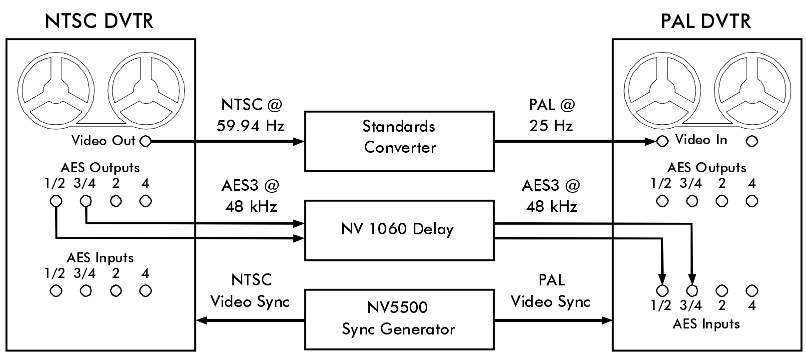


Figure 5.6 Synchronous delay compensation

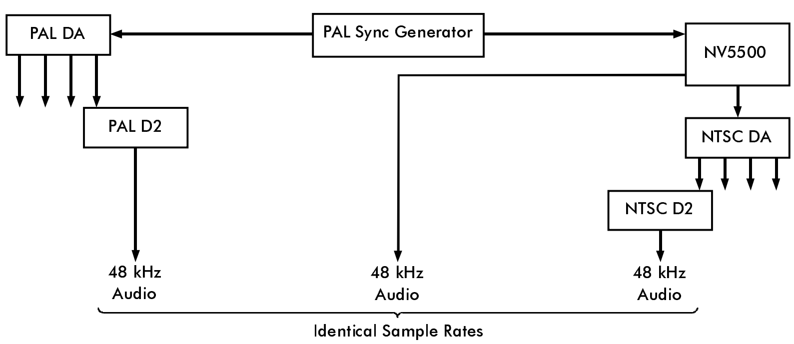


Figure 5.7 NV5500 Slave timing for system synchronization.

Sometimes the video timing is determined elsewhere, for example if two studios are genlocked for transfer over a distribution network. In this case slave synchronization can be used. Fig.5.7 shows that the NV5500 is connected as a slave and generates NTSC syncs from a PAL input or vice versa. When slave locking is used, frequency accuracy needs to be considered, as AES/EBU, NTSC and PAL have different tolerances. Thus in theory signals genlocked to an NTSC signal which was just within tolerance might be outside their stricter tolerances. In practice problems are unlikely to arise.

5.3 AES11-1991 documented standards for digital audio synchronization and requires professional equipment to be able to genlock either to a separate reference input or to the sampling rate of an AES/EBU input. The shift register in a receiver is generally buffered with a parallel loading latch which allows some freedom in the exact time at which the latch is read with respect to the serial input timing. Accordingly the standard defines synchronism as an identical sampling rate, but with a no requirement for a precise phase relationship. Fig.5.8 shows the timing tolerances allowed. The beginning of a frame (the frame edge) is defined as the leading edge of the X preamble. A device which is genlocked must correctly decode an input whose frame edges are within +/- 25 % of the sample period. This is quite a generous margin, and corresponds to the timing shift due to putting about a kilometer of cable in series with a signal. In order to prevent tolerance buildup when passing through several devices in series, the output timing must be held within +/- 5% of the sample period.

The reference signal may be an AES/EBU signal carrying program material, or it may carry muted audio samples; the so called digital audio silence or digital quiet signal. Digital quiet is useful as a router input whereas tone is useful for continuity checking and DAC alignment. The NV5500 and NV1000 family allow quiet or tone to be selected. Tone can be 500 or 1000Hz and the level may be set to FSD (0dBFs) or -20dBFs.

5.4 Although desirable, in practical situations, genlocking or backtiming is not always possible. In a satellite transmission, it is not really practicable to genlock a studio complex halfway round the world to another. When genlock is not achieved, there will be a slow slippage of sample phase between source and destination. In some cases a totally different standard sampling rate, e.g. 44.1 or 32 kHz will need to be handled. Laser discs are mastered on DVTRs

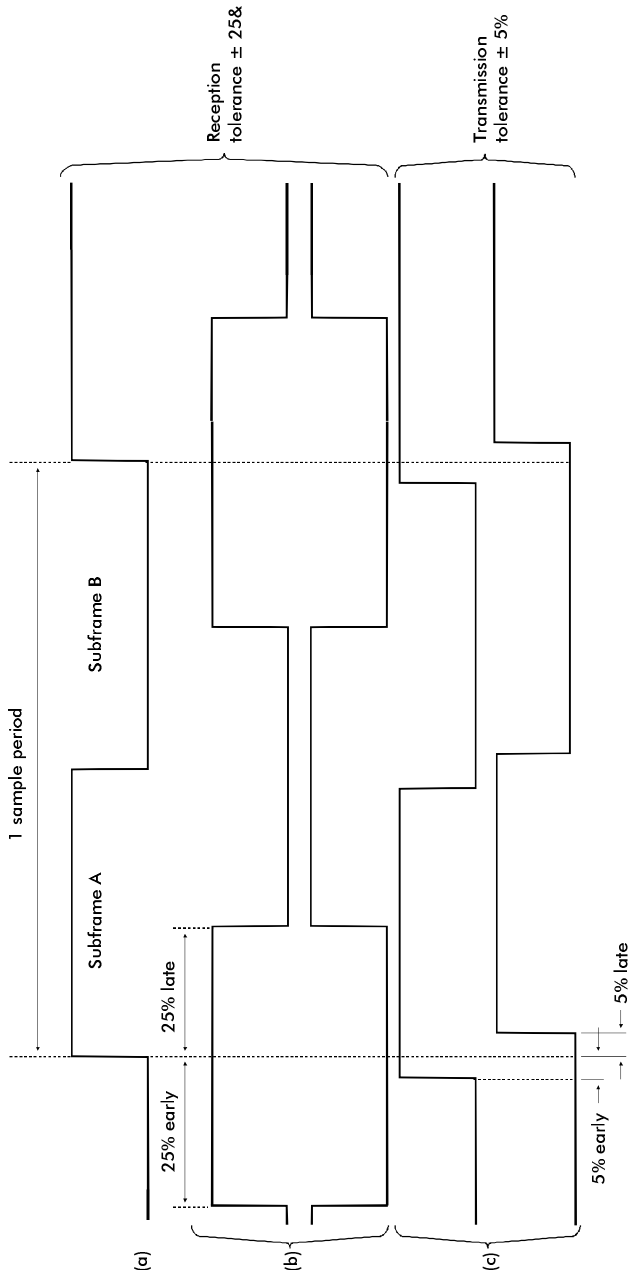


Figure 5.8 The timing accuracy required in AES/EBU signals with respect to a reference (a). Inputs over the range shown at (b) must be accepted, whereas outputs must be closer in timing to the reference shown at (c).

having 48 kHz audio, yet they require 44.1 kHz. The same problem will occur if it is required to make a CD master from a videotape soundtrack. These problems can be handled by sampling rate conversion.

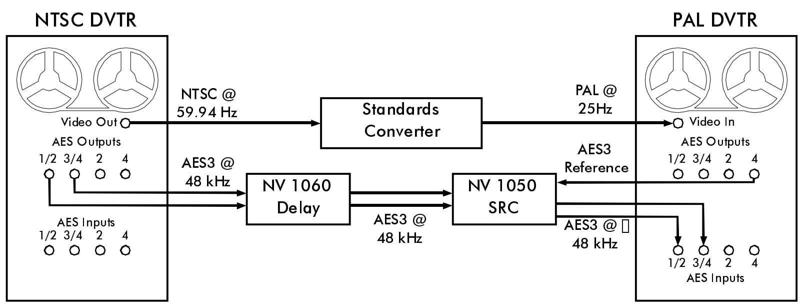


Figure 5.9 A standard NTSC to PAL transfer.

Fig.5.9 shows a standards conversion between two DVTRs. Whilst the ideal solution is to use an NV5500 as shown above, this may not be appropriate where only occasional use is contemplated. All video standards converters are designed to work without lock between the two standards and so the video transfer takes place with both DVTRs free running. This results in the audio sampling rate between the two machines being slightly different. The NV1050 sampling rate converter resolves the discrepancy by expressing the input audio waveforms in a different sampling structure. Note how the output sampling rate of the converter is determined by a reference taken from an audio output of the destination DVTR. The NV1050 can handle all four audio channels simultaneously.

Standards converters need to interpolate along the time axis between several video fields and this results in a significant processing delay. The NV1060 delay unit compensates by delaying the audio data a corresponding amount.

5.5 When film is converted to video in a 59.94 Hz system, the 24 Hz frame rate is handled by running the film 0.1% slow and converting alternate frames to two and three fields. This field repeating has the effect of portraying motion as two unevenly spaced jumps in a five field sequence. Whilst this is acceptable for 59.94 Hz television, it produces dreadful artifacts after standards conversion to 50 Hz. Even motion compensated standards converters fail on such material because they attempt to reproduce the bizarre motion. The solution is to use a special standards converter which is identifies the third field in the 3/2 sequence and deletes it. Pairs of fields are then de-interlaced to produce the original film frames at 23.97 Hz. These frames are interpolated to change to 625 lines and are recorded on a 50 Hz VTR which is modified to run 4.004% slow so that it actually runs at 23.97 Hz. The audio clocks change pro-rata resulting in an audio sampling rate of 46.0339 KHz. Fig.5.10 shows that the NV 1050 sampling rate converter can produce this frequency simply by genlocking to the audio output of the destination DVTR which multiplies up from 23.97 Hz. An NV 1060 audio delay compensates for the video processing delay.

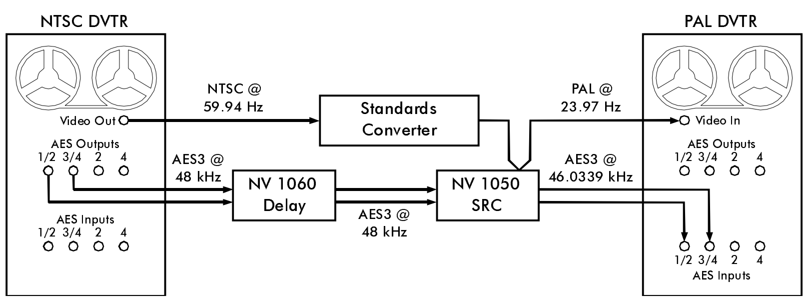


Figure 5.10 Sample rate conversion for complex video transfers.

When the resulting tape is played at 50 Hz and 48 kHz, the video has the same motion portrayal as if the original film had been played on a telecine at 25 Hz and the audio stays in sync with it.

5.6 Consumer DAT recorders are extremely inexpensive and portable, and make excellent field recorders. Although they have no timecode or genlock facilities, they are crystal controlled and stable and can be used unlocked with clapperboard synchronization. In a simple production, asynchronous field DAT tapes may be combined with effects from a 44.1 kHz Compact Disc player using the configuration shown in Fig.5.11. The NV 1055 Mixer needs all four inputs to be video-locked 48 kHz. The flexibility of the NV 1050 is demonstrated particularly well here because it can simultaneously accept two pairs of inputs at different rates and convert them to a common rate.

5.7 Sampling rate conversion is analogous to standards conversion

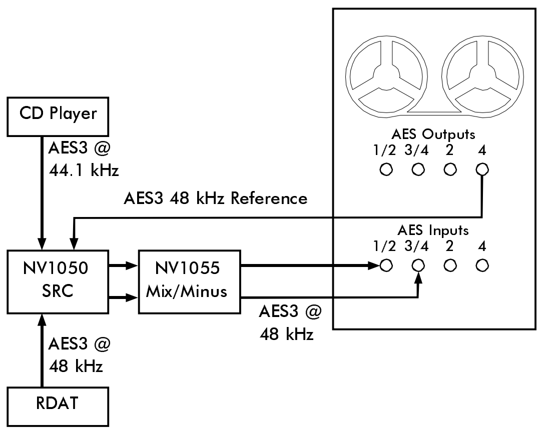


Figure 5.11 Multiple source rate conversion for audio production.

in video. It is an interpolation process using a low-pass filter to compute the value of the input waveform at some a point between the input samples. One of the most difficult aspects of the conversion is to estimate the position on the time axis of the output sample with respect to the input sample in order to generate the filter coefficients. Any inaccuracy here is the equivalent of having clock jitter in the original ADC. This estimation is especially tough if the two rates are totally unlocked as any sample phase relationship can occur. If, however the two rates are related by a ratio of two integers, the number of filter phases becomes finite and the performance can be enhanced because the coefficients are more accurate. This is known as synchronous rate conversion.

The NV4448 has this capability. It contains a phase locked gearbox which has a ratio of 160:147. In addition to accepting AES/EBU clocks, the NV4448 also accepts and generates SDIF-2 and SPDIF signals so it can be used to interface a wide range of equipment to the video environment. In addition to the synchronous conversion facility, the NV 4448 operates in asynchronous mode should it become unlocked.

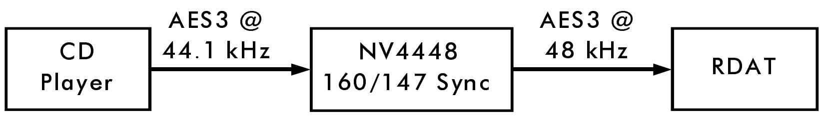


Figure 5.12 Synchronous conversion with AES3 clocking

Fig.5.12 shows a 44.1 kHz CD player driving the NV4448 which locks to the input data rate. The NV4448 then produces a synchronous 48 kHz output and rate converts the input to it. An RDAT machine locks to the 48 kHz output.

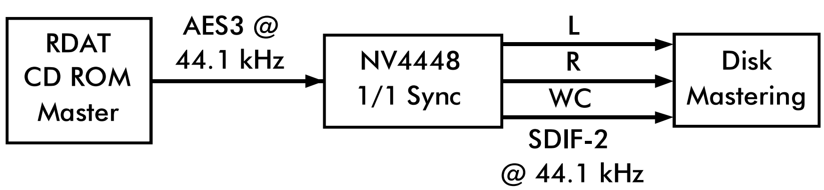


Figure 5.13 Synchronous sample rate conversion for video.

Fig.5.13 shows an alternative situation where a 44.1 kHz RDAT recording is synchronously converted to video-locked 48 kHz to suit a DVTR. The RDAT machine is simply fed with a video reference so that it produces video-locked 44.1 kHz. The DVTR produces video locked 48 kHz. The ratio between these is exactly 160/147.

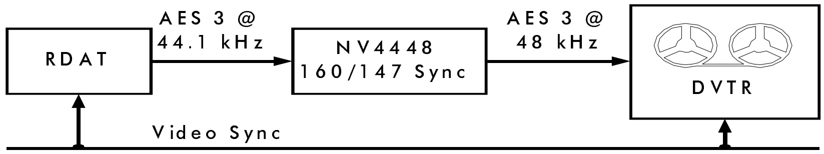


Figure 5.14 CD ROM mastering

Fig.5.14 shows an example of DVTR audio being rate converted to make a laser disc. In this case the flexibility of the NV4448 is useful because many laser disc cutters require SDIF-2 input.

5.8 In the digital domain, delay is easy and in most video installations delay is everywhere. Line delays in decoders, field or frame delays in noise reducers and DVEs, arbitrary delays in frame synchronizers all conspire to give you the sound before the lips moved. Less obvious is that the delays experienced by the video signal also shift it with respect to the timecode.

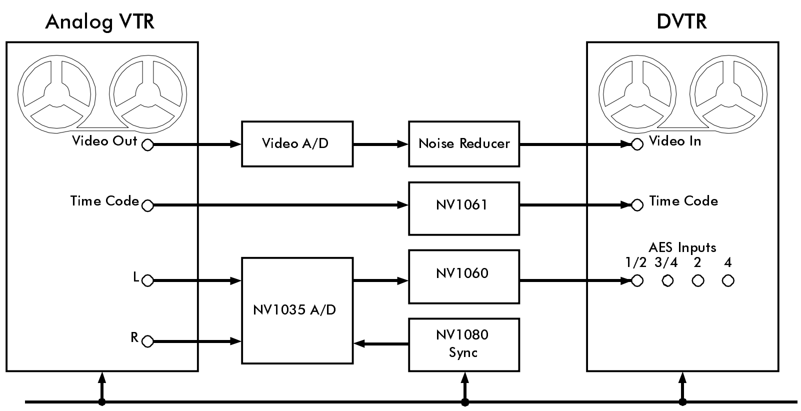


Figure 5.15 Delay compensation for analog to digital conversion with image enhancement.

The solution is to put compensating delays in the audio and timecode signals which cancel the differential delay. Fig.5.15 shows a typical application. An analog tape is being dubbed to a DVTR using video noise reduction. The noise reducer causes significant delay. The timecode passing between the machines is delayed by the NV1061 and the audio is delayed by the NV1060. This requires ADCs in the shape of an NV1035. A video-locked 48 kHz clock comes from a NV1080 fed with house reference.

In Section 5.2 above the use of a compensating audio delay in standards conversion applications was illustrated. In these cases the timecode is not transferred because SMPTE and EBU timecodes are incompatible. Instead the destination machine re-stripes using its own generator from a given starting code and a time code compensator is not required.

Section 6

Routing

6.1 There are two basic types of routing to consider: synchronous and asynchronous. Each has its own advantages and disadvantages. The first question to ask is whether "hot switches" are required. A hot switch is where a router output selects a different source on air. This may be done in a muted pause between, for example, programs and commercial breaks. Clearly during such a quality-critical switch no pops or clicks are allowed. In presentation suites, most transitions between sources are done with a mixer, but in the case of a mixer failure it is possible to stay on air by using the router to bypass the mixer. All transitions then have to be hot switches.

If hot switch capability is required, then only synchronous routing will do the job. However, for many purposes, such as in edit suites, routers will only need to change configuration when no wanted material is passing through. In this case pops and clicks at the switch point are acceptable as they never appear on the finished product. Asynchronous routing can then be used; an option which, in the right application, can offer savings over synchronous.

6.2 Imagine what would happen on a video tape if instead of performing an assemble edit the machine was just forced into record. An asynchronous AES/EBU switch does the same thing to the audio data. The AES/EBU serial data stream is precisely created so that the position of each bit after the sync pattern determines the function of that bit. Consider a switch made between two AES/EBU signals having the random phase relationship shown in Fig.6.1a). In the time after the switch point, but before the next sync pattern, the receiver is fooled into putting the bits from the second source in the wrong place in the sample. The result is a corrupted sample value which on conversion will result in an analog voltage which is likely to have a very different level to those either side - in other words a click. A further problem is that the random jump in sync pattern phase can cause downstream devices to lose lock momentarily causing further clicks or mutes.

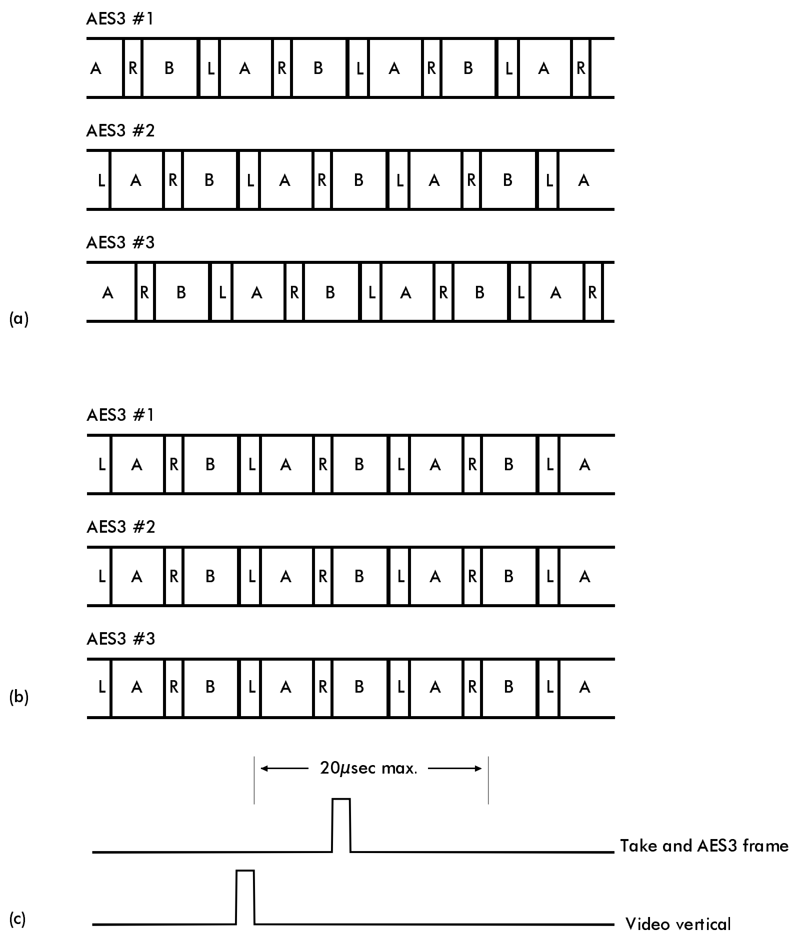


Figure 6.1 (a) Random inputs. (b) Phase alignment. (c) Video transition

6.3 Data corruption can only be avoided by having all inputs to the matrix genlocked as in Fig.6.1b). It is now easy to switch during the sync pattern or frame boundary so that the output changes effectively between samples. In this case no corruption of the sample value takes place. Having to genlock all inputs to the same phase is highly inconvenient and in practice is not necessary because a small buffer memory on each router input can be used to phase align all inputs to a single reference. When this is done, it is only necessary for the sampling RATES to be locked; the exact input phase does not matter. This is the operating principle of a re-phasing synchronous router such as the NV3512SA. With all the matrix inputs phase locked, switching does not disturb the output data phase. In the television environment a synchronous digital audio router will need to switch in the vertical interval to match any associated video material. Whilst the audio sampling rate is video locked, the audio phase is arbitrary. Consequently a video-slaved audio router will switch on the next sync pattern after video vertical, causing a delay of up to 20 microseconds as shown in Fig.6.1c). This is generally insignificant.

6.4 Asynchronous routers don’t have the re-phasing buffers and so are less expensive. Although they can pop and click on transitions, this may not matter for applications such as monitoring. They also have some advantages. There is no need for the audio sampling rate to be locked to video and in fact any sampling rate can be used. This can be useful for handling signal sources such as consumer DAT machines or CD players which have no genlock input. However, it is possible that a downstream device may not be able to lock to a free-running signal. Some devices have a very small lock-in range obtained by tweaking a crystal oscillator and are only really happy with a genlocked input.

6.5 When digital audio is embedded in serial digital video (SDI), the act of switching a video router usually interrupts the AES/EBU framing and results in a click. If hot switching is needed, the audio must be stripped out of the video (disembedded) first. Audio routing must be done in a synchronous audio router prior to re-embedding. Fig.6.2 shows how this is done. Assuming a cut is being made between two SDI signals, each signal must have a disembedder so that clean digital audio data are available from both SDI signals. The video router performs the video cut, and the audio cut is performed in a synchronous audio router prior to re-embedding. Installing a de-embedder on every input to an SDI

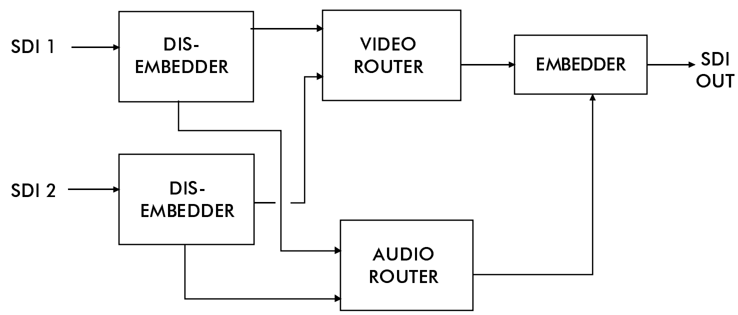


Figure 6.2 Disembedding prevents pops and crackles at SDI switches.

video router would be very expensive, and the alternative is have a smaller number (at least two) of de-embedders on dedicated SDI router outputs. In this way a de-embedder can be assigned in advance to a video signal to which the SDI router is about to cut. The de-embedded audio is then available to the parallel synchronous audio router which cuts to it on the next audio frame boundary after the video cut. After the cut, the outgoing de-embedder is no longer in use and can be re-assigned to the next source. Care is needed to ensure that the SDI router does not switch any de-embedder input when the output is in use. It should be recalled that the embedding and de-embedding processes cause a slight delay to the audio which may build up if these techniques are cascaded.

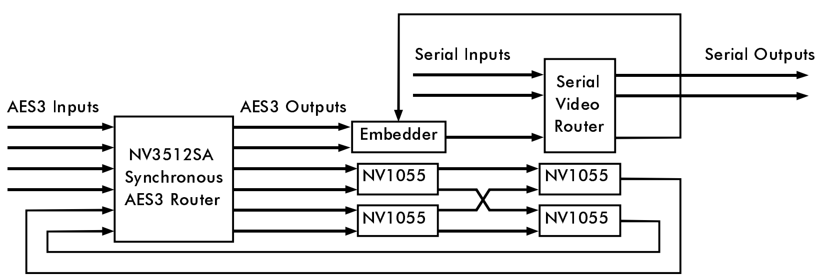


Figure 6.3 A large audio break-away configuration.

Synchronous routing does not cause sample corruption, and so a hot switch during a quiet section between programs will not be audible. However, a hot switch during program material could still cause a click if there is a disparity between the audio waveforms at the switch point. The same click would occur switching analog signals and the solution is the same for both: a short crossfade is needed. Fig.6.3 shows how a small digital mixer such as the NV1055 can be slaved to a synchronous digital audio router to perform crossfades instead of cuts.

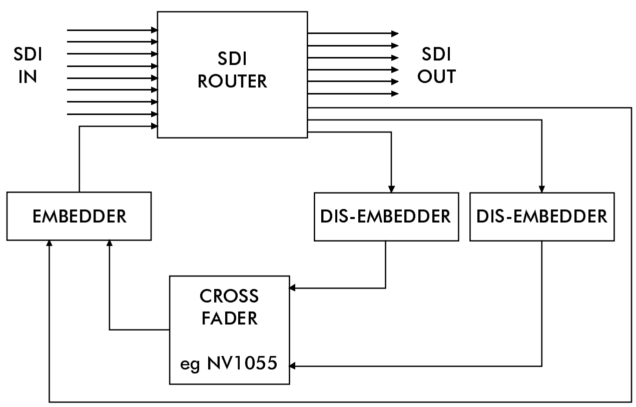


Figure 6.4 Hot cutter using cross-fades. Incoming audio is dis-embedded and fed to fader. After the cross-fade the previous input dis-embedder is no longer required and can be assigned to the next input.

Fig.6.4 shows the ultimate hot cutter for embedded audio. Two assignable de-embedders provide clean audio prior to video cuts and a crossfade is made between the audio signals prior to re-embedding.

6.6 Conventional routers operate on the entire AES/EBU signal which contains a pair of audio channels. What happens if you want to route individual audio channels? The answer is to use a sub-frame router. Fig.6.5 shows that in this useful device a pair of AES/EBU inputs carrying four audio channels are phase aligned and converted internally into four AES-like signals where the same audio channel appears in both sub-frames. This technique is used

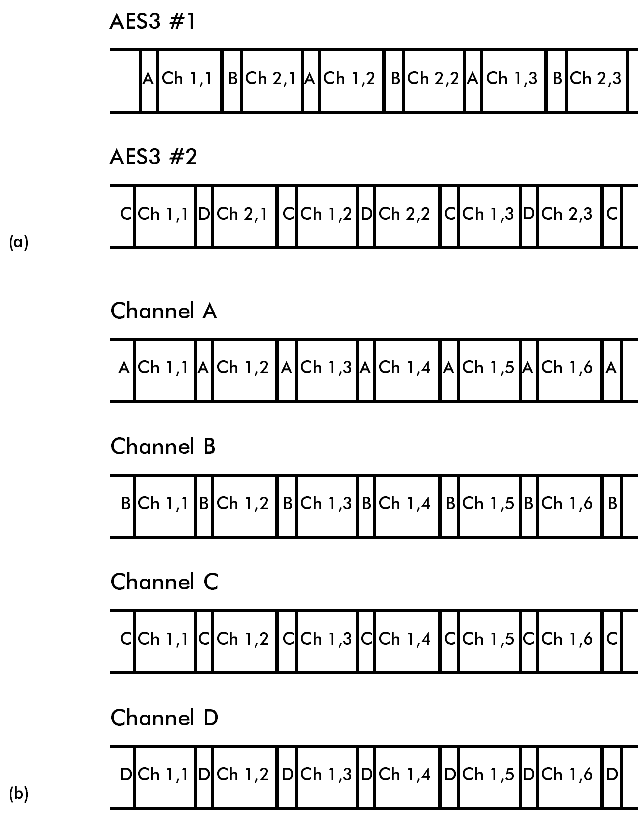


Figure 6.5 (a) Two AES3 inputs. **(b)** Phase aligned, four channel audio.

in the NV 1055 digital mixer which is shown in Fig.6.6. The NV 1055 is a four layer 4 x 1 mixer where each output is a linear mix of the four inputs. Using zero and unity gain coefficients the NV 1055 can route any audio channel to either sub-frame of either output. Because of the mixing facilities, the routing function can be implemented with true cross fades, eliminating audio waveform jumps on hot cuts.

6.7 Whilst major installations will require a large central synchronous router, there are many applications, such as small suites and project rooms which need smaller, simpler

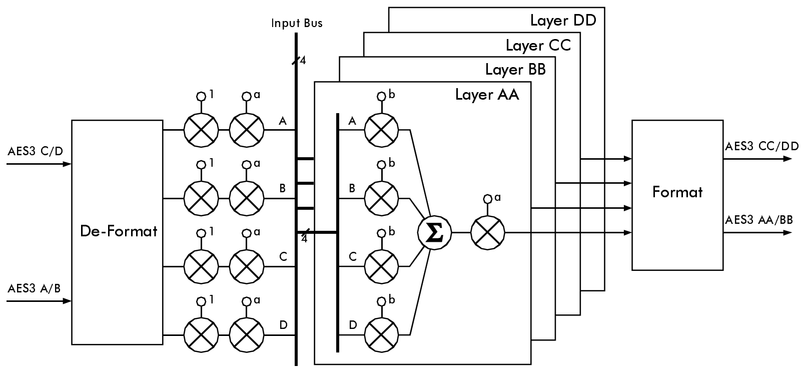


Figure 6.6 NV1055 conceptual architecture.

asynchronous routers. The NV1038 is an 8 x 8 router element which is easily cascaded to make larger routers. Fig.6.7 shows how a 16 x 16 router is built using four elements.

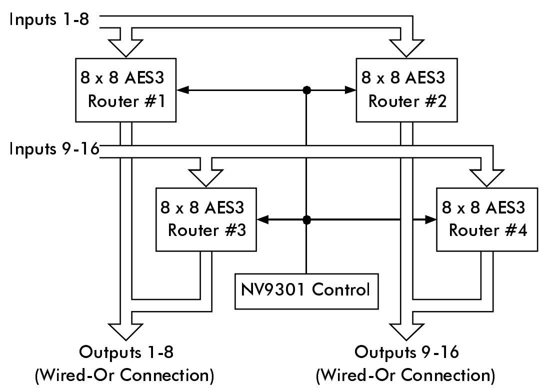


Figure 6.7 A 16 x 16 Matrix with 8 x 8 building blocks.

The input wiring requires some care in order to comply with transmission line practice. As section 4 showed, loop through can be used provided all but the last device is set to high impedance. Fig.6.8 shows the example of four elements in an 8 x 32 matrix.

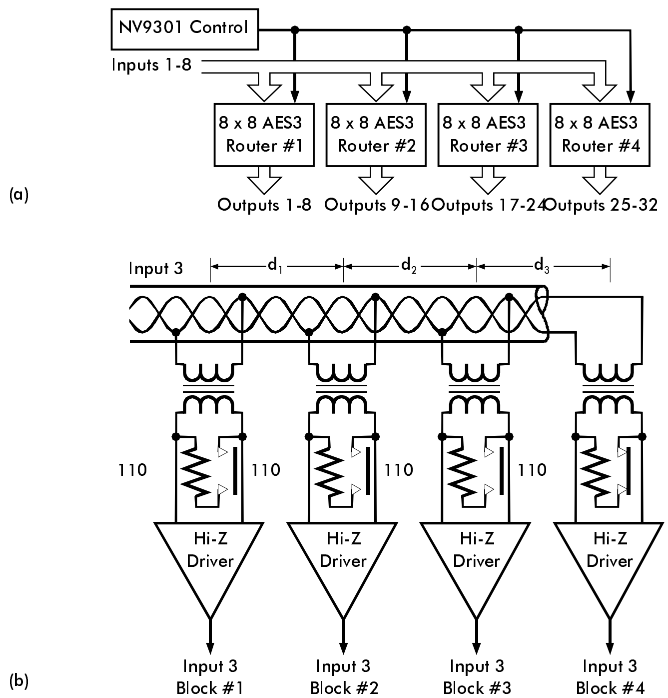


Figure 6.8 A conceptual 8 x 32 matrix (a). A detail of one AES3 input bus

Note how three of the elements have their terminating switch left open; the last one is closed to give the line 110 Ohm termination. Distances d1, d2 etch must be short compared to the cable run from the source.

Output cascading requires a little more thought because we can't have two outputs trying to drive the same cable. This problem is solved using special drivers which can adopt a high impedance state. Fig.6.9 shows that several drivers can then be wired to the same output provided that the control system set all but one to a high-Z state. Once more transmission line practice must be considered. When operating, a driver has an output impedance of 110 Ohms, however, all but the last driver sees an unterminated stub in parallel with the terminated output cable. This stub can cause reflections which potentially reduce the system data integrity unless distances d1, d2 etc. are kept to less than a couple of feet.

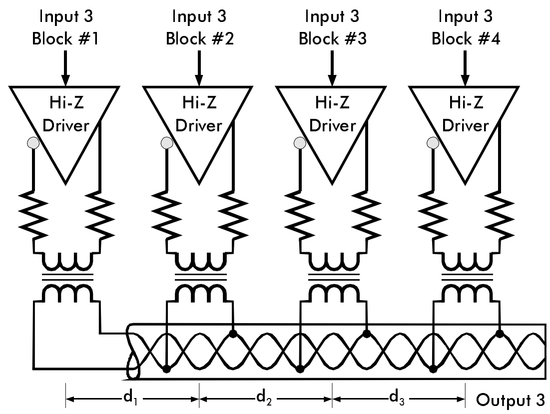


Figure 6.9 A detail of one AES3 output bus.

The control system is simplified because each 8 x 8 element can be set to a different address on the RS 422 control bus and is configured based on the input and output addresses associated with the whole matrix. In this way an external intelligent controller can operate the cascade as if it were one large router.

6.8 Timecode is a vital signal yet it is often taken for granted. Although timecode signals fit within the bandwidth of audio, it is often forgotten that they are actually digital signals. Time code has a data rate of 2.4 kBits/sec which is modulated with bi-phase coding to produce a DC-free signal where most of the power is concentrated in a bandwidth from 1.2 to 2.4 kHz. Good practice requires that digital signals are sliced and re-clocked at each stage of transmission to ensure data integrity. Reclockers output a clean constant amplitude trapezoidal waveform as specified in SMPTE 12M. Analog distribution and routing doesn’t do that and the result is signal rounding and jitter which impairs the signal. To make matters worse, many recorders don’t reclock the replayed T/C, they simply amplify the signal from the heads resulting in a signal amplitude that varies with tape speed. In the absence of standards for the electrical interface, timecode signals are found in balanced XLR and unbalanced BNC form.

6.9 Fortunately there is an alternative to this flaky, crossed fingers approach. All NVision T/C routers use correct slicing and reclocking. Fig.6.10 shows that the input signal is sliced to a binary signal, rejecting noise and waveform distortions. Routing is actually done with this binary signal in simple logic gates.

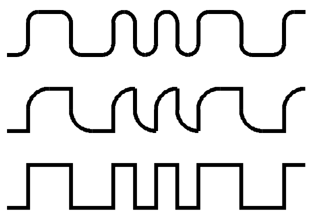


Figure 6.10 Time-code input signal processing.

Following the routing step, the binary signal is processed as shown in Fig.6.11 to re-create the textbook-perfect SMPTE 12M T/C signal having trapezoidal edges whose rise time is a function of frequency. This performance is maintained over a tape speed range from 1/30 normal to 100 x normal. The balanced/unbalanced dilemma is resolved by inputs and outputs which are designed to work with both. All inputs use a balanced amplifier which accommodates balanced signals directly or unbalanced signals by tying the inverting input to the shield. The output stages use an electronic virtual transformer which normally produces a balanced

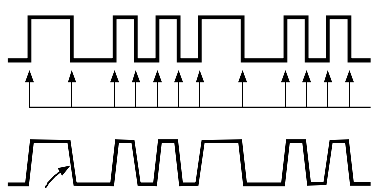


Figure 6.11 Time-code output signal processing.

output. However, if one output is tied to shield, the result is an unbalanced signal with the same amplitude. Fig.6.12 shows these simple unbalancing techniques.

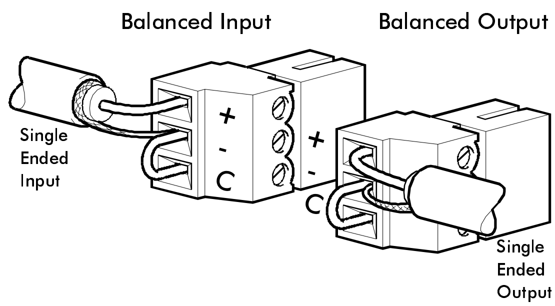


Figure 6.12 Balanced and single ended time-code connections.

Whilst giving optimum data integrity to T/C signals and interfacing balanced or unbalanced equipment, the NVision T/C routers actually use less space and power than traditional analog routers. For example the NV 3064TC provides a 64 x 64 matrix in only 6RU and requires 45 watts. An analog router might need three times that space and ten times the power. In practice the power saving is greater because you don't pay extra for the air conditioning to take the heat away. NVision T/C routers easily slave to your existing control system, eliminating changeover grief.

6.10 Unlike audio, video and timecode signals which only travel from the source to the destination, control signals such as the RS 422 standard used in SMPTE 207M are bi-directional. Commands can be issued on one signal pair and status can be returned on another. Instead of source/destination we need terms such as master/slave and controller/controlled as Fig.6.13 shows. Edit controllers are easy to grasp; they always want to be the master. In contrast, VTRs have split personalities. During the day a VTR might be a slave to an edit controller; by night it's the master for duplication. SMPTE 207M calls for commands to be transmitted on

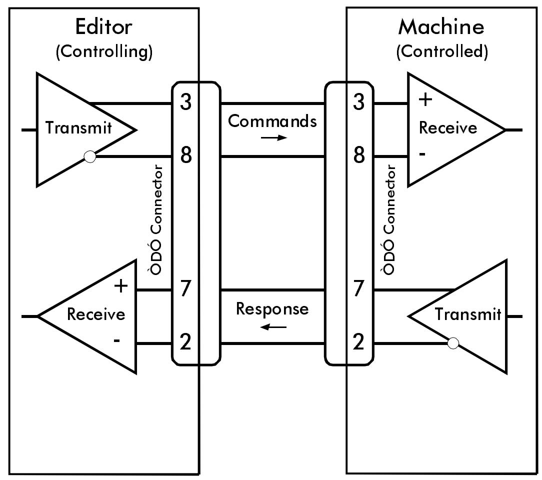


Figure 6.13 SMPTE 207M Port definition and pin-out.

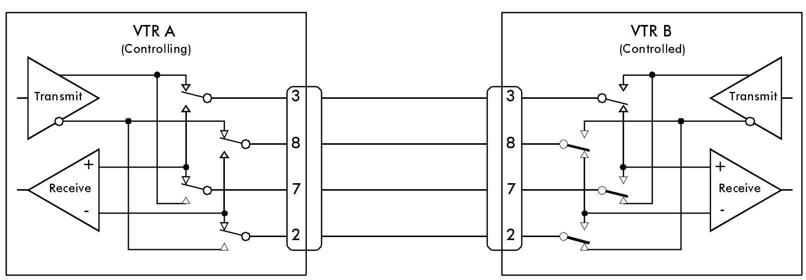


Figure 6.14 Controlled/ControllingVTR circuitry.

pins 3 and 8 and cables which are wired pin-for-pin. If a device wants to change from a slave to a master, it has to stop receiving on pins 3 and 8 and drive them instead. Fig.6.14 shows how a VTR changes its personality; it simply swaps over the pin pairs on the connector so that command and response signals exchange.

Unless the device at the other end follows suit we have two drivers fighting over the voltage on the wires. A conventional router can't change and often blown driver chips were the result. On later VTRs two connectors were fitted, and hardwired in such a way that one would have the right pinouts for master operation and one would have the right pinouts for slave operation. This concept is shown in Fig.6.15. The idea was that a two layer router would handle the master and slave signals separately. Unfortunately, the hardwiring causes command signals entering the slave socket to appear on the response input of the master socket. Unless the router output driving the response signals to the VTR can go to a high impedance state, once more there is a contention. Having a two layer router raises the cost and reduces flexibility.

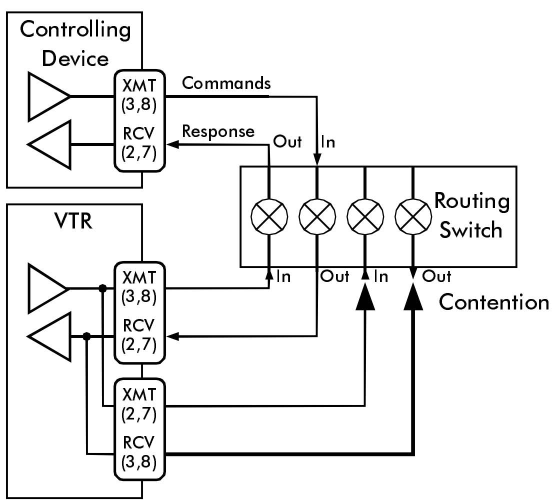


Figure 6.15 RS-422 data contention with standard routers.

The approach taken in the NV3128D machine control router is simply to provide a single layer bi-directional router which can dynamically change its ports between master and slave operation. Unless a control path is set up, the router remains in a high impedance state to prevent contention damage. Only a single cable is needed to each machine and because the master/slave switching is automatic, there are no switches to fiddle with and no time wasted. Fig.6.16 shows a dynamic port in detail.

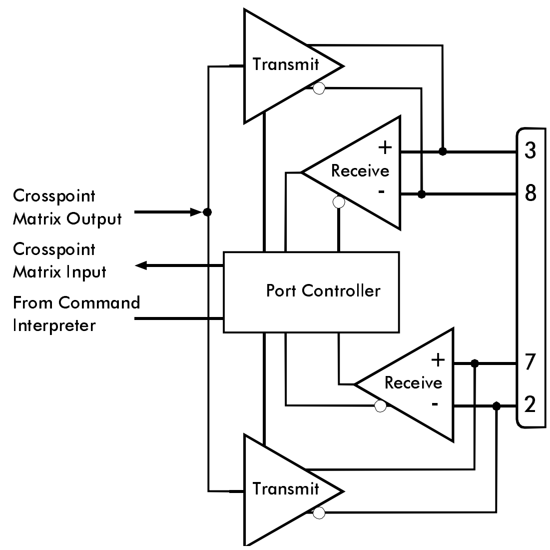


Figure 6.16 RS-422 Dynamic port detail.

A further machine control problem occurs when routers switch a different VTR to an editor. If the switch happens two quickly the edit controller doesn't realize that the VTR has changed and may carry on using the wrong protocol. The NV 3128 overcomes this problem by inserting a time-out pause between dropping the old machine and selecting the new one. The editor senses the communication loss and re-establishes protocol. Fig.6.17 shows a machine control router with dynamic ports.

6.11 In addition to the balanced RS 422 standard, there is a lot of RS 232 equipment around. RS 232 uses a single unbalanced wire for each direction. A master or controller device is called DTE (Data Terminal Equipment) whereas a slave or controlled device is called DCE (Data Communication Equipment). RS 232 ports do not change dynamically: if a piece of equipment is built as a DTE, that's what it always is. Fig.6.18 shows an RS 232 link. Note the

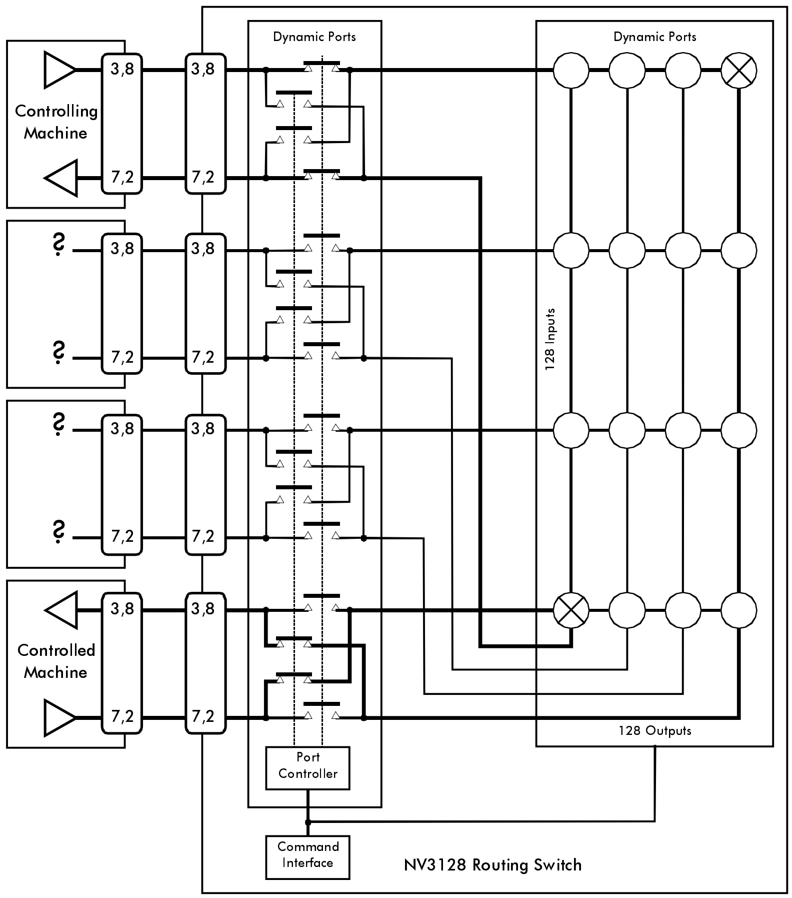


Figure 6.17 Dynamic ports in an X - Y architecture.

additional signals which are used for handshaking. If the software in the machine is designed for "smart modem" operation, these extra handshake lines can be tied up as shown in Fig.6.19 so that each device effectively shakes its own hands. The interface then defaults to a transmit line, a receive line and a common line which is needed because the signals are not differential.

When configuring a system with RS 232 devices, it is only necessary to know whether a given device is DTE or DCE. The

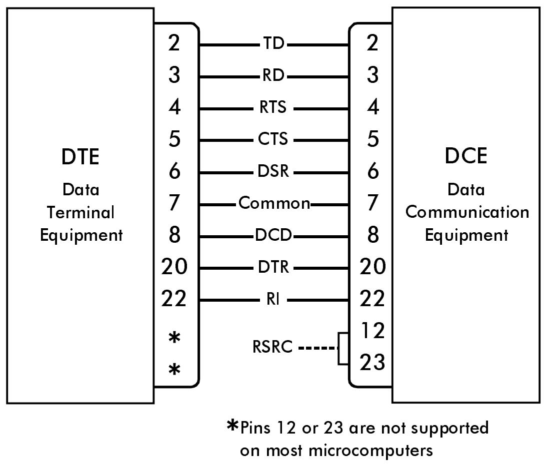


Figure 6.18 Industry standard 25 pin RS-232 connection.

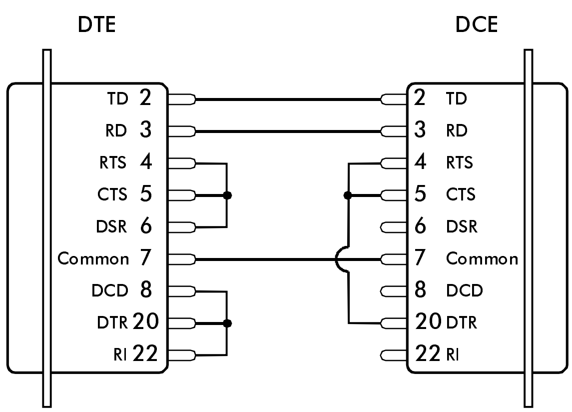


Figure 6.19 An RS-232 smart modem connection

router can then be programmed to act as the opposite type of device. In fact this can be done by the same dynamic port principle as is used for RS 422. The only difference is that the port never changes. The port type is set up using the NVUTILS configuration program and remains in non-volatile memory until the configuration needs to be changed.

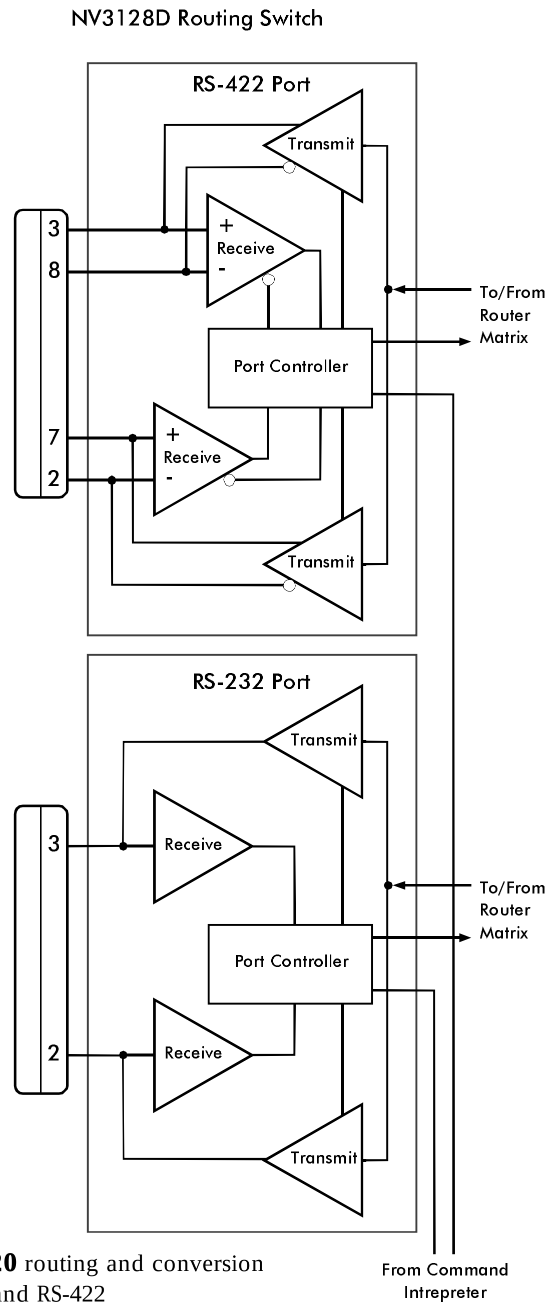


Figure 6.20 routing and conversion of RS-232 and RS-422

Fig.6.20 shows that it is easy to interface RS 422 to RS 232 equipment capable of "smart modem" operation.

6.12 Router control equipment can represent a significant investment. Configuring a large system to work the way you want it can take many man hours of effort. This effort should not be lost when the router is replaced. A further consideration is that the existing control system is familiar to the users. NVision philosophy is that a new router should be able to work with the existing control system.

All NVISION routers are provided with a Control Interface Module. This device is capable of interpreting the commands of a large number of router control systems. RS 422 and parallel interfaces are supported allowing your new NVision router to function at any level with any address space. Table 6.1 shows the currently supported protocols. Please call if your system is not listed here.

Serial RS-422

MFG.	FORMAT
BTS	Remote Binary Protocol ES-Bus
ProBel	RS-422/RS-485
Utah	AVS1, AVS2 PL320, PL160
Vistek	RS-422

Serial or Parallel
Hardware

MFG.	FORMAT
GVG	Horizon
PESA	37 Pin Parallel

Table 6.1 Currently supported protocols.

The router matrix status and configuration options are stored in SRAM with capacitor backup. Table 6.2 shows the dynamics of the SRAM. Should the control system fail the last map is retained indefinitely.

Matrix Refresh Time	512 outputs in less than 16 msec, continuous.
Matrix Storage Element Retention	20 minutes guaranteed, 12 hours typical.
Matrix Initialize on Power Up	Last known state/hard coded default, selectable.

Table 6.2 Matrix memory dynamics.

Router transition timing is critical and depends upon the references available. Table 6.3 shows the priority structure of the reference system.

Reference Priority 1	Video Vertical, NTSC or PAL
Reference Priority 2	AES3 External Frequency Reference
Reference Priority 3	Internal Free Running AES3 Oscillator
Reference Priority 4	Software Watchdog, Execute each Command as Received

Table 6.3 Transition timing priority structure.

6.13 To further speed the incorporation of an NVision router into your system, a software package known as NVUTILS is provided. This works on any IBM compatible PC running DOS 3.1 or above and communicates with the diagnostics port of the router. This is located near the front of the Control Interface Module. Possibly the most useful program in NVUTILS is called NVMAP. NVMAP allows the following functions:

1/ Configuring router levels. Allows setting up of two partitions which is useful for creating AES/EBU 1/2 and 3/4 layers or for routers which have asynchronous audio and timecode modules in the same frame.

2/ Physical and logical addresses can be assigned. A new router will probably not have the same physical addresses as the old. This is no problem because address reassignment allows the old control system to function as normal. All multi-level or multi-layer salvos programmed into the old system will still work.

3/ Layers can be logically linked. Any time a take command is issued, its linked source and output are automatically connected. AES/EBU 1/2 can be linked to 3/4. This is useful for audio-follow-video operations.

4/ For maximum flexibility, a router can be configured as 1/2 and 3/4 linked as in function 3, but retaining the ability to function as a flat router. The linked router and the flat router are logical levels. A priority structure is provided to determine which layer wins should a conflict arise.

5/ Because they have dynamic ports, NVision machine control routers can be configured for broadcasting. NV3128 routers allow many destinations to listen to one host but with only one destination responding. This is done with virtual layers as in function 4.

If you are starting out rather than replacing a router, NVTAKE is software which allows complete router control. Single takes and salvos can be executed. Salvos can be stored and retrieved. Execution is by a single keystroke.

6.14 The AES/EBU interface works well within its distance limitations, but for transmission over longer distances, perhaps between remote studios, an alternative is needed. The NV2000 system allows up to 10 channels of digital audio to be multiplexed into one bit stream which can be transmitted on digital coaxial or fiber optic lines, or on wideband analog video links.

Fiber optic technology uses two different techniques depending on the length of the link. When light pulses are launched down an optical fiber, they travel by internal reflection. Light traveling nearly parallel to the axis reflects less often than light traveling more obliquely and these light components travel different distances. At the receiver this multi-mode reflection shows up as edge smearing which gets worse with distance and limits the range to a few kilometers. The more expensive single mode fiber only allows propagation parallel to the axis of the fiber. This eliminates the edge smearing and allows much greater distances to be covered without a repeater; typically 50 kilometers.

Many Telcos have single mode fibers with enormous bandwidth. At each end, these fibers have multiplexers which allow several 6 MHz analog video signals to be combined into one optical signal. The NV 2000 can be used with such a single mode mux because it can create a 75 Ohm analog pseudo-video signal carrying the audio data.

Wave Division Multiplexing (WDM) is an exciting technology which allows more than one signal to share the same optical fiber without interference. Each signal uses a different wavelength, typically 1550 nm and 1300 nm. The wave division multiplexer is essentially an optical prism which separates light according to its wavelength. Fig.6.21 shows WDM being used to provide full duplex operation on a single optical fiber.
The NV 2000 produces a 18.432 Mb/sec data stream directly on optical fiber, on 75 Ohm digital coaxial cable, or as a pseudo-analog video signal. Each data stream can carry up to 10 audio

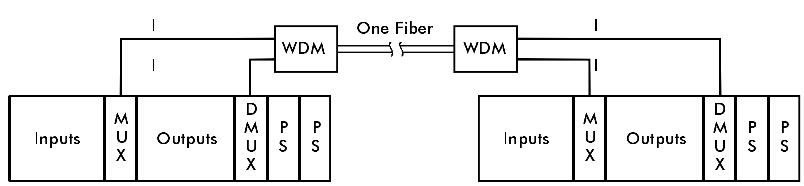


Figure 6.21 WDM's increase channel capacity.

channels. These can be input as analog or 48 kHz AES/EBU digital. The multiplexer locks to 48 kHz or to a video reference and all input modules get their clocks from the multiplexer. The demultiplexer automatically creates the same clocks from the transmitted signal which drive all the output modules. In duplex installations, the remote mux may be locked to the remote demux so that returning signals are synchronous.

In some applications, such as accessing a remote DVTR for audio layback production, it is necessary to transmit timecode and control signals in addition to the audio. Fig.6.22 shows that two audio channels in the multiplex can be replaced with an auxiliary channel carrying RS-422 control data, time code and cue audio.

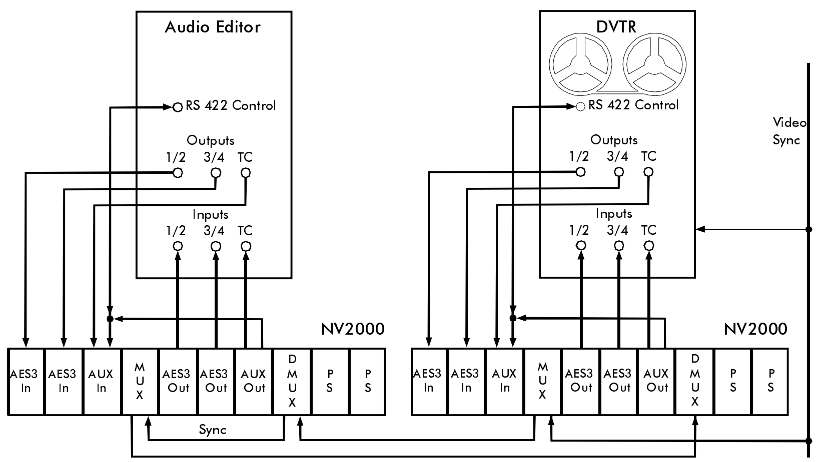


Figure 6.22 Remote audio layback.

Chapter 7

How do I go Digital?

If you are starting afresh, there is really only one choice today. Digital audio routing equipment offers significant savings over analog technology. It takes up less space, uses less power, needs less maintenance.

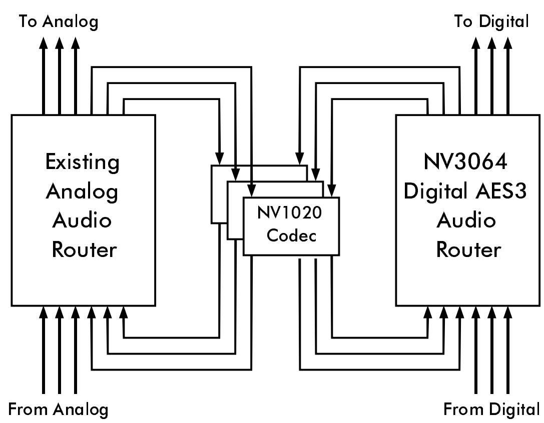


Figure 7.1 Mixed analog and digital audio routing.

Not many people are in the position to scrap an old router and start over. There is an investment in existing equipment and it is senseless to throw it out. In any case this is quite unnecessary. Since most new digital audio equipment will be equipped with AES/EBU interfaces, the best way of getting into the act is to add a digital audio router so that new devices can be incorporated into the system. Eventually, the digital router will replace the analog router, but initially, both work side by side. Fig.7.1 shows what happens. Instead of equipping all existing analog devices with ADCs and DACs at great cost, the analog router is retained and joined to the digital router with a smaller number of ADCs and

DACs. The control system of the digital router is slaved to the old router so that they work as one. When an old analog device is replaced, the new digital device goes on the new router. Eventually all devices are replaced and the old router goes.

Table 7.1 Mixed format and digital solutions

Step 1.	Determine the router size needed for analog equipment, digital equipment, and the analog/digital crossover. If the crossover exceeds 50%, use an all digital router. For calculation, use stereo audio crosspoints. AES3 digital accomodates stereo in one signal.
Step 2.	Add the number of crossover inputs and outputs to the dimensions of the analog and digital router matrices in Step.1
Step 3.	Calculate the cost of the stereo analog and digital routing matrices.
Step 4.	Add the cost of stereo A/D and D/A conversion used to bridge between the two formats. Use one A/D and one D/A for each stereo crossover.
Step 5.	This is the raw cost of the mixed format solution.
Step 6.	Add the number of stereo analog inputs and outputs to the number of digital audio inputs and outputs.
Step 7.	Calculate the cost of this digital audio router.
Step 8.	Add the cost of stereo A/D and D/A conversion for all analog equipment.
Step 9.	Compare the total in step 9. with that in step 5.
Step 10.	This ia the raw cost of the digital solution.

In order to judge which approach is best, try the assessment technique in Table 7.1. If the all digital approach turns out only slightly more expensive, bear in mind that it will offer the following:

a) Any material anywhere in the facility can be accessed without degradation.

b) Future proof; all future equipment will have AES/EBU interfaces.

c) Low expansion cost: digital crosspoints are cheap.

d) Analog equipment such as VTRs can be replaced by digital simply by bypassing the converters. No other change is needed.

e) Less power is required; less air conditioning load.

FURTHER READING

The following reading list (Focal Press) builds on the introduction given here, but covers all the necessary theory and practice of digital production.

Introduction to Digital Audio
Introduction to Digital Video
The Art of Digital Audio
The Art of Digital Video
The Digital Video Tape Recorder
The Digital Interface Handbook
Audio and Video Compression

References to products with ‘NV’ numbers refer to available NVISION products. Specific information on these items can be found in the 1995 NVISION catalog, available from NVISION Inc. or your nearest NVISION distributor.

NVISION®

125 Crown Point Court
Grass Valley, CA 95945
Phone +1 (530) 265 1000
Fax +1 (530) 265 1021